



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise de técnicas de clusterização em MMO com dados restritos: o caso de Final Fantasy XIV

Lucas Vanderlei Fernandes

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Orientador

Prof. Dr. Ricardo Pezzuol Jacobi

Brasília
2019



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise de técnicas de clusterização em MMO com dados restritos: o caso de Final Fantasy XIV

Lucas Vanderlei Fernandes

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Prof. Dr. Ricardo Pezzuol Jacobi (Orientador)
CIC/UnB

Prof. Dr. Bruno L. Macchiavello Espinoza Prof. Dr. Tiago Barros Pontes e Silva
Universidade de Brasília Universidade de Brasília

do Programa de Pós-graduação em Informática

Brasília, 21 de Dezembro de 2019

Dedicatória

Eu dedico esse trabalho aos meus pais, avós e tios que sempre acreditaram no meu potencial, assim como ao meu irmão que sempre me impulsionou a alcançar mais. Dedico esse trabalho aos meus amigos mais antigos: Henrique, Zerbini, Pedro, Bia e Felipão. Aos meus amigos de universidade que passaram dificuldades semelhantes e continuamos caminhando a frente: Dani, Rafael, GVT, Branco, Kell. Aos amigos que surgiram na vida e permanecem fazendo a diferença: Hoa, Fábio, Guilherme, Flavia, e Bahia. Aos companheiros do SILCE, que permitiram que eu pudesse concentrar neste trabalho, sem que nosso produto sofresse. E finalmente, à Lara, que foi minha inspiração e companheira por tantos anos.

Agradecimentos

Agradeço à Universidade de Brasília pela oportunidade de realizar meu mestrado. Assim como aos professores Prof. Dr. Carla Denise Castanho, Prof. Dr. Ricardo Pezzuol Jacobi, Prof. Dr. Mauricio Sarmet, e Prof. Dr. Tiago Barros Pontes e Silva.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

A utilização de dados de uma sessão de jogo para melhor compreensão do comportamento do jogador e os possíveis melhoramentos que podem ser realizados é um dos objetos de estudo da *Game Analytics*, domínio de pesquisa multidisciplinar que vem se difundindo amplamente entre pesquisadores da área de jogos eletrônicos. Entretanto, no caso dos MMOGs (Massive Multiplayer Online Games), os tipos de dados disponibilizados para análise não são padronizados, usualmente variando de um jogo para outro. Assim, um dos desafios desta área consiste em determinar o tipo de informação que pode ser obtida de um jogo MMO específico, assim como qual técnica de mineração de dados utilizar ou desenvolver em função da especificidade de sua base de dados. O objetivo deste trabalho é o estudo de técnicas de clusterização aplicadas ao Final Fantasy XIV, jogo que conta com uma base de milhões de jogadores mas disponibiliza apenas uma limitada quantidade de dados para análise e, portanto, tem sido pouco estudado na literatura. Os resultados obtidos poderão contribuir para uma melhor compreensão sobre os grupos de jogadores contidos em Final Fantasy XIV e fornecer uma base para o desenvolvimento de trabalhos futuros, além de prover um estudo de caso sobre técnicas de clusterização aplicadas sobre um limitado conjunto de dados de jogo.

Palavras-chave: Clusterização, Multijogador Massivo Online, *Game Analytics*, Final Fantasy XIV

Abstract

Using data from a game session to better understand player behavior and possible improvements that can be made in a game is one of the objects of study at Game Analytics, a multidisciplinary research domain that has been widely spread among researchers in the field of electronic games. However, for Massive Multiplayer Online Games (MMOGs), the types of data available for analysis are not standardized, usually varying from game to game. Thus, one of the challenges in this area is to determine the type of information that can be obtained from a specific MMO game, as well as which data mining technique to use or develop depending on the specificity of its database. The aim of this paper is the study of clustering techniques applied to Final Fantasy XIV, a game that has a player base of millions but provides only a limited amount of game data for analysis and, therefore, has been little studied in the literature. The results obtained may contribute to a better understanding of the Final Fantasy XIV player groups and provide a basis for future work, as well as provide a case study on clustering techniques applied over a limited set of game data.

Keywords: Clustering, Massive Online Multiplayer, Game Analytics, Final Fantasy XIV

Sumário

1	Introdução	1
1.1	Objetivos	3
1.2	Metodologia	3
1.3	Motivação e Relevância	3
1.4	Contribuições	5
1.5	Estrutura do texto	5
2	Fundamentação Teórica	7
2.1	Pesquisa	7
2.1.1	Game Analytics	7
2.1.2	Telemetria de Jogos	9
2.2	Jogos	9
2.2.1	Massive Multiplayer Online	10
2.2.2	Role Playing Game	10
2.2.3	Mecânicas de Jogos	11
2.2.4	Gameplay	11
2.2.5	<i>Patch</i> e Expansão	12
2.2.6	User Interface e HUD	13
2.3	Clusterização	13
2.3.1	Pré-Processamento	15
2.3.2	Abordagens e Algoritmos de Clusters	17
2.3.3	Avaliação do Cluster	31
3	Revisão Literária	34
3.1	Pesquisas no campo de Game Analytics	34
3.2	Pesquisas de <i>Game Analytics</i> sobre MMOs	36
3.3	Análises com foco no Jogador	38
3.4	Aquisição de dados através de <i>Game Log</i>	40
3.5	Aplicação de Clusterização	41

3.6	Pesquisas com o Final Fantasy XIV	43
4	Final Fantasy XIV	45
4.1	Trajetória do Jogo	45
4.2	Enredo	46
4.3	Criação de Personagem	48
4.4	Gameplay	50
4.5	Estilos de Jogo	51
4.5.1	<i>Tank</i>	53
4.5.2	<i>Healer</i>	55
4.5.3	<i>DPS</i>	55
4.5.4	<i>Crafters</i>	56
4.5.5	<i>Gatherers</i>	57
4.6	Atividades	58
4.6.1	<i>Main Scenario Quests</i>	58
4.6.2	<i>Side Quests</i>	59
4.6.3	<i>Dungeons</i>	61
4.6.4	<i>Trials</i>	61
4.6.5	<i>Raids</i>	62
4.6.6	<i>PvP</i>	62
4.6.7	<i>High End</i>	63
5	Base de Dados	64
5.1	Descrição	64
5.2	Estrutura	66
5.3	Conteúdo	66
5.3.1	Informações Contidas	68
5.3.2	Meses Disponíveis	71
5.3.3	Limitações	71
6	Proposta e Metodologia	73
6.1	Proposta	73
6.2	Recursos	74
6.3	Coleta de Dados	75
6.4	Pré-Processamento	76
6.4.1	Valores Ausentes	76
6.4.2	Conversão de Atributos	76
6.4.3	Agregação de Atributos	77

6.4.4	Normalização	78
6.4.5	Diminuição de Dimensão	80
6.4.6	Diminuição do Tamanho do Domínio	80
6.4.7	Escolhendo o número de Clusters	82
6.5	Validação de Clusters	82
6.5.1	Estatística Hopkins	83
6.5.2	Coeficiente de Silhueta	84
6.5.3	Índice de Calinski-Harabasz	85
6.5.4	Índice de Davies-Bouldin	85
6.6	Atribuição de Perfil	86
7	Resultados	87
7.1	Pré-Processamento	87
7.1.1	Diminuição do Tamanho do Domínio	87
7.1.2	Verificação de Dados Ausentes	88
7.1.3	Diminuição de Dimensão	88
7.1.4	Número de Clusters	92
7.2	Resultados dos Clusters	92
7.2.1	K-means	92
7.2.2	WARD	95
7.2.3	DBSCAN	96
7.2.4	BANG	96
7.2.5	SOM	96
7.2.6	Spectral	97
7.2.7	Fuzzy C-means	98
7.3	Avaliações	99
7.3.1	Estatística Hopkins	100
7.3.2	Avaliação dos Clusters Resultantes	100
7.3.3	Performance	101
7.4	Atribuição de Perfis	103
7.5	Perfis Não Identificados	109
8	Considerações Finais	111
8.1	Nova Base de Dados	111
8.2	Comparação dos Algoritmos	112
8.3	Perfis de jogadores de <i>Final Fantasy XIV</i>	113
8.4	Perspectiva	114

Referências	116
Anexo	127
I Tabelas de Clusters Resultantes	128

Lista de Figuras

2.1	Captura de Tela do jogo Mega Man	12
2.2	Captura de Tela do jogo Kingdom Hearts 0.2 Birth by Sleep - A fragmentary passage	14
2.3	Resultado de uma Clusterização Particional	18
2.4	Resultado de uma Clusterização Hierárquica	20
2.5	Resultado de uma Clusterização Baseada em Densidade	23
2.6	Resultado de uma Clusterização Baseada em Grade	26
2.7	Resultado de uma Clusterização Baseada em Modelo	28
2.8	Resultado de uma Clusterização <i>Spectral-Based</i>	29
2.9	Resultado de uma Clusterização <i>Fuzzy</i>	30
3.1	Abordagens de Clusterização por Artigos	42
4.1	Criação de Personagem	49
4.2	Configuração Padrão da HUD do PC	52
4.3	Configuração Padrão da HUD do Console	53
5.1	FFXIV Census	66
6.1	Método do Cotovelo	83
7.1	Execução da verificação por valores ausentes nos meses analisados de 2018	88
7.2	Mapa de Calor com as correlações entre atributos nos meses analisados de 2017	89
7.3	Mapa de Calor com as correlações entre atributos nos meses analisados de 2018	90
7.4	Mapa de Calor com as correlações entre atributos nos meses analisados de 2019	91
7.5	Método do Cotovelo nos meses analisados de 2017	93
7.6	Método do Cotovelo nos meses analisados de 2018	93
7.7	Método do Cotovelo nos meses analisados de 2019	93

7.8	Porcentagens média de cada Cluster Resultante do K-means	94
7.9	Porcentagens média de cada Cluster Resultante do WARD	95
7.10	Porcentagens média de cada Cluster Resultante do DBSCAN	97
7.11	Porcentagens média de cada Cluster Resultante do SOM	98
7.12	Porcentagens média de cada Cluster Resultante do Spectral	99
7.13	Porcentagens média de cada Cluster Resultante do C-means	100
7.14	Estatística Hopkins dos meses analisados	101
7.15	Média dos Coeficientes de Silhueta	102
7.16	Média dos índices de Calinski-Harabasz	102
7.17	Média dos índices de Davies-Bouldin	103
7.18	Média do tempo de execução dos algoritmos de Clusterização	104
7.19	Porcentagem Média dos Atributos do Perfil Iniciante	105
7.20	Porcentagem Média dos Atributos do Perfil Casual	106
7.21	Porcentagem Média dos Atributos do Perfil Dedicado	107
7.22	Porcentagem Média dos Atributos do Perfil Hardcore	108
7.23	Porcentagem Média dos Atributos do Perfil Intermitente	108
I.1	Descrição do domínio clusterizado em cada mês	129
I.2	Descrição dos Cluster Resultantes do K-means no mês 06/2017	130
I.3	Descrição dos Cluster Resultantes do K-means no mês 10/2017	130
I.4	Descrição dos Cluster Resultantes do K-means no mês 02/2018	131
I.5	Descrição dos Cluster Resultantes do K-means no mês 06/2018	131
I.6	Descrição dos Cluster Resultantes do K-means no mês 10/2018	132
I.7	Descrição dos Cluster Resultantes do K-means no mês 02/2019	132
I.8	Descrição dos Cluster Resultantes do K-means no mês 06/2019	133
I.9	Descrição dos Cluster Resultantes do WARD no mês 06/2017	133
I.10	Descrição dos Cluster Resultantes do WARD no mês 10/2017	134
I.11	Descrição dos Cluster Resultantes do WARD no mês 02/2018	134
I.12	Descrição dos Cluster Resultantes do WARD no mês 06/2018	135
I.13	Descrição dos Cluster Resultantes do WARD no mês 10/2018	135
I.14	Descrição dos Cluster Resultantes do WARD no mês 02/2019	136
I.15	Descrição dos Cluster Resultantes do WARD no mês 06/2019	136
I.16	Descrição dos Cluster Resultantes do DBSCAN no mês 06/2017	137
I.17	Descrição dos Cluster Resultantes do DBSCAN no mês 10/2017	138
I.18	Descrição dos Cluster Resultantes do DBSCAN no mês 02/2018	139
I.19	Descrição dos Cluster Resultantes do DBSCAN no mês 06/2018	140
I.20	Descrição dos Cluster Resultantes do DBSCAN no mês 10/2018	141
I.21	Descrição dos Cluster Resultantes do DBSCAN no mês 02/2019	142

I.22	Descrição dos Cluster Resultantes do DBSCAN no mês 06/2019	143
I.23	Resultados da clusterização do algoritmo BANG	144
I.24	Descrição dos Cluster Resultantes do SOM no mês 06/2017	145
I.25	Descrição dos Cluster Resultantes do SOM no mês 10/2017	145
I.26	Descrição dos Cluster Resultantes do SOM no mês 02/2018	146
I.27	Descrição dos Cluster Resultantes do SOM no mês 06/2018	146
I.28	Descrição dos Cluster Resultantes do SOM no mês 10/2018	147
I.29	Descrição dos Cluster Resultantes do SOM no mês 02/2019	147
I.30	Descrição dos Cluster Resultantes do SOM no mês 06/2019	148
I.31	Descrição dos Cluster Resultantes do Spectral no mês 06/2017	148
I.32	Descrição dos Cluster Resultantes do Spectral no mês 10/2017	149
I.33	Descrição dos Cluster Resultantes do Spectral no mês 02/2018	149
I.34	Descrição dos Cluster Resultantes do Spectral no mês 06/2018	150
I.35	Descrição dos Cluster Resultantes do Spectral no mês 10/2018	150
I.36	Descrição dos Cluster Resultantes do Spectral no mês 02/2019	151
I.37	Descrição dos Cluster Resultantes do Spectral no mês 06/2019	151
I.38	Descrição dos Cluster Resultantes do C-means no mês 06/2017	152
I.39	Descrição dos Cluster Resultantes do C-means no mês 10/2017	152
I.40	Descrição dos Cluster Resultantes do C-means no mês 02/2018	153
I.41	Descrição dos Cluster Resultantes do C-means no mês 06/2018	153
I.42	Descrição dos Cluster Resultantes do C-means no mês 10/2018	154
I.43	Descrição dos Cluster Resultantes do C-means no mês 02/2019	154
I.44	Descrição dos Cluster Resultantes do C-means no mês 06/2019	155

Lista de Tabelas

4.1	Discípulos de Guerra e Mágica	54
4.2	<i>Jobs</i> lançadas em Expansões	54
4.3	Discípulos da Mão e Terra	57
5.1	Estrutura da Tabela <i>tblplayers</i>	67
5.2	Relação de Atributos e Itens Físicos	70
6.1	Agregação de Atributos	78
6.2	Cont: Agregação de Atributos	79
7.1	Lista dos atributos do jogador a serem considerados na clusterização	92
7.2	Tabela de Identificação de Perfil por Algoritmo	105

Lista de Abreviaturas e Siglas

HUD Heads-up display.

MMO Massive Multiplayer Online.

MMORPG Massive Multiplayer Online Role-Playing Game.

MUD Multi-user dungeon.

RPG Role-Playing Game.

SOM Self Organizing Maps.

Capítulo 1

Introdução

Igualmente a qualquer empreendimento no mercado, a indústria do entretenimento busca o sucesso financeiro contínuo. Dentro destas atividades, o setor de jogos é um dos setores que mais cresce nos últimos anos [1]. Entretanto, com esse crescimento e com as evoluções tecnológicas, a competitividade nunca foi tão grande. A existência de diversas plataformas de jogo, tais como dispositivos móveis, consoles e computadores, tornaram o lançamento de novos jogos algo rotineiro [2]. O reflexo deste crescimento se manifesta no aumento do público e dos jogadores, cujo números estão cada vez maiores [3].

Um gênero que se beneficiou em decorrência desse crescimento de jogadores é o MMO, do inglês *Massive Multiplayer Online games*, jogos estes que necessitam de uma ampla base de jogadores para poderem serem rentáveis. Muitos MMOs foram criados nos últimos anos, porém poucos alcançaram o sucesso [4]. Diferente de outros gêneros, o sucesso financeiro de tais jogos dependem demasiadamente da permanência de antigos jogadores e conquista de novos. O alto consumo de tempo que eles exigem pode ser um problema para diversos indivíduos. Muitos jogadores tendem a parar de jogá-los em favor de outras alternativas, situação que ocorre no momento em que o jogador perde o interesse no jogo. Logo, o contínuo interesse do jogador é vital para sobrevivência deste gênero.

A fim de estabelecer uma vantagem sobre os demais desenvolvedores, muitos realizaram análises do processo de desenvolvimento do jogo e da experiência do usuário. A análise de dados provenientes de jogos digitais, ou *Game Analytics*, é uma tendência que está sendo adotada não apenas pela indústria, mas também pelos pesquisadores como uma ferramenta de investigação facilmente adaptável para muitos estudos [5, 6]. O uso de *Game Analytics* pode ser empregado em diversas etapas do ciclo de vida de um jogo, independente de seu gênero. Na etapa de concepção e *design*, seu uso pode se concentrar na procura de elementos de sucesso em jogos semelhantes ao que está sendo conceitualizado. Na etapa de desenvolvimento, a otimização do desempenho e tomadas de decisão sobre os mecanismos do jogo são alvos preferíveis de *Game Analytics*. O teste pode aplicar

as técnicas deste processo para validar melhorias e apontar falhas no jogo. Finalmente, quando o jogo já tiver sido lançado, *Game Analytics* pode prover informações valiosas sobre como melhorar o produto em uso. O uso e aplicação deste tipo de ferramenta de análise resulta em um produto com mais apelo ao consumidor final, no caso os jogadores.

Uma das técnicas mais utilizadas tanto em *Game Analytics* quanto em análises genéricas de dados é a clusterização [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Esta técnica agrupa objetos em análise a partir de características em comum, permitindo uma classificação de grandes volumes de dados, aspecto muito útil tratando-se de dados provenientes de jogos MMO. Entretanto, apesar de ser uma técnica muito utilizada, não existe um único algoritmo de clusterização que possa ser aplicado com sucesso para qualquer tipo de dados [25, 26]. Consequentemente existem diferentes abordagens e algoritmos para realização desses agrupamentos, gerando assim a necessidade de estudos para estabelecer a melhor abordagem para cada caso.

Neste cenário competitivo de jogos, a aplicação de técnicas de *Game Analytics* requer a coleta por vezes extensiva de dados sobre os jogadores para análise. Entretanto, tais informações não são sempre de fácil acesso. Uma fonte de dados acessível aos desenvolvedores é a base de dados gerada pelas ações do jogador e pela evolução do seu personagem ao longo do jogo. Enquanto a empresa que produz o jogo tem acesso a todas as informações coletadas, a totalidade desses dados não é sempre disponibilizada publicamente. Além das restrições ao tipo de informação disponível, outro problema igualmente fundamental é a interpretação desses dados, cujo propósito é a geração de informações úteis para análise. Este trabalho visa contribuir para a resolução do seguinte problema: como analisar jogadores de MMORPG a partir de um conjunto restrito de dados?

Neste contexto, o objeto de estudo desta pesquisa é *Final Fantasy XIV*, um jogo MMO com milhões de personagens [27] que, entretanto, disponibiliza uma limitada quantidade de dados abertos e tem sido pouco pesquisado na literatura. Apesar dessas restrições, ele possui uma quantidade expressiva e crescente de personagens que atualmente pode rivalizar até com o *World of Warcraft*, jogo mais popular do gênero [28]. Outro fator peculiar de *Final Fantasy XIV* é a liberdade que o jogador tem para explorar todas as formas disponíveis no jogo, ou seja, um personagem pode possuir todas as classes do jogo, as quais determinam o conjunto de ações que o jogador pode executar para interagir com o mundo virtual. É uma abordagem diferente da adotada pelo *World of Warcraft*, que limita o personagem de acordo com a classe escolhida, não permitindo troca de classes como ocorre em *Final Fantasy XIV*. Essa liberdade permite que o estilo de jogo do jogador se reflita mais precisamente nas ações do personagem, não exigindo a criação de outros personagens para explorar outras classes, evento comum em *World of Warcraft*. Reunindo milhões de jogadores, o *Final Fantasy XIV* apresenta grande potencial de interesse para

futuras pesquisas.

1.1 Objetivos

O objetivo principal desta pesquisa é determinar a(s) técnica(s) de clusterização de dados mais adequada(s) ao conjunto de dados disponibilizados pelo jogo *Final Fantasy XIV*. Um objetivo secundário, porém relevante, é expandir o domínio de jogos estudados nesta área de pesquisa. O incentivo para a inclusão do *Final Fantasy XIV* no repertório de pesquisadores será fornecido através da validação da base de dados estudada, assim como pela exploração das peculiaridades presentes em suas mecânicas, estas que fornecem uma maior liberdade para o jogador jogar como desejar.

1.2 Metodologia

Para que os objetivos discutidos acima possam ser alcançados será necessária a realização de diversos objetivos específicos no processo. Esses objetivos seguem a metodologia padrão de descoberta de novas informações [5]. Após a definição do objeto de estudo, isto é, dados dos personagens de *Final Fantasy XIV*, será feita a coleta dos mesmos através de um site que os disponibiliza mensalmente [27]. Com os dados armazenados, eles deverão passar por um processo de pré-processamento para permitir melhores resultados após a clusterização. Essa etapa será executada por meio de um algoritmo, em *Python*, que realizará o tratamento do conjunto de dados auxiliado por bibliotecas presentes na literatura para análise de clusters. Após a realização do tratamento dos dados coletados, o algoritmo desenvolvido irá aplicar as clusterizações de diferentes metodologias. Posterior a aplicação dos algoritmos de clusterização, o passo seguinte será a avaliação do resultado, assim como sua visualização. A avaliação do resultado será feita com auxílio de índices usados na literatura e a visualização será exibida com auxílio de gráficos e tabelas. A metodologia se encerra pela elaboração de um relatório com o resultado e aplicação do novo conhecimento. Por falta de acesso ao desenvolvimento do objeto de estudo, a etapa de aplicação não será parte do escopo desta pesquisa.

1.3 Motivação e Relevância

Pesquisas que visam a compreensão do comportamento humano enfrentam o desafio da generalização: quais técnicas empregar sobre qual amostragem da população alvo para que as conclusões possam ser estendidas para uma significativa gama de pessoas. Uma

das principais motivações dessa pesquisa deriva do alcance que MMOs possuem. O comportamento de milhões de usuários, assim como suas progressões, podem ser analisados, permitindo um melhor aferimento das conclusões em comparação com outros gêneros. A base de dados que será utilizada neste trabalho possui uma quantidade expressiva de registros. Por outro lado, existem muitas lacunas que dificultam a análise desse conjunto, tais como a ausência de dados sobre o período jogado, itens adquiridos e atividades feitas. Tal fato constitui uma grande parte da justificativa para esta pesquisa.

Sendo um dos MMOs altamente conhecidos, *World of Warcraft* é o alvo de diversas pesquisas, possivelmente o MMO mais pesquisado. Tais circunstâncias acabam gerando uma tendência que possibilita diversas novas pesquisas, estas que devem visar aspectos fora do domínio do *World of Warcraft*, como liberdade de troca de classes. Outra motivação deste trabalho é o incentivo para pesquisa acerca de um jogo com grande público e pouco explorado cientificamente. Possuindo uma história de redenção com seus fãs, após um lançamento tão negativo que acarretou no pedido de desculpas da desenvolvedora *Square Enix*, o jogo *Final Fantasy XIV* possui um grande potencial para pesquisa sobre seus jogadores, além de contar com uma base crescente de mais de quinze milhões de personagens.

Além das motivações citadas acima, trabalhos neste campo de pesquisa revelam sua relevância ao se observar a reação dos jogadores ao emprego desse conhecimento no desenvolvimento de novos jogos. A satisfação decorrente dos jogadores tende a resultar em uma maior confiança na desenvolvedora, que, por consequência, gera maiores vendas tanto para o jogo em questão quanto para futuros jogos.

Um exemplo dessa confiança pode ser observado nas vendas iniciais do jogo *Destiny*. Sua desenvolvedora, Bungie, foi responsável pelos principais jogos da série *Halo* dos anos 2001 a 2010. O legado criado sobre a recepção positiva dos jogos dessa série [29] refletiu em uma confiança na qualidade dos da Bungie. Consequentemente, o lançamento da nova Propriedade Intelectual, *Destiny*, acarretou em vendas milionárias [30]. Este foi o primeiro jogo desenvolvido após a desvinculação da empresa com a série Halo.

Em contrapartida, no ano de 2018, a desenvolvedora Activision Blizzard realizou um evento destinado a seu público alvo que atraiu atenção pelo impacto negativo do anúncio do jogo *Diablo Immortal*. Ao anunciar um novo jogo da franquia *Diablo* em seu evento, o público presente não correspondeu com agrado, incluindo até vaias e questionamentos sobre a validade do anúncio. A razão desta péssima recepção está possivelmente ligada à falta de compreensão por parte da desenvolvedora acerca do perfil de seu público alvo. O público do evento, composto principalmente por jogadores de computador, não aprovou o anúncio realizado na apresentação principal, que indicava que a nova interação da franquia *Diablo* seria um jogo exclusivamente para celulares [31].

Estes exemplos ilustram a relevância do conhecimento do público alvo para a indústria de jogos eletrônicos. Falhas na compreensão de seus interesses e necessidades podem resultar em perdas financeiras consideráveis e comprometimento do grau de confiança do usuário nas empresas, fator fundamental para o sucesso neste nicho altamente competitivo de mercado. Neste contexto, pesquisas que contribuam para a compreensão do perfil dos jogadores se revestem de demasiada importância.

1.4 Contribuições

A principal contribuição deste trabalho será o fornecimento de um embasamento sobre quais abordagens seguir ao clusterizar dados do *Final Fantasy XIV*, assim como a introdução de sua base de dados de personagens ao cenário científico. Com essas contribuições, é esperado que mais pesquisadores sejam persuadidos a realizarem pesquisas utilizando o *Final Fantasy XIV*, permitindo um maior aferimento de seus resultados quando corroborados por pesquisas com outros MMORPGs. Outra importante contribuição é a expectativa de que sejam identificados os diferentes agrupamentos de jogadores presentes no *Final Fantasy XIV*, assim como a dimensão de cada um desses perfis.

A abrangência do resultado desta pesquisa não está limitada ao domínio do jogos. Existem evidências de que o comportamento virtual dos jogadores pode refletir comportamentos do mundo real, possibilitando até a simulação da realidade por intermédio do jogo. Logo os *clusters* resultantes podem configurar características do jogador, algo muito valioso para análises comportamentais da psicologia. Analogamente, o estudo realizado sobre clusterização com dados limitados pode ser replicado com dados comportamentais de pessoas reais, aplicações que são beneficiadas por esta pesquisa.

1.5 Estrutura do texto

Esta dissertação está organizada da seguinte forma. O próximo capítulo tem como objetivo o nivelamento do conhecimento dos leitores com relação aos conceitos teóricos envolvidos nesta pesquisa. O terceiro capítulo apresenta o estado da arte neste campo de pesquisa, principalmente envolvendo *Game Analytics* e *MMORPG*. Após sua leitura, o leitor terá uma visão contemporânea sobre os assuntos pesquisados neste campo, permitindo a percepção da posição deste trabalho neste cenário.

O Capítulo 4 fará uma exploração do jogo *Final Fantasy XIV*, discursando sobre o legado da franquia, a história do lançamento do jogo, seu desenvolvimento, sua recepção e sua evolução como produto comercial. O mesmo também investiga o mundo digital do jogo, tais como, suas mecânicas, *gameplay*, atividades e histórias. O capítulo posterior

visa analisar e descrever a base de dados do *Final Fantasy XIV* que será usada, bem como discutir o conteúdo e as limitações dessa base.

O sexto capítulo expande a problemática e proposta que serão o foco desta pesquisa e, em seguida, descreve a metodologia que o trabalho adotará. O capítulo seguinte é composto pelos resultados das clusterizações utilizadas, juntamente como uma comparação da validação de cada um. Por fim, o último capítulo tem o intuito de analisar o resultado reportado e guiar contribuições de futuras pesquisas.

Capítulo 2

Fundamentação Teórica

Os conceitos abordados neste capítulo, essenciais para compreensão deste trabalho, estão estruturados em três seções. A primeira seção detalhará conceitos relacionados ao campo de pesquisa em geral deste trabalho, tais como *Game Analytics* e telemetria de jogos. A segunda está associada ao âmbito de jogos digitais e expressões usadas por jogadores. Finalmente o capítulo encerrará examinando cada metodologia e algoritmo de Clusterização que serão utilizados nesta pesquisa.

2.1 Pesquisa

As expressões presentes nessa seção, *Game Analytics* e Telemetria de Jogos, estão relacionadas com o âmbito científico deste trabalho.

2.1.1 Game Analytics

Com intuito de auxiliar na resolução de problemas de negócio, o processo de *Analytics* causou um impacto substancial na indústria [5]. Esse processo consiste na exploração de padrões presentes em dados, tanto em suas descobertas, quanto na comunicação de seus resultados. Outra aplicação de *Analytics* foca na utilização dos padrões de dados para previsão de estados futuros, estes que deverão ser usados para auxiliar em decisões corporativas, provocar ações ou melhorar o desempenho [5].

Este termo, *Analytics*, é um termo genérico que abrange toda a metodologia de definição e comunicação de padrões em dados. Uma investigação específica em um conjunto de dados não constitui todo o domínio de *Analytics*, sendo apenas uma análise em instâncias individuais aplicadas [25].

Os seguintes fundamentos são os pilares de *Analytics*: estatística, mineração de dados, matemática, programação, pesquisa operacional e visualização de dados [5].

Game Analytics é um domínio de aplicação específico da *Analytics*, constituído pela sua aplicação no contexto de desenvolvimento e pesquisa de jogos. O benefício direto obtido com a adoção da *Game Analytics* é o suporte para a tomada de decisões em todos os níveis e áreas de uma organização de desenvolvimento de jogos - do *design* à arte, da programação ao marketing, da gestão à pesquisa do usuário [25]. Assim, a *Game Analytics* pode ser vista como um processo para entender a experiência de jogo com o objetivo de melhorá-la [32].

Atualmente uma das principais áreas de análise de jogos se caracteriza por dar suporte a *Game User Research* (GUR). GUR é a aplicação de várias técnicas e metodologias de Psicologia experimental, Inteligência Computacional, Aprendizado de Máquina e Interação Humano-Computador para avaliar como as pessoas jogam e a qualidade da interação entre jogador e jogo. Em consequência, a análise do comportamento do usuário tem sido um importante direcionador para a evolução da *Game Analytics* na última década [5].

Game Analytics segue o processo padrão para a descoberta de conhecimento em dados [33], amplamente utilizado em análises orientadas por dados para descobrir conhecimento útil. As etapas para realização desse processo incluem:

- Definição de atributos;
- Aquisição de dados;
- Pré-processamento de dados;
- Desenvolvimento de métricas;
- Análise e Avaliação;
- Visualização;
- Relatórios;
- Implantação de conhecimento.

Duas das maiores desenvolvedoras de consoles para jogos, Microsoft e Sony, podem rastrear jogadores com seus sistemas de gravação de dados, assim como analisá-los com seus próprias ferramentas analíticas internas [32].

Atualmente, existem meios para jogadores acessarem gráficos analíticos de quantos jogadores ficam *online* durante o dia, ou quantas vezes uma certa arma é usada, mas estas são exposições superficiais, tendo em vista que os jogadores não têm acesso direto aos próprios registros. Com o objetivo de aumentar a sensação de comunidade em torno de um jogo e expandir sua relevância, será essencial para jogadores e desenvolvedores que empresas forneçam meios diretos de acesso a essas informações [32].

2.1.2 Telemetria de Jogos

Um dos pré-requisitos para qualquer aplicação de *Analytics* é a aquisição de dados. Neste contexto, o termo telemetria de jogos é usado para representar qualquer fonte de dados obtidos à distância. Dentre as variadas telemetrias presentes no âmbito de jogos, a telemetria de usuário se destaca por ser a mais utilizada [5]. Essa telemetria contém informações sobre o comportamento de jogadores, como por exemplo, quais interações o jogador fez com outros jogadores, ou com qual conteúdo o jogador interagiu [5].

A observação de dados não tratados de telemetria geralmente não agrega a compreensão do objeto de estudo [5]. O armazenamento e consequente ordenação desses dados pode fornecer diversas medidas interpretáveis a partir do domínio estudado. Uma banco de dados com informações de horas jogadas e conquistas atingidas pode gerar uma medida de tempo médio por conquista. Essas medidas interpretáveis são denominadas métricas de jogos [5]. As métricas podem ser variáveis simples, agregados mais complexos ou valores calculados, por exemplo, a soma de múltiplas variáveis.

Seguem os três tipos de categorias de métricas de jogos que já foram definidas [5]:

- Métricas de Usuário: são métricas relacionadas ao comportamento virtual e real de jogadores ao interagir com jogos.
- Métricas de Desempenho: são métricas cujo foco envolve a infraestrutura técnica responsável pela execução do jogo, assim como em softwares que permitem tal execução.
- Métricas de Processo: são métricas que visam avaliar o processo de desenvolvimento de jogos.

Para diferentes gêneros de jogos, podem existir diferentes prioridades em suas métricas. Este estudo abordará um jogo do gênero MMORPG, neste gênero as métricas de jogabilidade comumente utilizadas em análises são: progressão do personagem, missões completas, tempo para completar uma missão, uso de recursos, uso de itens ou habilidades, estatísticas de combate, desempenho contra inimigos de AI, progressão de história, interações personagens do jogo, desempenho de habilidade, dano recebido e fontes de dano e itens coletados [5]. A utilização dessas métricas é justificada pela característica de desenvolvimento do personagem essencialmente presente nesse gênero de jogo [5].

2.2 Jogos

Esta seção consiste de termos e conceitos relacionados aos jogos digitais.

2.2.1 Massive Multiplayer Online

Jogos do gênero *Massive Multiplayer Online Games* fornecem um mundo virtual para centenas ou milhares de jogadores se entreterem e interagirem simultaneamente, movendo-se através das áreas do jogo e completando diversos objetivos. Tanto a interação com o mundo virtual quanto a comunicação com outros jogadores é feita por intermédio de um avatar, ou personagem, que representa o jogador dentro do jogo [34].

Este gênero de jogo originou-se de um gênero anteriormente conhecido como *Multi-user dungeon* (MUD). Tais *mundos virtuais* não são simplesmente jogos no sentido tradicional baseado em regras. Dentro de seu domínio são encontrados mundos persistentes, sociais e materiais, vagamente estruturados por narrativas abertas (fantasia), onde os jogadores possuem, em grande parte, a liberdade para agirem como quiserem [35]. Existe uma variedade de gêneros associados aos MMOs, incluindo fantasia, ação, aventura, luta, quebra-cabeça, RPG(*role-playing game*), simulação, esportes e estratégia.

MMOs comerciais passaram a dominar a indústria de entretenimento digital e tornaram-se cada vez mais a forma dominante de entretenimento, particularmente para crianças e adolescentes por volta dos anos 2000s [36]. Como evidência de popularidade, a população de jogadores dos MMOs mais famosos variava de 70.000 a 10 milhões. Apesar de números impressionantes, outro fator atrativo a esse gênero é o nível de interação social disponível para os usuários. [37]

MMOs são distintos de outros gêneros, não apenas em relação ao conteúdo ou à vastidão dos mundos virtuais, mas também em relação à expansividade de seus canais de comunicação. Os jogadores podem transmitir mensagens para toda àrea, para uma equipe em particular ou diretamente para outro jogador. Normalmente, o sistema de bate-papo é integrado diretamente à interface do usuário do jogo. Em decorrência da existência de diversas atividades complexas, é praticamente uma exigência que os jogadores formem equipes cooperativas para superar desafios, que seriam insuperáveis caso não houvesse tal comunicação entre os usuários [38].

2.2.2 Role Playing Game

Role-Playing Game (RPG)(RPG) consiste em um jogo no qual os jogadores assumem os papéis dos personagens em um cenário fictício [39]. Esses mesmos jogadores assumem a responsabilidade de representar esses papéis dentro de uma narrativa, seja por meio de atuação literal ou por meio de um processo estruturado de tomada de decisão do desenvolvimento do personagem [40]. As ações tomadas dentro de muitos jogos apresentam sucesso ou falha de acordo com um sistema formal de regras e diretrizes [39]. Por meio de respostas do Mestre, o mundo do jogo fornece uma reação que é alimentada de volta aos

jogadores, que posteriormente processam a resposta, antes de fazer uma decisão sobre o que fazer a seguir [41]. Assim é criado um ciclo de interação entre jogadores e Mestre.

Existem várias versões de RPGs em mídia eletrônica, como MUDs baseados em texto e seus sucessores baseados em gráficos, MMORPG (*massively multiplayer online role-playing games*). Os jogos de interpretação de papéis também incluem jogos de RPG de jogador único nos quais os jogadores controlam um personagem ou equipe que realiza missões, e estes podem incluir parâmetros que evoluem usando mecanismos estatísticos para aumentar a probabilidade de sucesso. Esses jogos geralmente compartilham configurações e regras com RPGs de mesa, mas enfatizam mais o avanço do personagem do que a narrativa colaborativa [39, 42]. O meio eletrônico por outro lado elimina a necessidade de um Mestre, como em RPGs de mesa clássicos, e aumenta a velocidade de resolução do combate.

2.2.3 Mecânicas de Jogos

É comum ouvir jogadores, desenvolvedores, pesquisadores e *designers* de jogos tratarem acerca de mecânicas de jogos. Entretanto, se questionados sobre a definição de mecânica, os mesmos fornecem uma série de definições usadas em diferentes contextos [43]. Essa divergência demonstra que não existe um consenso acerca da definição de mecânica de jogo.

A partir da realização de um levantamento sobre diferentes definições utilizadas pela comunidade científica chegou-se a uma definição mais abrangente, que engloba as outras. Siscart [43] definiu a mecânica do jogo como métodos invocados por agentes para interagir com o mundo do jogo.

Dessa maneira, ações oferecidas pelo jogo ao jogador para superar desafios, assim como ações mapeadas a dispositivos de entrada, podem ser consideradas mecânicas de jogo.

2.2.4 Gameplay

Outro conceito explorado na indústria de jogos é o *gameplay*. Mesmo sendo considerado como um conceito que tem por objetivo descrever um jogo, não existe uma definição universalmente aceita. Entretanto, um estudo que questionou jogadores sobre o que consideram ser *gameplay*, apontou que as respostas sempre remetiam ao que poderia ser feito no âmbito do jogo, com foco no que o jogador conseguiria realizar e no que outras entidades podem fazer em resposta a ações do jogador [44].

Utilizando essa definição, a título de exemplificação, o *gameplay* de Mega Man, representado na Figura 2.1, consiste em movimentos de andar para direita e esquerda, pular, atirar, pausar o jogo, mudar de arma e escolher um estágio.



Figura 2.1: Captura de Tela do jogo Mega Man

2.2.5 *Patch* e Expansão

A manutenção de software foi facilitada pela proliferação da internet pelo mundo. Desenvolvedores de Software disponibilizam *Patches*, pela internet, para realizar diferentes ações nos produtos em produção. Um *Patch*, como definido pela Technopedia [45], é uma atualização de software o qual consiste em um código a ser inserido no código de um programa. Tal código geralmente consiste em concertos de *bugs* conhecidos podendo, entretanto, cumprir outros propósitos. Um *Patch* pode ter a funcionalidade de:

- Corrigir *bugs*;
- Disponibilizar conteúdo novo para ser jogado;
- Alterar parâmetros do jogo;
- Oferecer melhorias de Qualidade de vida;
- Melhorar segurança do produto.

Semelhante ao *Patch*, uma *Expansão* contém código para ser inserido em um programa. Entretanto seu propósito principal é a disponibilização de novo conteúdo e sempre é opcional, ao contrário do *Patch* que possui a liberdade de ser obrigatório caso os desenvolvedores optem por tal. Alguns jogos inserem o conteúdo da expansão dentro de um *Patch* e apenas disponibilizam no momento em que o jogador compra a expansão. Utilizando o jogo *Final Fantasy XIV* como exemplo, o *Patch 5.0* continha conteúdo novo em demasia, exclusivo para jogadores que adquiriram a expansão *Shadowbringers*. Por outro lado, correções de

bugs e melhorias de qualidade de vida eram feitas para todos os usuários [46]. Por ser um jogo totalmente *online*, por meio dos *Patches* os desenvolvedores tornaram obrigatória a atualização do jogo para os usuários que desejarem continuar jogando.

2.2.6 User Interface e HUD

A interface de usuário, ou *User Interface*, consiste em uma parte essencial de muitos softwares, pois ela administra o diálogo entre usuário e aplicação. Tais sistema podem ser descritos como uma composição de instâncias de objetos gráficos interativos [47].

Ao observar o ambiente de jogos na web e mobile, nota-se que a interface é adaptada para as opções de interação que o usuário possui. Em celulares ocorre a adaptação do jogo para que suas interações envolvam a tela sensível ao toque. Em jogos na web, a interação ocorre usualmente com *mouse* e teclado.

Outro conceito gráfico adotado em jogos é o Heads-up display (HUD). Um HUD é constituído por elementos visuais presentes na tela de jogo cujo propósito é indicar atributos do jogador [48]. Tais elementos são utilizados comumente para mostrar a vida do jogador, sua direção, quantos itens ele possui e sua posição em uma corrida, entre outros.

A Figura 2.2 demonstra vários elementos da HUD provenientes do jogo *Kingdom Hearts 0.2 Birth by Sleep - A fragmentary passage*. No canto superior direito, é exibido um mini-mapa para auxiliar navegação. No canto inferior direito, são exibidos atributos do jogador, como a vida, MP e *Focus* e um indicativo de que o estado do jogo está sendo salvo. No canto superior esquerdo são exibidas informações do objetivo para o jogador e um indicativo de uma nova conquista que ele conseguiu. Finalmente, no canto inferior esquerdo, são exibidas informações de direitos autorais e possíveis ações que o personagem pode executar. A interface do usuário na Figura 2.2 está presente nos comandos que o usuário pode interagir para fazer com que o personagem execute-os. O botão exibido ao lado da conquista também indica outra parte da *User Interface*. Outra interface do usuário neste jogo é o controle com o qual ele está controlando as ações de seu personagem.

2.3 Clusterização

Esta seção explora a definição de Clusterização, assim como aborda técnicas de pré-processamento, diferentes metodologias de agrupamento e avaliação de clusters resultantes.

A clusterização, análise de *clusters*, agrupamento ou armazenamento em *cluster*, refere-se ao processo de particionar um conjunto de objetos de dados em subconjuntos [25]. Ou seja, tal processo tem como objetivo resolver o problema básico: dado um conjunto de



Figura 2.2: Captura de Tela do jogo Kingdom Hearts 0.2 Birth by Sleep - A fragmentary passage

dados, particione-os em um conjunto de grupos onde seus membros são o mais semelhantes possíveis [26]. Esses grupos, ou subconjuntos, são chamados de *clusters*, de modo que os elementos dentro de cada *cluster* apresentam mais semelhança entre si do que com elementos de outros *clusters* [25]. A clusterização é uma técnica essencial na mineração de dados [49].

Tal processo faz parte da categoria de mineração de dados chamada de aprendizado não supervisionado, porque geralmente é realizado quando não há dados ou informações disponíveis relacionados à associação dos dados [49], podendo ser considerado o mais importante problema de aprendizado não supervisionado [50]. Uma das principais vantagens oferecidas pela clusterização é o fornecimento de uma melhor representação dos dados, pois todos os objetos em um *cluster* têm menos variabilidade em seus atributos e podem ser resumidos com eficiência [51].

O particionamento dos objetos em um *cluster* não é realizado mecanicamente, mas por um algoritmo de clusterização [25]. Existem vários algoritmos de clusterização, no entanto, não há uma solução única para todos os problemas [51]. A falta de um consenso acerca do "melhor" algoritmo é dada em razão de que cada algoritmo de *cluster* é projetado com certas suposições e possui seus próprios vieses [51].

A utilidade da clusterização é apresentada a partir das possibilidades de descoberta de grupos anteriormente desconhecidos de dados [25]. Com a intenção de capitalizar essa vantagem, muitas áreas aplicam diferentes técnicas de clusterização, incluindo: Filtragem colaborativa, Segmentação de Clientes, Resumo de dados, Detecção dinâmica de

tendências, Análise de dados multimídia, Análise de Dados Biológicos, Análise de Redes Sociais [26].

O pré-processamento necessário para clusterização será detalhado a seguir, juntamente com as metodologias seguidas pelas técnicas que serão utilizadas neste trabalho.

2.3.1 Pré-Processamento

Um analista de dados tem como desafios diversos obstáculos ao utilizar bases reais. As informações contidas nessas bases são altamente suscetíveis a inconsistências, ausências e ruído, devido ao seu tamanho [25]. Diversas técnicas podem ser usadas para superar esses desafios, cada uma focando em um desafio específico visando melhorar a qualidade dos dados que é definida em termos da acurácia, completude, consistência, credibilidade, interpretabilidade e período [25].

As técnicas citadas anteriormente podem ser agrupadas em diferentes etapas durante o pré-processamento dos dados: *Data Cleaning*, *Data Integration*, *Data Reduction*, *Data Transformation* e *Data Discretization* [25].

Acerca das etapas de *Data Integration* e *Data Discretization* as mesmas não serão tratadas durante o pré-processamento deste trabalho, tão somente serão apresentados seus conceitos. *Data Integration* envolve a integração de dados de diferentes fontes [25], algo que não será feito neste estudo pois apenas uma base de dados será analisada. A etapa *Data Discretization* não será detalhada no pré-processamento de dados devido ao objetivo deste trabalho que engloba a análise de agrupamentos. Esta etapa é caracterizada pela transformação de dados numéricos, feita mediante o mapeamento dos valores em intervalos ou conceitos [25]. Englobada por esse etapa está a análise de *clusters*, análise de histograma, análise de correlação e análise de árvores de decisão [25].

Data Cleaning

O objetivo de rotinas de *Data Cleaning*, ou limpeza de dados, é preencher os valores ausentes, suavizar o ruído ao identificar discrepâncias e corrigir inconsistências nos dados [25].

O tratamento de dados ausentes pode seguir diferentes métodos. O analista pode escolher ignorar a ausência do dado ou resolver preencher cada dado ausente com seu valor verdadeiro caso conhecido. Entretanto o primeiro método produz um resultado capaz de perder sua acurácia e o segundo método é custoso em demasia quando a base de dados é muito grande [25]. Outra abordagem comum envolve o preenchimento do dado ausente com um valor globalmente constante, um valor médio global, um valor médio local ou o valor mais provável obtido por técnicas de inferência [25]. Ressalta-se o fato

de que nem todo valor ausente é um erro na base, dados opcionais podem existir em seu interior.

Ruídos em dados consistem em erros randômicos ou variações em uma variável mensurável [25]. As técnicas de *Binning*, Regressão e Análise de *Outliers* são utilizadas para realizar o tratamento desses ruídos [25]. *Binning* suavizam um sequência de dados ordenada utilizando uma consulta a seus valores vizinhos. Tal sequência é dividida em *bins*, e todos valores dentro desses subconjuntos são substituídos pela média dos valores locais ou pelo valor de fronteira mais próximo do valor original [25]. Regressões lineares ou multilíneas podem ser usadas para transformar os valores em funções lineares, permitindo a dedução de próximos valores [25]. Análise de *Outliers* envolve o agrupamento dos dados e eliminando aqueles que não pertencem a nenhum agrupamento [25].

Data Reduction

Análises de base de dados imensas podem inviabilizar um estudo devido ao tempo e complexidade derivados de seus tamanhos. Técnicas de *Data Reduction*, ou redução de dados, buscam obter uma representação com menor volume da base de dados original que possua uma integridade semelhante aos dados originais, permitindo a produção de resultados analíticos úteis com menor custo [25].

Existem três estratégias para redução de dados: redução de dimensão, redução de numerosidade e compressão de dados [25].

As técnicas de redução de dimensionalidade procuram diminuir o número de variáveis e atributos aleatórios em consideração evitando que tempo e esforço sejam gastos para analisar variáveis ou atributos não correlacionados [25].

Para reduzir a numerosidade da amostra, as técnicas desta estratégia exploram a substituição dos dados originais por representações menores. Tais métodos podem ser paramétricos, nos quais um modelo é usado para representar a amostra e apenas os parâmetros são armazenados, ou não-paramétricos, onde são armazenados as representações de um conjunto de dados [25].

Finalmente, em técnicas de compressão de dados, são aplicadas transformações para comprimir a representação dos dados originais. Se essa compressão pode ser desfeita sem perda de informação, ela é considerada *lossless*, e *lossy* caso contrário [25].

Data Transformation

A última etapa do pré-processamento de dados é composta por transformações ou consolidações dos dados para que o resultado da mineração de dados seja de fácil entendimento e torne mais eficiente a mineração aplicada. Estão incluídas nesta etapa as estratégias de suavização (mencionada também na etapa de *Data Cleaning*), construção de atributos,

agregação, normalização, discretização e geração de hierarquia de conceito para dados nominais [25].

A construção de atributos visa auxiliar no processo de mineração criando novas características e adicionando-os aos dados originais. Técnicas de agregação visam resumir ou agregar dados com base em alguns critérios, um exemplo seria agregar dados mensais. A normalização é utilizada para fazer com que os dados trabalhados estejam localizados em um intervalo menor de valores. A discretização resulta na transformação de valores exatos em intervalos ou rótulos, permitindo a organização deles em hierarquias. A estratégia de geração de hierarquia de conceito para dados nominais conclui as estratégias de transformação de dados. Esta estratégia tem como objetivo a utilização da hierarquia de conceitos para generalizar conceitos presentes na base de dados original, trabalhando assim com menos conceitos, estes sendo de maior nível [25].

2.3.2 Abordagens e Algoritmos de Clusters

Nesta seção será feita o detalhamento das abordagens e algoritmos de clusterização que serão utilizados nesta pesquisa, tais como: *K-means*, WARD, DBSCAN, *Spectral*, BANG, SOM e *Fuzzy C-means*.

Abordagem Particional

A primeira metodologia de clusterização a ser explorada é *Partitional Clustering*, ou Clusterização Particional. As técnicas que seguem esta metodologia pretendem criar os *clusters* com base na construção de uma partição de dados. Cada partição representa um *cluster* e o número de clusters deve ser menor ou igual ao número de dados [25].

Existem dois critérios que algoritmos de particionamento devem garantir em seu resultado [49]:

- Cada *cluster* deve conter no mínimo um único elemento;
- Todo elemento deve estar relacionado a absolutamente um *cluster*.

Seguindo os critérios acima, os métodos desta metodologia realizam a realocação de elementos de um *cluster* para outro, iniciado de um particionamento inicial, até conseguirem atingir um limite definido de interações ou até nenhum elemento mudar de *cluster* após uma interação do algoritmo [52]. Esse particionamento inicial é pré-especificado pelo usuário, configurando esse método como semi-supervisionado [51]. A Figura 2.3 mostra o resultado de uma clusterização particional, utilizando o algoritmo k-means. Cada cor representa um *cluster* e a linha divisória entre os elementos representa a divisão das partições.

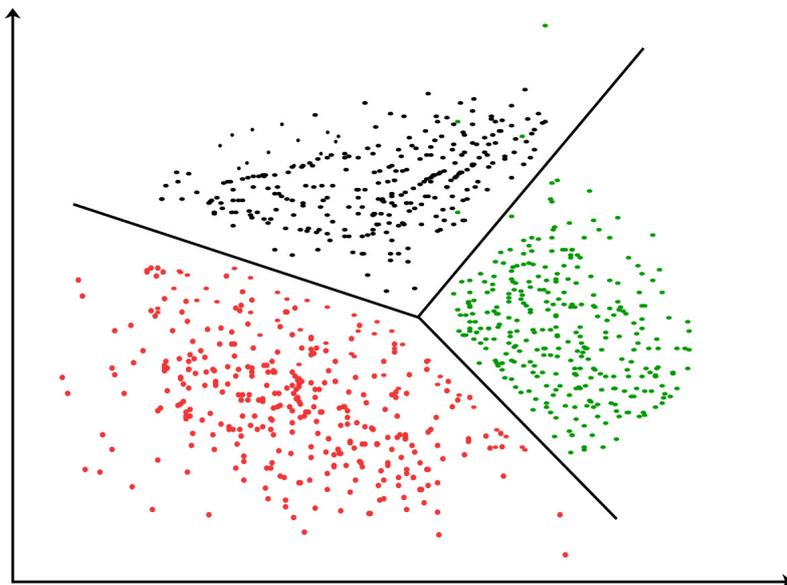


Figura 2.3: Resultado de uma Clusterização Particional [53]

Os métodos mais comuns da metodologia de clusterização particional são métodos k-means [49].

K-means

O algoritmo escolhido para representar métodos particionais de clusterização é o algoritmo mais utilizado nos trabalhos relacionados no capítulo sobre a revisão literária deste trabalho [7, 8, 9, 10, 11, 12, 13, 14, 15, 16], o *k-means*.

O fato deste algoritmo também ser um dos mais usuais em análises de dados com clusters [25, 26] gerou reflexos neste campo de pesquisa. Essa abrangência pode justificada pela sua simplicidade e efetividade [26].

Inicialmente, o algoritmo considera k pontos randômicos na dimensão dos dados na qual a clusterização foi aplicada. Estes serão os centros de cada cluster nesta primeira interação [26, 25]. A segunda etapa atribui cada ponto de dado ao centro mais próximo, essa aproximação é baseada em uma medida de distância escolhida, a distância euclidiana de n -dimensões é usada na implementação deste trabalho [54]. Tal distância pode ser calculada através da seguinte fórmula, onde p e q são pontos com n dimensões:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.1)$$

Em seguida, cada centro de cluster é alterado para a média dos valores de seus membros. As interações subsequentes repetem a atribuição de cada ponto de dado aos novos clusters e alteração do centros destes. O algoritmo irá repetir tais interações até que

o centro dos clusters não altere ou que alcance algum critério pré-determinado [26, 25]. A implementação utilizada irá realizar tais interações até atingir o valor padrão de 300 repetições [54]. Este algoritmo almeja minimizar o critério de soma dos quadrados dentro do cluster, ou seja, diminuir a inércia geral. Tal cálculo pode ser representado através da seguinte fórmula, onde o conjunto de dados é formado por $x_0 \dots x_n$, existem K clusters representados pelo conjunto C com os centros μ [54]:

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2) \quad (2.2)$$

O algoritmo *K-means* que será utilizado pode ser resumido nos seguintes passos, recebendo o parâmetro k como entrada em um conjunto de dados de n dimensões:

1. Inicialize k pontos n -dimensionais, marcando-os como centro de seus clusters;
2. Atribua cada ponto de dados ao centro de cluster mais próximo usando a distância euclidiana;
3. Altere o centro de cada cluster para a média de todos os valores de cada dimensão dos pontos de dados contidos nele;
4. Repita os passos 2 e 3 até que os novos centros não sejam distintos dos centros anteriores ou até que ultrapasse 300 repetições.

Abordagem Hierárquica

Devido ao fato já discutido de não existir uma solução universal para todos os problemas de clusterização [51], algoritmos hierárquicos foram desenvolvidos como alternativas aos métodos de *cluster* baseados em partições. O intuito era criar um mecanismo mais determinístico e flexível para agrupar os objetos de dados [26]. Além da abordagem particional um dos métodos mais populares é a metodologia Hierárquica [18, 17, 15, 19].

Os *clusters* obtidos por esse métodos são conhecidos como dendrogramas que mostram como os *clusters* estão relacionados entre si [49]. O relacionamento representado mostra os níveis de similaridade nos quais os agrupamentos são alterados. Um *cluster* dos objetos de dados é obtido cortando o dendrograma no nível de similaridade desejado [52].

Essa hierarquia de *cluster* também pode ser interpretada usando a terminologia de árvore [55]. A raiz representa todos os conjuntos de objetos de dados a serem agrupados em *cluster* e isso forma o ápice da hierarquia. Em cada nível, as entradas (ou nós) filhos, que são subconjuntos de todo o conjunto de dados, correspondem aos *clusters*. As entradas em cada um desses *clusters* podem ser determinadas atravessando a árvore do nó do *cluster* atual para os pontos de dados *singleton* base [26]. Cada nível na hierarquia corresponde a algum conjunto de *clusters* [26].

A Figura 2.4 mostra um conjunto de *clusters* obtidos por métodos hierárquicos, representados pelas letras de A até F. O dendrograma resultante é exibido no lado direito da imagem, com ele é possível descobrir a qual *cluster* um elemento aleatório irá pertencer.

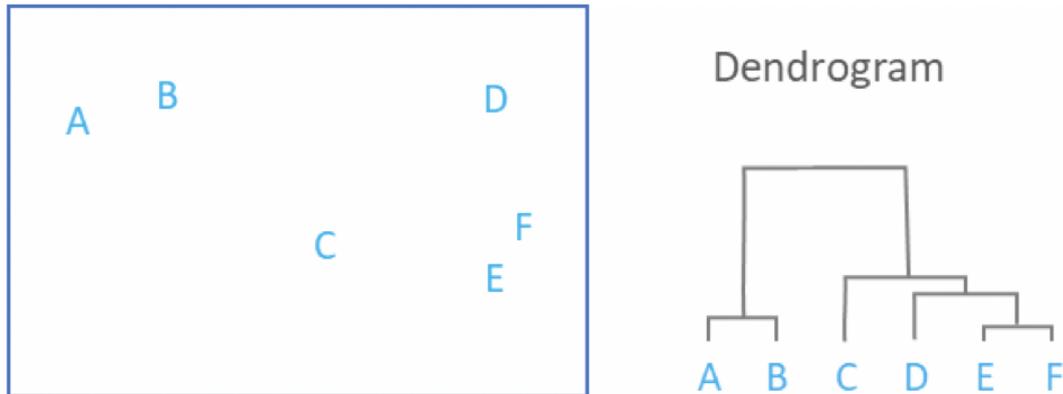


Figura 2.4: Resultado de uma Clusterização Hierárquica [56]

Os algoritmos desta metodologia de clusterização são divididos em duas classificações, dependendo da escolha de qual abordagem será usada. As duas possíveis abordagens são aglomerativo e divisivo [52, 49]. Ambas constroem os *clusters* particionando recursivamente as instâncias de maneira descendente ou ascendente. A abordagem aglomerativa considera, inicialmente, cada objeto como um *cluster* próprio. Em seguida, o algoritmo combina vários *clusters* menores em um *cluster* maior. Esse processo é repetido até que a estrutura de *clusters* desejada seja obtida, caracterizando-a como uma abordagem ascendente [52, 49]. A abordagem descendente, também conhecido como algoritmo hierárquico divisivo, parte do princípio de que, inicialmente, todos objetos pertencem ao mesmo *cluster*. Esse *cluster* será dividido em *clusters* menores, que são sucessivamente divididos em seus próprios *sub-clusters*. Esse processo é executado até que a estrutura desejada seja alcançada [52, 49].

A fusão ou divisão que ocorre em cada etapa do processo hierárquico, é realizada seguindo alguma medida de similaridade, escolhida para otimizar algum critério [52]. Considerando essa medida de similaridade, a maioria dos algoritmos desta metodologia podem ser categorizados em como ela é calculada:

- WARD: minimiza a soma das diferenças quadráticas em todos os *clusters*.
- *cluster de link único*: métodos que consideram a distância entre dois *clusters* igual à menor distância entre qualquer membro de um *cluster* e qualquer membro do outro *cluster* [52].

- *Cluster de link completo*: métodos que consideram a distância entre dois *clusters* igual à maior distância entre qualquer membro de um *cluster* e qualquer membro do outro *cluster* [52].
- *Cluster de link médio*: métodos que consideram a distância entre dois *clusters* igual à distância média de qualquer membro de um *cluster* a qualquer membro do outro *cluster* [52].

A maioria dos algoritmos de clusterização hierárquica são variantes dos algoritmos de *link* único, *link* completo e *link* médio [57].

Uma inconveniência do método hierárquico é o fato de que, uma vez feita a mesclagem ou divisão de clusters, ela nunca pode ser desfeita. Essa inconveniência pode ser útil, pois, conseqüentemente, um número combinatório de opções diferentes não é considerado, fazendo com que o custo computacional seja menor [50]. Entretanto, existem algoritmos que visam melhorar a qualidade do cluster hierárquico realizando análises das ligações dos objetos em cada particionamento, como CURE e Chameleon [50].

WARD

Dentre as possíveis combinações dos algoritmos hierárquicos a vertente mais aplicada foi a aglomerativa [18, 17, 15], com a utilização do critério de *Ward* como medida de similaridade mais frequente [18, 17]. Logo a implementação desta combinação será aplicada como representante da abordagem hierárquica [54].

O pseudocódigo descrito no Algoritmo 1 descreve um esboço básico de um algoritmo de clusterização hierárquico aglomerativo. Este inicia o processo calculando a matriz de dissimilaridade, utilizando alguma medida de similaridade, para todos pontos do conjunto de dados D . Em seguida é iniciado um loop onde pares de clusters são mesclados repetidamente, em concordância com metodologia aglomerativa, ou *bottom-up*. Após tal mesclagem, a matriz de dissimilaridade é atualizada. Na próxima interação do loop, tais operações são feitas na matriz de dissimilaridade atualizada. A última linha do Algoritmo

1 indica a condição de finalização [26].

Algoritmo 1: Pseudocódigo Clusterização Hierárquica Aglomerativa [26]

Data: D : Conjunto de n objetos de dados

Result: Conjunto de Clusters baseados em Hierarquia

```
1 compute a matriz de dissimilaridade entre todos pontos de dados de  $D$ ;  
2 do  
3   funda clusters como  $C_{a \cup b} = C_a \cup C_b$ ;  
4   defina a cardinalidade do novo cluster como  $N_{a \cup b} = N_a + N_b$ ;  
5   insira uma nova linha e coluna contendo as distâncias entre o novo cluster e os  
   clusters restantes;  
6 while não existir apenas um cluster maximal;
```

A medida de similaridade, relatada no algoritmo acima, aplicará o critério de *Ward* para computar a matriz de dissimilaridade na implementação utilizada neste trabalho [54].

O critério de *Ward* é empregado na medição da distância entre dois clusters. O cálculo do valor dessa distância é feito medindo o aumento no valor do critério da soma da estimativa quadrática dos erros para o cluster obtido por meio da fusão em $C_a \cup C_b$. O critério de Ward é representado pela seguinte forma [26], onde N representa a cardinalidade do grupo:

$$W(C_{a \cup b}, c_{a \cup b}) = -W(C, c) = \frac{N_a N_b}{N_a + N_b} d(c_a, c_b) \quad (2.3)$$

Portanto, o critério de *Ward* pode ser interpretado como a distância euclidiana ao quadrado entre os centróides dos clusters mesclados C_a e C_b ponderados por um fator proporcional ao produto das cardinalidades dos clusters mesclados [26]. O critério de Ward minimiza a soma das diferenças ao quadrado em todos os clusters. Logo, é uma abordagem que minimiza a variância [54].

Abordagem *Density-Based*

A distância entre objetos de dados é geralmente a base da maioria dos métodos particionais. Utilizando essa métrica, tais métodos encontram dificuldades em descobrir *clusters* com formatos arbitrários [25]. Para suprir essa dificuldade, algoritmos de clusterização baseados em densidade foram desenvolvidos [50, 49, 26, 51, 25]. Uma das vantagens dessa abordagem é a sua capacidade de administração de dados com ruído e a verificação dos dados em apenas uma varredura [49, 26]. Entretanto, esses algoritmos possuem uma restrição, que é a necessidade de inicialização dos parâmetros de densidade [55].

A clusterização baseada em densidade realiza uma separação entre regiões de baixa densidade e regiões de alta densidade de objetos na definição dos *clusters* [49]. Os métodos

desta abordagem assumem que os pontos que pertencem a cada *cluster* são extraídos de uma distribuição de probabilidade específica. Logo, a distribuição geral dos dados é assumida como uma mistura de várias distribuições [52]. O objetivo desses métodos é identificar os *clusters* e seus parâmetros de distribuição [52].

Os *clusters* resultantes são conectados, compostos por áreas densas no espaço de dados separados uns dos outros por áreas mais esparsas [26]. Por consequência, é correto assumir que a densidade dentro das áreas de ruído seja menor que a densidade em qualquer um dos *clusters* [26]. DBSCAN e SSN são dois algoritmos típicos desse tipo [50].

O resultado de uma Clusterização Baseada em Densidade é exibido graficamente na Figura 2.5. Diferentes formas de *cluster* podem ser observadas na figura: uma esférica, duas circulares e duas retangulares. Alguns objetos de dados classificados com ruído são exibidos como pontos pretos.

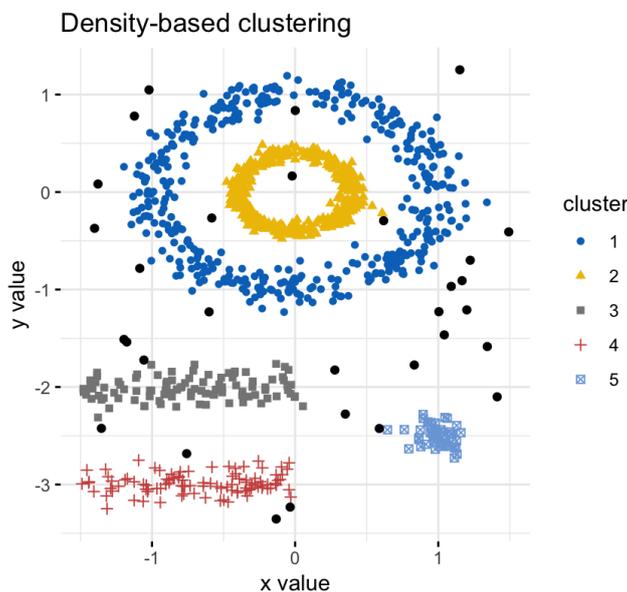


Figura 2.5: Resultado de uma Clusterização Baseada em Densidade [58]

Para conseguir chegar no resultado desejado, regiões densas de objetos de dados são combinadas para formar clusters [49]. O algoritmo baseado em densidade continua a crescer o cluster fornecido, desde que a densidade na vizinhança seja maior que determinado limite [55, 51, 52, 25]. Entretanto, não é obrigatório que todos objetos de dados pertençam à um cluster. Áreas esparsas no espaço de dados são tratadas como ruído e não são atribuídas a nenhum cluster [26].

DBSCAN

O algoritmo escolhido para representar clusterizações baseadas em Densidade é o algoritmo DBSCAN, (Density-Based Spatial Clustering of Applications with Noise) [25, 16].

Este algoritmo conecta objetos de dados e suas vizinhanças para formar regiões densas como agrupamentos [25].

Essa conexão é feita através das seguintes regras [26]:

- Um ponto é chamado de ponto central se a vizinhança do raio e contiver pelo menos $MinPts$ pontos, ou seja, a densidade na vizinhança deve exceder algum limite.
- Um ponto q é diretamente alcançável pela densidade a partir de um ponto central p se q está dentro da vizinhança e de p , e a alcançabilidade da densidade é dada pelo fechamento transitivo da alcançabilidade direta da densidade.
- Dois pontos p e q são chamados de densidade conectada se houver um terceiro ponto r a partir do qual p e q são alcançáveis pela densidade.
- Um cluster é então um conjunto de pontos conectados à densidade que é o máximo em relação à alcançabilidade da densidade.
- O ruído é definido como o conjunto de pontos no banco de dados que não pertencem a nenhum de seus clusters.

O algoritmo inicia marcando cada objeto de dados do conjunto D como "não visitado". Em seguida o algoritmo irá executar um loop até que todos objetos de dados sejam marcados como "visitados", finalizando o algoritmo após o encerramento dessas marcações. O primeiro passo dentro do loop é a escolha randômica de um objeto "não visitado" e sua marcação como "visitado". Após essa marcação é feita uma verificação se a vizinhança do ponto possui pelo menos $MinPts$ objetos. Em casos negativos, o ponto é marcado como ruído, ou *noise*. Em casos positivos, um novo cluster é criado, C , e o ponto é adicionado a este. Posteriormente, é feito outro loop onde cada ponto dentro da vizinhança do ponto em análise é avaliado se ele foi "visitado" antes e se ele pertence à algum cluster. Caso não pertença, será adicionado ao cluster criado. Caso não tenha sido visitado, será marcado como "visitado" e se sua vizinhança tem pelo menos $MinPts$ objetos, tal vizinhança é adicionada na vizinhança do ponto em análise original. Após esse loop interno, o loop externo reiniciará com a verificação se ainda existe pontos "não visitado" [25]. Esse algoritmo está representado no Algoritmo 2.

A implementação utilizada permite execução deste algoritmo com valores padrões, estes sendo $e = 0.5$ e $MinPts = 2 * nDimensoes$ [54, 59].

Abordagem *Grid-Based*

Algoritmos *Grid-Based*, ou Baseados em Grades, quantizam o espaço dos objetos de dados em um número finito de células que caracterizam uma estrutura de grades [55, 25]. Nessas

Algoritmo 2: Pseudocódigo DBSCAN [25]

Data: D : Conjunto de n objetos de dados, e : parâmetro de raio, $MinPts$: limite de densidade da vizinhança

Result: Conjunto de Clusters baseados em Densidade

```
1 marque todos objetos como "não visitado";
2 while existe um objeto "não visitado" do
3   |   randomicamente escolha um objeto "não visitado" $p$ ;
4   |   marque  $p$  como "visitado";
5   |   if a  $e$ -vizinhança de  $p$  tenha pelo menos  $MinPts$  objetos then
6   |   |   crie um novo Cluster  $C$  e adicione  $p$  em  $C$ ;
7   |   |   seja  $N$  o conjunto de objetos na  $e$ -vizinhança de  $p$ ;
8   |   |   for cada ponto  $p'$  em  $N$  do
9   |   |   |   if  $p'$  é "não visitado" then
10  |   |   |   |   marque  $p'$  como "visitado";
11  |   |   |   |   if a  $e$ -vizinhança de  $p'$  tem pelo menos  $MinPts$  objetos then
12  |   |   |   |   |   adicione aqueles pontos em  $N$ ;
13  |   |   |   |   end
14  |   |   |   end
15  |   |   |   if  $p'$  não é membro de nenhum cluster then
16  |   |   |   |   adicione  $p'$  em  $C$ ;
17  |   |   |   end
18  |   |   end
19  |   else
20  |   |   marque  $p$  como noise;
21  |   end
22 end
```

estruturas são executadas todas as operações para clusterização [55, 50]. Essas estruturas podem possuir vários níveis de células correspondentes a diferentes níveis de resolução [50]. Os clusters resultantes correspondem a regiões que são mais densas em pontos de dados do que em seus arredores [26].

Ao contrário dos métodos discutidos anteriormente, que agrupam os pontos de dados diretamente, as abordagens baseadas em grades agrupam a vizinhança ao redor dos pontos de dados representados pelas células [26]. Logo, a complexidade da clusterização é baseada no número de células de grade preenchidas e não depende do número de objetos de dados analisados [55]. A Figura 2.6 mostra o resultado de uma Clusterização Baseada em Grades, com cinco clusters, onde cada círculo representa o centro do agrupamento.

Esta abordagem é geralmente eficiente para muitos problemas de mineração de dados espaciais, incluindo clusterização [25] e mineração de grandes conjuntos de dados multidimensionais [26]. Esta é a principal vantagem desta metodologia, pois apresenta uma redução significativa na complexidade de tempo, especialmente para conjuntos de dados

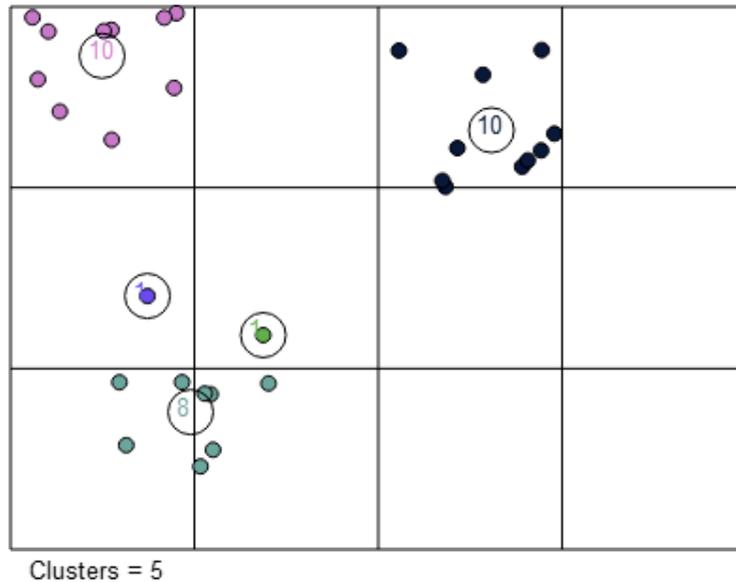


Figura 2.6: Resultado de uma Clusterização Baseada em Grade [60]

muito grandes [26, 52, 55]. Isso é consequência do tempo de processamento rápido, que normalmente é independente do número de objetos de dados e depende apenas do número de células em cada dimensão no espaço quantizado [25], aprimorando significativamente seu desempenho [26]. Duas consequências negativas desses métodos são que as formas dos clusters são limitadas à união das células da grade e são criadas em pontos de dados resumidos [55].

BANG

A abordagem de clusterização baseada em grades não foi aplicada nos trabalhos relacionados neste estudo. Entretanto, razão para tal fato não foi descoberta. Portanto um algoritmo desta abordagem será aplicado neste trabalho apesar de sua falta de popularidade aparente, este sendo o algoritmo BANG [61].

A implementação deste algoritmo pode ser resumida através dos seguintes passos [26]:

1. Particione o espaço do recurso em blocos retangulares que contêm até um máximo de dados p_{max} pontos.
2. Crie uma árvore binária para manter os blocos preenchidos, nos quais o nível da partição corresponde à profundidade do nó na árvore.
3. Calcule o dendrograma no qual os índices de densidade de todos os blocos são calculados e classificados em ordem decrescente.

4. Começando com o maior índice de densidade, todos os blocos vizinhos são determinados e classificados em ordem decrescente. O agrupamento BANG coloca as regiões encontradas no dendrograma à direita de os blocos originais.
5. Repita o Passo 4 para os blocos restantes do dendrograma.

Algoritmos desta metodologia utilizam estruturas de grades para agrupar os objetos de dados. Neste algoritmo, uma árvore binária é usada para armazenar essa estrutura da grade multidimensional para que a pesquisa de vizinhos possa ser feita com mais eficiência. A partir desta árvore no diretório de grade e das densidades de blocos classificados, o dendrograma é calculado. Centros de clusters são os blocos mais densos na fase de agrupamento [26].

Abordagem *Model-Based*

A Clusterização Baseada em Modelo realiza seus agrupamentos considerando a hipótese de que os dados a serem analisados foram gerados a partir de uma mistura de múltiplas distribuições de probabilidade, cujos parâmetros significam que a matriz de covariância deve ser estimada [51, 50]. Com base nessa hipótese, os métodos de clusterização baseados em modelo tentam otimizar o ajuste entre os dados fornecidos e algum modelo matemático [50, 52]. O modelo matemático geralmente usado para modelar cada componente são as distribuições normal ou gaussiana [50].

As técnicas empregadas nesta metodologia seguem duas abordagens principais para otimizar o ajuste a algum modelo matemático: abordagem estatística ou abordagem de rede neural [50]. Os métodos de indução mais frequentemente usados são árvores de decisão e redes neurais [52]. Exemplos bem conhecidos de Rede Neurais Artificiais usadas para a realização da clusterização incluem a quantização do vetor de aprendizagem de Kohonen (LVQ) e o mapa auto-organizado (SOM) [57]. A Figura 2.7 mostra o resultado de uma clusterização baseada em modelo utilizando mapas auto-organizados, com os objetos de dados representados no lado esquerdo da imagem e resultado ilustrado no lado direito.

A principal desvantagem desses algoritmos é que, semelhante ao *K-means*, ele também pode convergir para a solução ideal local, dependendo da atribuição inicial das k sementes [51]. Entretanto, os critérios de informação bayesiana ou Akaike podem ser usados para a seleção do número ideal de clusters [51].

SOM

Utilizado em outras pesquisas neste campo de pesquisa [23, 24], o algoritmo de clusterização baseado em Mapas Auto-Organizáveis, ou SOM, será aplicado neste trabalho [61].

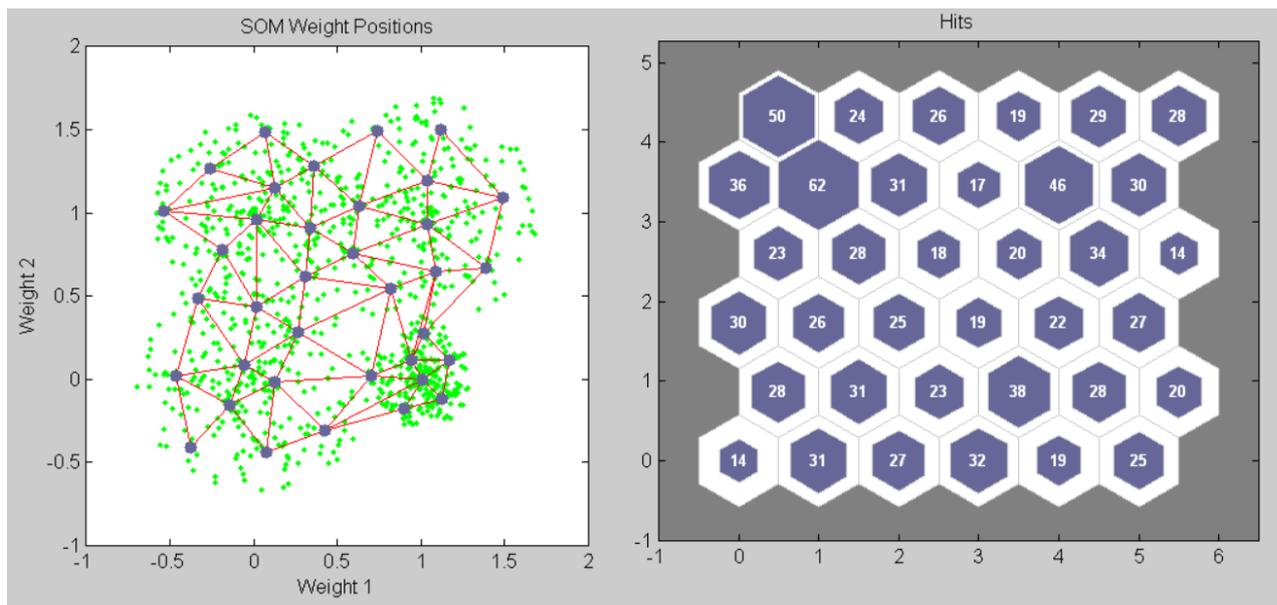


Figura 2.7: Resultado de uma Clusterização Baseada em Modelo [62]

Uma das vantagens deste algoritmo é sua eficiência na visualização de dados. Sua sequência de processos segue de forma semelhante a sequência dos algoritmos *K-means*, onde cada ponto é atribuído ao cluster cujo centro está mais próximo. Entretanto, durante a etapa de atualização do centro de cada cluster, os vizinhos deste ponto central também são atualizados. Em sua finalização, este algoritmo gera uma rede neural, cuja exploração permite a compreensão dos relacionamentos entre os diferentes objetos envolvidos no cluster [26]. Este algoritmo pode ser resumido como sendo a implementação de rede neural do algoritmo *K-means* [61].

Abordagem *Spectral-Based*

A Clusterização Espectral pode ser resumida como uma classe de técnicas que se baseia nos Autovalores de uma matriz de similaridade [55]. Além dessa base para realizar essa divisão, algoritmos desta metodologia são caracterizados por possuírem três estágios principais em suas execuções [55, 26]:

- Pré-processamento: trata da construção da matriz de similaridade para todos os pontos de dados.
- Mapeamento espectral: os pontos de dados são incorporados em um espaço, no qual os *clusters* são mais "óbvios", com o uso dos autovetores do gráfico Laplaciano.
- Pós-processamento: lida com os pontos de dados de agrupamento aplicando um algoritmo clássico de *cluster* para particioná-los.

Algoritmos espectrais possuem muitos atrativos para qualquer analista de dados. São de fácil implementação e produzem rapidamente resultados estatisticamente consistentes [55]. Outras vantagens desse método são que eles não consideram ótimos locais e não realizam suposições fortes sobre a forma do *cluster* resultante.

Essas vantagens tornam aplicações de dados de alta dimensão em alvos adequados para a clusterização espectral, um exemplo que pode ser visualizado na Figura 2.8 como processamento da imagem de uma maçã [25]. Além de alta dimensão, estes algoritmos podem resolver problemas em cenários com formatos de *clusters* muito mais complexos, como espirais entrelaçadas ou outras formas não lineares. Essa liberdade é consequência da ausência de suposições sobre as formas dos *clusters* resultantes [26].

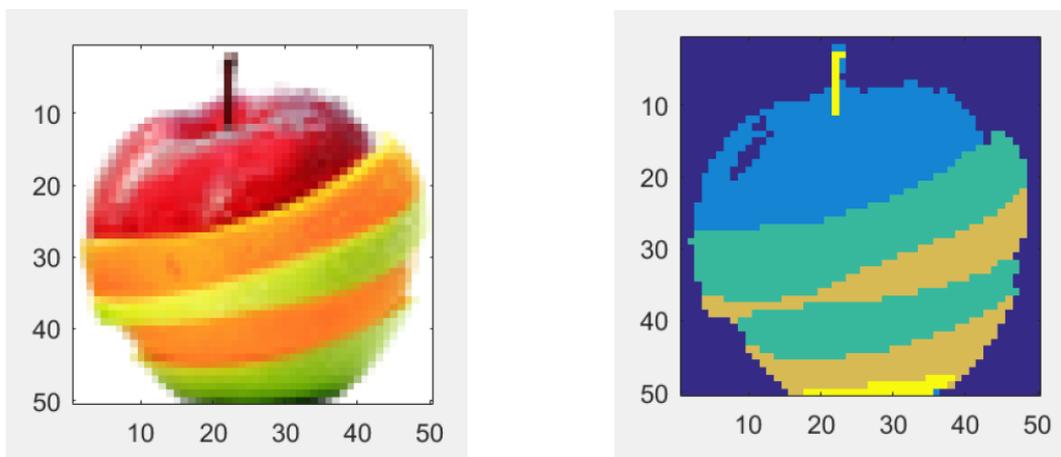


Figura 2.8: Resultado de uma Clusterização *Spectral-Based* [63]

Em contrapartida, a principal desvantagem dessa abordagem é que ela exhibe alta complexidade computacional [55]. Fato que é consequência do cálculo de autovetores, principalmente para uma matriz de grande tamanho [25]. A implementação que será utilizada constrói a matriz de afinidade utilizando uma função do Kernel, como o Kernel Gaussiano da distância euclidiana. O mesmo também requer que o número de clusters seja especificado [54].

Abordagem *Fuzzy Clustering*

A clusterização difusa, ou *Fuzzy Clustering*, possui uma peculiaridade se comparada com os outros métodos discutidos, a saber, o fato de que os objetos de dados podem pertencer a mais de um *cluster* no resultado final. Por este motivo, esses métodos também são referenciados como agrupadores flexíveis [25]. Este tratamento envolve a atribuição, pelos métodos difusos, de graus de associação em vários *clusters* a cada padrão de entrada [57]. Ou seja, para cada objeto de dado, existe um conjunto de coeficientes de associação que

correspondem ao grau de associação a um dos *clusters* do resultado final, caracterizando um agrupamento suave, ou *soft clustering* [51].

Essa propriedade difusa é dada pois cada *cluster* resultante é um conjunto difuso. Um conjunto difuso S é um subconjunto de X que permite que cada objeto em X tenha um grau de associação entre 0 e 1, onde X é um conjunto de dados fornecido, $X = x_1, \dots, x_n$ [25]. Formalmente, um conjunto difuso, S , pode ser modelado como uma função, $FS : X \rightarrow [0, 1]$ [25]. O resultado de uma clusterização difusa é exibido na Figura 2.9, onde é possível observar o grau de associação à um *cluster* representando pelo grau da cor de cada *cluster*.

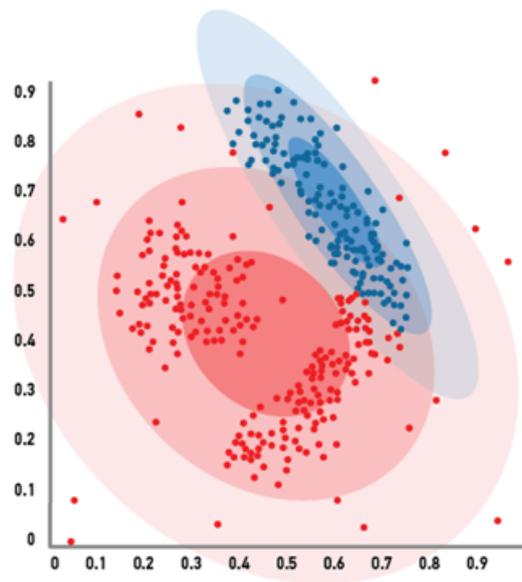


Figura 2.9: Resultado de uma Clusterização *Fuzzy* [64]

Se analisarmos as características dos métodos de clusterização mais clássicos (rígidos), estes que atribuem a apenas um *cluster* exclusivamente, é razoável deduzir que eles pertencem a um caso especial do *cluster* difuso [25]. Este pode ser convertido em um *cluster* rígido, apenas fazendo a atribuição de cada objeto de dado ao *cluster* com que ele possui o maior grau de associação [57]. Entretanto essa conversão não deve ser feita em qualquer caso, pois métodos rígidos não são viáveis em conjuntos de dados complexos onde existem *clusters* sobrepostos, domínio onde a clusterização difusa se destaca [26].

Fuzzy C-means

O algoritmo que representará a metodologia *Fuzzy Clustering* é o *Fuzzy C-means* [61]. A implementação deste algoritmo minimiza a Soma de Quadrados do Erro (SQE), iterativamente seguido pela atualização do peso de associação de cada ponto w à cada *cluster* c , semelhantemente ao algoritmo K-means. Tais interações são repetidas até a convergência dos centros dos agrupamentos [26].

A função Soma de Quadrados do Erro(SQE) para este algoritmo é calculada a partir das seguintes equações [26]:

$$SQE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} x_{wik}^\beta \|x_i - c_k\|^2 \quad (2.4)$$

$$w_{xik} = \frac{1}{\sum_{j=1}^K \left(\frac{x_i - c_k}{x_i - c_j}\right)^{\frac{2}{\beta-1}}} \quad (2.5)$$

$$c_k = \frac{\sum_{x_i \in C_k} w_{xik}^\beta x_i}{\sum_{x_i \in C_k} x_{xik}} \quad (2.6)$$

Nessas equações, x_{xik} é o peso da associação do ponto x_i pertencente a C_k . Esse peso é usado durante a etapa de atualização de C-médias difusas. O centro do cluster (representado por c_k) é calculado ponderando de acordo com os pesos difusos para C_k . Neste algoritmo, a associação de pontos a diferentes clusters pode variar no intervalo de 0 a 1 [26].

2.3.3 Avaliação do Cluster

Em análises de clusterização, não basta simplesmente a aplicação de um algoritmo para alcançar o objetivo pretendido. Por serem uma tarefa de aprendizado não supervisionada, é necessário encontrar uma maneira de validar a qualidade do resultado após a separação em *clusters* [26]. Além da qualidade dos clusters resultantes, uma análise do domínio estudado é necessária para averiguar a viabilidade da análise de *clusters*, pois a clusterização de uma amostra randomicamente ou uniformemente distribuída não fornece informações válidas. Essas validações são reconhecidas como dois dos problemas vitais e essenciais para o sucesso das aplicações de clusterização e, apesar da grande quantidade de pesquisas envolvendo-as, não há solução consistente e conclusiva para esses problemas [26].

As principais tarefas da avaliação de Clusterização incluem o seguinte: Avaliação da tendência de *cluster*, determinação do número de *clusters* e medição da qualidade do *cluster* [25]. Estas que serão detalhadas a seguir.

Avaliando a tendência de *cluster*

Apesar da utilidade da clusterização, sua aplicação não é universal. A aplicação cega de um método de clusterização em um conjunto de dados retornará necessariamente um conjunto de *clusters*; no entanto, o resultado adquirido pode ser inexato. A clusterização requer distribuição não uniforme de dados [25]. Caso contrário qualquer *cluster* formado não fará sentido pois seus limites serão falsos, ou seja, os objetos de dados fora e dentro do limite do *cluster* ainda serão muito parecidos ou próximos.

Para garantir que algoritmos de clusterização possam ter uma chance de produzir resultados significativos, pesquisadores devem medir a probabilidade de que o conjunto de dados tenha sido gerado por uma distribuição uniforme antes de iniciarem seu trabalho. Essa medição é feita mediante testes estatísticos para aleatoriedade espacial [25].

Determinando o número de *clusters*

Outra medida de avaliação do processo de clusterização envolve o cálculo prévio do provável número de *clusters* que serão produzidos. Alguns algoritmos, como *k-means*, requerem o número de *clusters* em um conjunto de dados como parâmetro, tornando a determinação do número de *clusters* essencial para o seu correto funcionamento [25]. Além disso, o número de *clusters* pode ser considerado como uma estatística resumida interessante e importante de um conjunto de dados [25].

Essas considerações evidenciam a utilidade da estimativa desse número antes mesmo que um algoritmo de agrupamento seja usado para derivar *clusters* detalhados. Entretanto, calcular esse número está longe de ser fácil, geralmente devido a ambiguidade em estabelecer o número "certo" [25]. Uma heurística simples é estimar o número de *clusters* como cerca de $\sqrt{n/2}$ para um conjunto de dados de n pontos, na expectativa de que cada *cluster* terá $\sqrt{2n}$ pontos [25].

Medindo a qualidade do cluster

Após a aplicação do algoritmo e antes da interpretação dos resultados é necessário o julgamento da qualidade do resultado produzido, ou seja, queremos avaliar quão bons são os *clusters* resultantes [25]. Uma série de medidas podem ser usadas. Alguns métodos medem o quão bem os *clusters* se ajustam ao conjunto de dados, enquanto outros medem o quão bem os *clusters* correspondem à verdade básica, se essa verdade estiver disponível [25]. Logo, as medidas de validação da qualidade do *cluster* podem ser categorizadas em dois tipos principais: validação de *cluster* externo e validação de *cluster* interno [26]. A principal diferença é se informações externas são usadas para validação de *cluster*. Por outro lado, medidas internas avaliam a qualidade de uma estrutura de *cluster*, sem considerar as informações externas [26].

Métodos extrínsecos, ou externos, fazem a utilização de uma verdade básica sobre o conjuntos de dados, ou outra informação que permita a avaliação dos *clusters* resultantes, para descobrir se o resultado produzido condiz com o esperado ou combina com alguma estrutura externa [26, 25].

No entanto, a existência de informação externa ao conjunto de dados não é uma certeza, principalmente com mecanismos de aprendizado não-supervisionado. Nesses casos, apenas os métodos os métodos intrínsecos, ou internos, podem ser aplicados nos *clusters*.

Estes métodos avaliam a qualidade de um *cluster* considerando o quão bem os *clusters* são separados e o quão compactos estão [25].

O critério de Separação de *clusters* mede o quanto os agrupamentos são distintos entre si. Esse critério pode ser medido utilizando-se a distância entre os centros de *clusters*, distâncias mínimas entre objetos em diferentes *clusters*, assim como a medição da densidade em alguns índices [26].

A Compacidade de um *cluster* mede o grau de relacionamento interno dos objetos em um *cluster*. Existem diversas medidas que fazem a avaliação dessa compactação com base na variação de seus membros, com a menor variação indicando uma melhor compactação [26].

Além dos métodos citados anteriormente, existem técnicas que podem ser usadas para comparar dois conjuntos de resultados de clusterização aplicados no mesmo conjunto de dados [25]. Uma dessas técnicas é o cálculo do coeficiente de silhueta, que usa o valor médio do coeficiente de silhueta de todos os objetos no conjunto de dados para medir a qualidade do mesmo [25]. Esse coeficiente será mais detalhado no Capítulo 6.

Capítulo 3

Revisão Literária

Estudos de *Game Analytics*, MMOs e clusterização não são novidades no âmbito científico. A evolução das plataformas de computação e o aumento do tamanho de dados trabalhados abriram muitas oportunidades para aplicar técnicas de *Analytics* em diversas áreas. Sua aplicação aos jogos foi inicialmente dificultada pela dificuldade em se obter uma grande quantidade de dados do usuário. Entretanto, essa limitação foi superada graças ao desenvolvimento de jogos em rede e pela comunicação por meio da Internet. Além disso, MMOs tiveram um desenvolvimento extraordinário na última década, impulsionados pelos avanços tecnológicos no processamento e comunicações de computadores [6]. Tais superações e desenvolvimentos permitiram a aplicação de técnicas de mineração de dados em um volume imenso de dados provenientes de jogos. Este capítulo relatará trabalhos realizados sobre *Game Analytics*, MMOs e clusterização. Uma seção final deste capítulo mencionará também trabalhos que utilizaram o *Final Fantasy XIV* como objeto de estudo.

3.1 Pesquisas no campo de Game Analytics

Game Analytics não é um assunto novo, isto é, pesquisas explorando suas vertentes já são realizadas desde, no mínimo, 2004, sendo que mais de trinta artigos foram publicados nesse período [6]. Além do interesse constante manifestado pela publicação de ao menos 2 artigos por ano [6], esse campo de pesquisa apresenta uma referência chave que permite que cientistas a utilizem como ponto de partida para seus trabalhos. Com cerca de trezentas e setenta citações no momento da escrita desta monografia, o livro *Game Analytics* [5], dos autores Anders Drachen, Magy Seif El-Nasr e Alessandro Canossa, é a publicação com maior relevância e citações no Google Scholar ao se pesquisar pelos termos *Game Analytics* [65].

Contido no início do livro [5] estão noções básicas de *Game Analytics*, incluindo sua definição, terminologias, seus benefícios e casos de estudos. Em seguida, o livro relata as coleções de telemetria de jogos, assim como exibição de ferramentas para realização desta etapa do processo. A terceira parte do livro foca em técnicas e diretrizes para realização da mineração de dados, contendo também um caso de estudo e uma entrevista com membro de uma desenvolvedora de jogos. Meios para visualização das métricas utilizadas e métodos para avaliação de jogos são relatados e explorados nas partes seguintes, respectivamente. A sexta parte do livro relata sobre *Analytics* no contexto de comunidades de jogadores. A parte seguinte visa fornecer uma melhor compreensão sobre métricas utilizadas no processo de aprendizagem em jogos com tal objetivo. Finalizando o livro, a oitava parte entra em detalhes sobre métricas para melhorar desafios analisando um estudo de caso sobre o *level design* automatizado realizado em um jogo de *solitaire*. Esta dissertação baseia sua fundamentação teórica principalmente nos conceitos e diretrizes constituídos pela publicação de Anders et al.

Dentre as diversas vertentes contidas em *Game Analytics*, muitos pesquisadores escolhem explorar o indivíduo que se entretém com jogos digitais, empenhando-se em investigar diferentes aspectos do jogador e do ato de jogar. Exemplos de trabalhos desta vertente incluem uma investigação sobre a força motivacional intrínseca dos jogos, assim como quais efeitos são causados na saúde do jogador [66]. Tal estudo foi realizado a partir da análise das respostas de um questionário realizado com os jogadores. Outro estudo analisou como e porquê um indivíduo se identifica como *Gamer* por meio da utilização de questionário como recurso [67].

O uso de questionários é um recurso recorrente em pesquisas de *Game Analytics* para coletar dados sobre jogadores. Com o intuito de avaliar a utilidade e conforto de controles, foi realizada uma pesquisa sobre o efeito do controle na experiência do usuário [68]. Enquanto este trabalho analisou os impactos na experiência do jogador, outros estudos analisaram a influência do gênero do jogo no estado do jogador [69].

Uma utilidade adicional de questionários é o auxílio a pesquisadores que buscam criar escalas e modelos para classificar e medir aspectos do jogador. Um exemplo desta prática, inclui a proposta de uma escala para medir o comportamento social em jogos de time, tal estudo explorou jogadores do jogo *League of Legends*, no qual dois times de cinco de enfrentam em uma arena, e todos podem se comunicar entre si [70]. Outros dois exemplos abrangem os trabalhos que buscaram compreender e explorar a motivação e satisfação de jogadores, com propósito de medir ambas [71, 72].

Com o aumento de pesquisas nessa vertente, muitas definições e propostas de classificação de jogadores emergiram. Com intuito de reunir esse conhecimento em apenas uma produção e relatar as diferentes classificações de jogadores presentes na literatura

científica, foi realizado um estudo exploratório [73].

Entretanto, questionários não são as únicas fontes de dados para pesquisas nesse ramo. Os pesquisadores já utilizaram avaliações de jogos comerciais para legitimar modelos de avaliação do aproveitamento do jogador [74]. Outra pesquisa envolvendo avaliações de jogos investigou que aspectos impactam os *reviews* em jogos presentes na plataforma de venda de jogos, STEAM [75]. Este artigo utilizou técnicas de Aprendizado de Máquina, Rede Neural Artificial e Árvore de Regressão para analisar essas avaliações e observou a acurácia dos resultados para avaliar a resolução atingida. Dados sobre mecânicas e características também podem ser usados para analisar diferentes partes do processo de desenvolvimento de jogos. Um exemplo do uso desses dados é um estudo realizado para explorar a evolução da mecânica de salvar dados em jogos, desde a permanência do *High Score* [32].

Apesar dos inúmeros guias, diretrizes e pesquisas realizadas, o campo de pesquisa de *Game Analytics* ainda não atingiu a maturidade [6]. Sua natureza inter e multidisciplinar abre um amplo espectro de oportunidades de pesquisa que ainda não foram exploradas [6]. Essa diversidade fica evidente ao examinar os diversos trabalhos relatados nesta seção.

3.2 Pesquisas de *Game Analytics* sobre MMOs

O gênero MMO atraiu a atenção do mundo quando o *World of Warcraft* [21] foi lançado em 2004 e obteve um sucesso de engajamento nunca antes observado. Entretanto, antes mesmo do gigantesco sucesso que obteve, outros MMOs já existiam e pesquisas sobre eles já eram realizadas como, por exemplo, com o *Everquest* [76, 77] e *Lineage* [35].

O sucesso e interesse geral atribuído ao gênero MMO gerou um reflexo no cenário científico. Um reflexo que se estende a diversos campos de pesquisa. A investigação sobre o indivíduo alvo de tais jogos, o jogador, foi objeto de diversos trabalhos. Foram determinadas suas características estruturais [78]. Descrições da etnografia cognitiva em andamento de um MMO em expansão foram realizadas [35]. Métodos para categorizá-los automaticamente foram propostos [79], sendo que tentativas envolvendo as técnicas de clusterização serão relacionadas na Seção 3.5. Possíveis efeitos relacionados com a saúde do jogador foram o alvo de algumas pesquisas, sob o receio de efeitos negativos [80]. Uma das pesquisas dessa área, examinou a racialização, praticada por jogadores, de trabalhadores chineses que recebiam compensação monetária pela mineração de recursos no jogo de RPG multi-jogador *online World of Warcraft* [81]. Outro trabalho examinou a existência de um estereótipo de gênero em relação a classes do jogo entre jogadores iniciantes e experientes [82].

Além da quantidade massiva de jogadores, a possibilidade de interação entre os jogadores fornece oportunidades únicas de pesquisa. A dinâmica de amigos e inimigos ainda existe dentro desses ambientes digitais e as estruturas de como ela são executadas foi um objeto de estudo [83]. Outras investigações procuraram examinar estruturas de grupos virtuais e compará-las com grupos especiais sociais reais [84] e analisar seus impactos nas realizações individuais do jogador por intermédio de resoluções colaborativa de problemas [34]. As interações sociais também foram alvos de pesquisas [77], sendo que uma delas verificou como os aspectos sociais do jogo imitam o mundo real em termos de escolhas que os jogadores fazem ao interagir com os outros [85].

Independentemente do atrativo de se estudar uma quantidade massiva de jogadores, muitas pesquisas realizaram suas investigações com foco no jogo, suas mecânicas e seu processo de desenvolvimento. O interesse em avaliar aspectos técnicos foi identificado durante a pesquisa realizada, cujo trabalho investigou o impacto de falhas na rede na Qualidade de Serviço percebida [86]. Também foi conduzido um estudo sobre como é gerado o tráfego de rede e até que ponto depende da ação que o jogador executa e do contexto virtual [87]. Uma proposta caracterizada pela intenção de auxiliar desenvolvedores em aspectos técnicos, envolve um *middleware* de monitoramento não intrusivo que lida com esses desafios particionando dinamicamente o mapa virtual geográfico em termos de objetos do mapa e alterações de estado do jogo [88].

Além das iniciativas mencionadas anteriormente, diversos estudos foram feitos tendo o jogo ou mecânicas, como ponto focal. Cientistas relataram o primeiro estudo a longo prazo de um MMORPG popular desde seu lançamento [89]. Dois trabalhos analisaram o processo de progressão do personagem, tanto no processo de *design* [90], quanto avaliando a adição da terceira dimensão à sua árvore típica de progressão do jogo [91]. Outro estudo se atentou ao equilíbrio de elementos de perigo, principalmente com recursos de segurança durante negociações de transações de jogadores [92]. Baseando-se na teoria de que jogos virtuais podem ser espaços onde comportamentos humanos serão espelhados, cientistas projetaram um sistema de gerenciamento e disseminação de doenças que monitora interações e aciona eventos de infecção, usando personagens não-jogadores como *proxies* humanas, para espalhar uma doença entre uma população virtual, estudando se sua taxa de contaminação se assemelha com padrões reais [93].

Diferentemente dos trabalhos relacionados anteriormente, uma abordagem possível para coleta de dados envolve o desenvolvedor do jogo. Dois estudos investigaram como os profissionais que desenvolvem MMOs raciocinam sobre os requisitos de jogabilidade e como estes são tratados em seus projetos [94, 95].

Outra vertente explorada pelos pesquisadores desta área é a relação com a área educacional. Uma proposta já foi feita para o desenvolvimento de jogos educacionais multi-

jogador *online*, com base na teoria da atividade [96]. Tais intenções de aplicação no cenário pedagógico são reforçadas por análises de como a estrutura em MMORPGs pode informar o *design* de ambientes de aprendizado interativo e de aprendizado baseado em jogos [97, 98]. Escolhendo usar tais diretrizes no ensino de uma matéria em particular, um grupo de cientistas propôs uma estrutura que guiará a incorporação de mecanismos de análise da aprendizagem na educação em programação de computadores [99] e depois examinou os efeitos do uso de um MMORPG educacional para múltiplos jogadores online no ensino e aprendizagem de programação de computadores [100].

Uma inferência importante da pesquisa realizada até agora [6] é a predominância do *World of Warcraft* como jogo focado em diversos trabalhos [13, 86, 101, 102, 103, 23, 104, 81, 18, 105, 78, 106, 90, 82, 87, 107, 21]. Cientistas afirmam que essa escolha é justificada pela popularidade expressiva do jogo, assim como pela disponibilidade e dimensão de seus dados de personagens e jogadores a qualquer um com interesse.

3.3 Análises com foco no Jogador

Evidenciado na seção anterior, trabalhos realizados com MMOs possuem uma diversidade de finalidades, entretanto, nota-se que existem alguns propósitos que são mais recorrentes. A intenção desta análise é compartilhada com muitas pesquisas [6] e tem por objetivo final a compreensão dos jogadores.

Essa compreensão do jogador pode ser atingida mediante uma definição de perfis de jogadores. Um trabalho em particular explorou traços de personalidade de jogadores utilizando o *Big Five Inventory*, totalizando quase vinte mil jogadores em sete gêneros de jogos [108], enquanto outra pesquisa explorou o estado de arte de trabalhos que relacionaram os perfis de jogadores com as emoções provocadas por ações no jogo [109].

Uma abordagem que está sendo averiguada por alguns estudos é o efeito de jogos na saúde do jogador. A identificação de jogadores viciados ou com uma tendência ao vício foi explorada em estudos que visaram realizar tal identificação por meio da auto-identificação com o avatar do jogador [110], estudo da relação entre o vício, agressividade, autocontrole e traços de personalidade narcísicos [111], e uma proposta de medição de hábitos e preferências individuais no uso de videogames [112].

Outras pesquisas escolhem investigar aspectos específicos de jogadores, como a utilização do visual do personagem como indicador de performance usada pelos jogadores [103]. Alguns jogadores transformaram sua experiência em uma atividade remunerada. O jogador recebe compensação financeira pelo esforço realizado dentro de um jogo. A análise desse tipo de jogador foi um dos focos de pesquisas realizadas neste meio [113]. Entretanto, nem todas as interações de troca envolvendo valores monetários são bem vistas pelos

jogadores, sendo algumas até consideradas passíveis de punição pela desenvolvedora. Um tipo de transação foi foco de uma pesquisa que visou identificar jogadores que realizam vendas assim [114]. Seja para engajar em atividades remuneradas ou apenas para melhorar o próprio personagem, alguns jogadores criam *bots*, ou *scripts* que jogam o jogo automaticamente de acordo com uma série de comandos, para contornar o processo normal e mais custoso que ter que realizar tais interações manualmente. Focando em jogadores com tal intenção, pesquisas foram realizadas para detectar quais personagens não estariam sendo jogados por pessoas reais [105].

Dois importantes aspectos de jogadores que tanto pesquisadores quanto desenvolvedores possuem um grande interesse em avaliar são a vontade de jogar e o envolvimento do jogador. O primeiro aspecto possui pesquisas que visaram identificar os elementos psico-estruturais de jogos [115], enquanto o segundo possui pesquisas discutindo sobre como os jogadores podem controlar a interpretação da influência do personagem, facilitando assim a sensação de envolvimento dentro de um jogo [116]. Quando ambos aspectos de vontade e envolvimento estão baixos, o abandono do jogo é uma consequência direta, logo, pesquisas foram feitas para identificar quais fatores afetam negativamente ambos, como, por exemplo, se o senso de justiça dentro do jogo afeta esse abandono [117]. A diversão do jogador pode ser um agente chave no envolvimento do mesmo com o jogo, assim sendo, uma pesquisa explorou a relação entre os atributos da transação de itens virtuais, a diversão percebida pelos usuários e o custo da transação [118].

Além da exploração de perfis de jogadores e aspectos que os afetam, muitas pesquisas buscaram analisar o fenômeno social que ocorre quando o jogo permite múltiplos jogadores. Semelhante as outras pesquisas que buscaram verificar se o cenário virtual pode espelhar comportamentos do mundo real, um trabalho foi desenvolvido para explorar as interações entre jogadores e descobrir se configura como tribalismo [102]. Estudos buscaram propor modelos e sistemas para investigar comportamentos individuais dentro de organização virtual [119] ou interações com amigos [120]. Proposto na literatura, um método estatístico, automatizado e escalonável, objetiva prever essas reuniões inesperadas de usuários [121]. Semelhante a este, uma abordagem para rastrear grupos ao longo do tempo, à medida que participam de redes maiores de grupos, foi estudada em outra pesquisa [122]. Diferentemente desses, um dos trabalhos desse campo propôs a classificação de jogadores com base na distribuição de graus dos rede de relacionamento [107]. Outro aspecto social pesquisado foi o comportamento de jogadores com maiores níveis de experiência em um MMORPG [38].

Um aspecto diferente que foi analisado por algumas pesquisas, e mencionado na seção anterior, é a associação dos jogos com a educação. A relação com ensino e aprendizagem de matemática em um curso à distância foi estudada a partir de um jogo educacional de

RPG online [123]. Enquanto o artigo anterior explorou cursos a distância, um trabalho realizado posteriormente focou na interação entre aluno e professor por intermédio de um computador [124], analisando a imersão do aluno e os efeitos da substituição do professor pelo computador.

As pesquisas envolvendo o jogador são abundantes nesse campo de pesquisa evidenciando o interesse da comunidade científica no assunto. Entretanto não foi descoberto um concordância no melhor método para analisá-los, nem quais dados são obrigatórios.

3.4 Aquisição de dados através de *Game Log*

Para conseguir realizar uma pesquisa em *Game Analytics*, o pesquisador deve ter acesso a algum tipo de dado que possa ser vinculado ao jogo de algum meio. Os dados em questão, úteis para análise de RPGs, costumam ser a progressão do personagem, missões completas, interações sociais, tempo de jogo, habilidades do personagem, estatísticas de combate, progressão de história, itens, desempenho de habilidades, dano recebido e infringido [5].

Uma das fontes de dados mais utilizada [6] neste campo de pesquisa é o *Log* do Jogo. Esse *Log* pode ser disponibilizado pela desenvolvedora ou mecanismos de captura podem ser desenvolvidos para gerar o próprio *Log*. *World of Warcraft* é famoso por fornecer uma quantidade expressiva de dados com a dimensão necessária para quase qualquer análise.

Dentro desses *Logs* podem existir uma gama variada de informações. Alguns trabalhos utilizaram as mensagens armazenadas nesses *Logs*, estes que foram trocadas entre jogadores para investigarem tais interações sociais [34, 79, 124, 83, 119, 69]. Dentro dessa gama, também existem variáveis que relacionam o jogador com itens do jogo e com o comércio, pesquisas envolvendo essas variáveis tendem a estudar a economia virtual e seus desafios [18, 114].

Dentre tais variáveis pertencentes a *Logs* de jogo, as mais estudadas são aquelas que revelam as estatísticas do personagem. Alguns pesquisadores escolhem poucas variáveis dentro das disponíveis, como um título de Mentor [22] ou variáveis de amizade, login e grupos de jogadores [120].

Muitos trabalhos utilizaram os *Logs* do jogo para fazer uma análise principalmente utilizando o *level* do personagem, seu tempo de jogo, ações, seu equipamentos e atributos [13, 125, 14, 126, 104, 105, 106]. Muitas dessas análises só foram possíveis pela abundância de variáveis disponíveis dentro do *Logs* e por conta da permissão da desenvolvedora. Algo que nem sempre é possível, como no caso do *Final Fantasy XIV*, cujos dados possuem muitas limitações de atributos. Por fim, existem aqueles que visam o jogo como um todo, decidindo apenas trabalhar com ações de grupos de jogadores, como quantos existem, quantos estão jogando e outros [89].

3.5 Aplicação de Clusterização

Um dos pontos focais deste trabalho consiste na comparação de metodologias de clusterização, entretanto a aplicação de tais técnicas já vem sendo realizada desde no mínimo onze anos atrás [24]. Nesta seção serão identificadas quais técnicas e algoritmos são usados e inferir lacunas e tendências na literatura científica.

Ao finalizar a pesquisa por artigos que realizam a clusterização em seus dados de jogos, a primeira conclusão tirada foi que o algoritmo *K-means* é o algoritmo de clusterização mais utilizado. Um exemplo de tal aplicação envolve uma pesquisa que propôs o uso de realidade virtual para visualização do resultado de uma análise de dados, esta sendo a clusterização dos dados [8]. Outro uso deste algoritmo envolve a análise da efetividade de uma campanha de marketing em jogos *mobile free-to-play* [9]. Entre os diferentes usos da clusterização, um dos mais comuns em *Game Analytics* é a identificação de perfis de jogador. O propósito desse tipo de aplicação varia entre pesquisas. Por exemplo, em [11], os autores investigaram o agrupamento dos perfis de jogadores com base em algum comportamento em específico, como padrões de compra. Outra pesquisa quis analisar qual técnica seria apropriada para identificar esse perfis, comparando modelos de regressão com a clusterização *K-means*, finalizando com a avaliação do resultados utilizando o Coeficiente da Silhueta [10].

Ainda que uma pesquisa tenha aplicado o algoritmo *K-means*, nada impede que sejam aplicados outros algoritmos, tanto para validar o resultado quanto para realizar outras análises. Para agrupar jogadores em relação a seus comportamentos, uma das pesquisas sobre o tema utilizou *Simplex Volume Maximization* para determinar os perfis extremos presentes no conjunto de dados e complementou o resultado com os clusters formados pelo *K-means* [12]. Outra pesquisa aplicou, sobre o resultado da clusterização, modelos de regressão para identificar perfis que no futuro fariam pagamentos da mensalidade em *World of Warcraft* [13]. Com o objetivo de melhorar o sistema de pareamento de partidas de jogador contra jogador no jogo *Destiny*, uma pesquisa utilizou tanto *Archtypal Analysis* quanto *K-means* para agrupar os perfis semelhantes de habilidades de jogador [14].

Dois trabalhos realizaram suas investigações aplicando duas técnicas de clusterização. Ambas possuíam o intuito de melhorar o cenário de clusterização de dados comportamentais de jogadores. Uma dessas pesquisas aplicou uma abordagem hierárquica aglomerativa [15] e a outra aplicou o algoritmo DBSCAN da abordagem baseada em densidade [16].

Apesar da quantidade de artigos que aplicaram o *K-means*, ela não engloba todos artigos estudados. Dentre os dezoito artigos agrupados nesta seção, oito trabalhos não aplicaram agrupamentos particionais. Dois trabalhos realizaram seus agrupamentos utilizando a abordagem baseado em modelos, utilizando o algoritmo de mapeamento auto-

organizado (SOM) para conseguir agrupar comunidades online dentro do jogos *World of Warcraft* [23] e *Ragnarok* [24].

Sendo uma das três abordagens mais utilizadas em trabalhos desta seção, a clusterização hierárquica foi aplicada em quatro artigos. Desses quatro, um desses estudos aplicou o algoritmo CURE, como proposta para modelar jogadores [19]. Os outros três estudos aplicaram algoritmos aglomerativos com diferentes objetivos [15, 17, 18]. Um deles aplicou essa clusterização para analisar as decisões de compra de usuários e permitir uma avaliação do mercado virtual [18]. Outra pesquisa buscou descobrir padrões de ações de jogadores em um jogo *Free-to-play* [17].

Abordagens baseadas em densidades foram utilizadas em quatro pesquisas, sendo uma delas já discutida anteriormente [16]. Procurando agrupar os caminhos percorridos por jogadores, um dos trabalhos mencionados aplicou dois algoritmos diferentes, *Average Distance of Corresponding Points-Density Clustering* e *Longest Common Subsequence Density Clustering*, validando seus resultados por meio do coeficiente de silhueta [20]. Os outros dois trabalhos utilizaram o algoritmo de maximização de expectativa para determinar perfis de jogador com base em conquistas virtuais [21] e estudar o fenômeno de mentoria em *Everquest 2* [22], respectivamente.

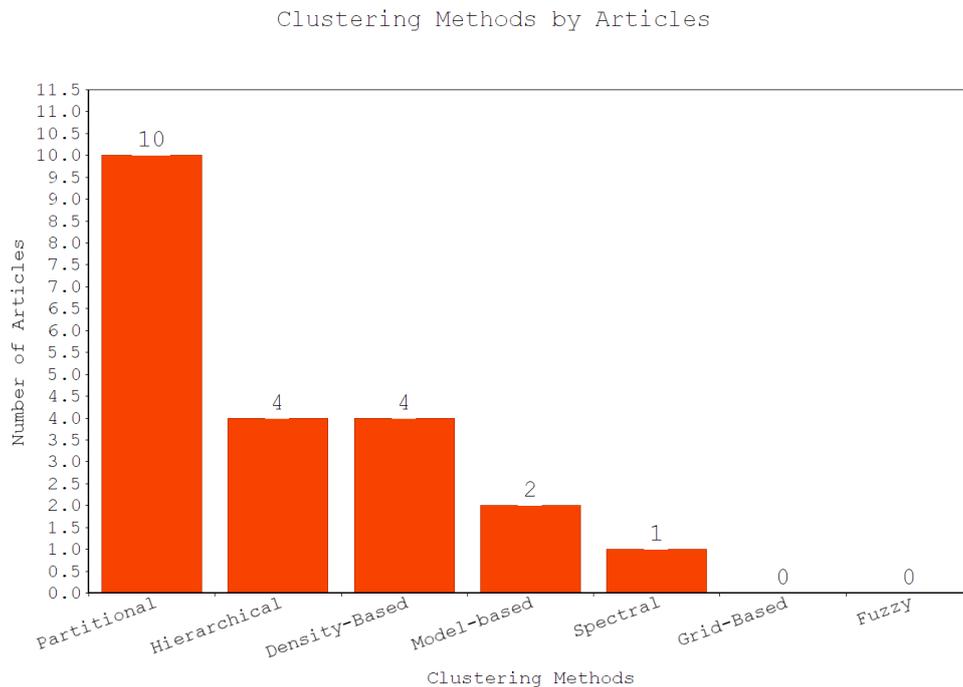


Figura 3.1: Abordagens de Clusterização por Artigos

Esta sumarização de trabalhos com clusterização em *Game Analytics* permitiu a criação do gráfico presente na Figura 3.1, onde está registrado o número de artigos que aplicou determinada abordagem de clusterização.

Ao analisar a Figura 3.1, determinadas conclusões foram retidas. Primeiramente, foi revelado uma tendência ao algoritmo *K-means*, evidenciando-o como um algoritmo base de clusterização e primeira escolha ao se aplicar a abordagem particional. Em segundo lugar e com menos de metade de artigos onde são aplicados, estão as abordagens hierárquica e baseada em densidade. Dentre as abordagens hierárquicas, foi notado uma preferência por abordagens aglomerativas. Algoritmos baseados em densidade possuem uma maior variedade entre si, com quatro algoritmos usados em quatro artigos.

Abordagens baseadas em modelos e spectral foram pouco exploradas, com dois artigos aplicando a clusterização de modelos e apenas um artigo explorando a abordagem spectral, sendo este um artigo que faz um estudo sobre clusterização em dados de jogos [7]. Finalizando, não foram encontrados artigos que aplicaram as metodologias baseadas em grade e difusas para analisar dados de jogos.

3.6 Pesquisas com o Final Fantasy XIV

Ao analisar a bibliografia científica relacionada ao jogo *Final Fantasy XIV* foram encontrados poucos artigos com este tema em foco. Uma dessas pesquisas é um trabalho de graduação que analisa as estratégias utilizadas pela equipe de *Final Fantasy XIV* para tradução e localização do jogo fora do Japão [127]. O objetivo geral era compreender parte do processo de desenvolvimento do jogo. O estudo analisou o texto original e traduzido de falas de personagens não-jogáveis do jogo, assim como seus perfis de personagem.

Outro trabalho se concentra nos aspectos sociológicos do *Final Fantasy XIV*, escolhendo explorar a encenação performativa da identidade do jogador utilizando seu avatar virtual. O pesquisador teve a intenção de compreender como o jogo possibilita essa encenação e personificação do avatar, assim como as interações do jogo complementam tais ideais [128]. O estudo foi realizado utilizando respostas de um questionário à jogadores sobre o tema, e de observações feitas pela autora ao jogar o jogo.

O último trabalho que será listado aqui realizou uma entrevista com 905 jogadores para investigar a relação entre a percepção do jogador (dimensões de realizações dentro do jogo, imersão e orientação social) e os aspectos do avatar (social, econômico, cultural e simbólico). Este estudo compilou as respostas e as analisou mediante uma técnica de modelagem de equações estruturais baseada em componentes. Os pesquisadores concluíram que os aspectos do avatar podem ser previstos analisando o percepção do jogador. Por exemplo, os resultados do trabalho mostram que o capital cultural do avatar é previsto por orientação de imersão (personalização, descoberta e interpretação de papéis), orientação de realização (mecânica) e orientação social (relacionamento e socialização).

Para validação de seus resultados, os pesquisadores aplicaram técnicas de variância média extraída e confiabilidade composta [129].

Não foi encontrado nenhum trabalho na literatura que tenha utilizado a base de dados disponível por intermédio de uma API do *Final Fantasy XIV*. Esse fato em conjunto com os poucos trabalhos realizados em cima de *Final Fantasy XIV* evidenciam uma notável oportunidade de pesquisa: uma quantidade imensa de dados e praticamente sem análise investigativa.

Capítulo 4

Final Fantasy XIV

A fonte de dados, onde serão aplicados os algoritmos de clusterização, foi originada do jogo *Final Fantasy XIV*. O conhecimento sobre todos os aspectos do jogo é essencial para compreensão desta pesquisa. Inicialmente será feita uma descrição da história do desenvolvimento do jogo, assim como sua evolução como produto comercial. Aspectos do jogo, como enredo, *gameplay* e atividades serão especificadas neste capítulo posteriormente.

4.1 Trajetória do Jogo

A franquia *Final Fantasy* contribuiu muito para a popularização do gênero RPG japonês no comércio ocidental, conquistando um lugar entre as séries de jogo mais vendidas de todos os tempos, com mais de 142 milhões de jogos vendidos. Entretanto, a série também explorou outras mídias, como quadrinhos, livros e até filmes. A série é melhor conhecida por suas composições musicais e elementos narrativos que exploram os conflitos internos de seus protagonistas [130]. Apesar de focar em RPG, a série também divergiu para diferentes gêneros, como luta, jogos de ritmo, corrida, mini-games e até jogo de tiro. A série iniciou em 1987, com *Final Fantasy I* e hoje possui 15 jogos em sua linha principal e numerada.

Dessa linha principal, apenas dois jogos pertencem ao gênero MMO, *Final Fantasy XI* e *Final Fantasy XIV*. O décimo primeiro jogo numerado foi lançado em 2002 e foi o primeiro MMORPG a possuir o mesmo mundo integrado com jogadores de PC e *Playstation 2*. Quando uma versão do jogo foi lançada para *Xbox 360*, ele foi o primeiro MMORPG de console. O jogo possuía um sistema de mensalidades e expansões, onde o jogador precisa pagar para jogar no mês e a compra das expansões é exigida para ter acesso ao novo conteúdo contido nela. A recepção deste jogo foi altamente positiva, além do sucesso financeira dessa versão do *Final Fantasy*, tornado-se o jogo mais lucrativo da história da empresa em 2012.

Em Setembro de 2010, foi lançado o segundo *Final Fantasy* do gênero MMORPG, *Final Fantasy XIV*. Entretanto, ao contrário de seu predecessor, o jogo não obteve sucesso com sua recepção inicial. Foi altamente criticado pelo seu *gameplay*, com alguns críticos afirmando que o jogo não parecia finalizado. A reação do público foi tamanha, que a empresa *Square Enix* suspendeu a cobrança de mensalidades, pospôs a versão de *Playstation 3*, substituiu a liderança do desenvolvimento e pediu desculpas pelo produto.

Liderada por Naoki Yoshida, a equipe se propôs a continuar a desenvolver conteúdo para o jogo original, assim como iniciar o desenvolvimento de um novo jogo, que iria possuir novos mecanismos de jogo, estruturas de servidor melhoradas, melhor interface e histórias renovadas, para suprir as deficiências da primeira versão. Após um longo período de desenvolvimento paralelo, aproximadamente dois anos, os servidores da versão inicial foram fechados em Novembro de 2012, evento que fez com que jogadores experimentassem a destruição do mundo virtual por um dragão divino. Em Agosto de 2013, *Final Fantasy XIV: A Realm Reborn* foi lançado para uma recepção fortemente positiva, superando o estigma criado pelo seu predecessor e resgatando o prestígio do nome da franquia.

Após essa saga de erros, reconhecimento, cobrança e entrega, o *Final Fantasy XIV* reconquistou a confiança de seus fãs, assim como atraiu novos jogadores. Entretanto, como a falta de conteúdo pode significar a morte de um MMORPG, a *Square Enix* manteve o desenvolvimento de novos conteúdos para seu principal jogo desse gênero. A qualidade desse conteúdo deve agradar os jogadores, incentivando-os a manter a sua assinatura.

Após cinco atualizações com conteúdo inédito a cada três meses aproximadamente, foi anunciada a primeira expansão do jogo, *Heavensward*, em 2015. Seguindo o modelo do jogo base, foram lançadas cinco grandes atualizações antes do lançamento da segunda expansão, *Stormblood*, em 2017. Esse ciclo de lançamento de atualizações e expansões está persistindo devido as respostas positivas que a equipe tem recebido na forma de um consistente crescimento de personagens no jogo. Crescimento que em 2018 atingiu treze milhões de personagens registrados. Em 2018 foi anunciado que a terceira expansão, *Shadowbringers*. Esta que foi lançada em 2019, dois anos após a segunda expansão.

4.2 Enredo

A história do jogo *Final Fantasy XIV: A Realm Reborn* se passa após uma calamidade que destruiu o mundo do *Final Fantasy XIV* original. Após tal evento, o personagem do jogador, denominado *Warrior of Light*, é chamado à um novo mundo por uma deusa, *Hydaelyn*. Neste novo mundo, tal calamidade foi evitada, entretanto, as principais cidades: *Ul'dah*, *Gridania* e *Limsa Lominsa*, enfrentam graves problemas com tribos bestiais que ameaçam invocar seus deuses, conhecidos como *Primals*, para combater a ameaça que

sentem dessas cidades. Um grupo chamado de *Scions of the Seventh Dawn* recruta o personagem principal para auxiliar à combater esses deuses.

Por outro lado, uma outra entidade representa a maior ameaça às alianças da região de *Eorzea*, na forma do Império de *Garlemald*. Este império pretende acabar com qualquer possibilidade de invocação de *Primals* extinguindo as tribos bestiais e conquistando toda região sob sua bandeira. Para combater tal ameaça, as três alianças se unem com os *Scions* para defender a região. Durante os combates entre essas facções rivais, um novo inimigo é descoberto na forma de entidades extra dimensionais denominadas *Ascians*. Essas entidades possuíram um membro dos *Scions* e manipularam o império para provocar as invocações de *Primals*.

No fim da história do jogo base, o personagem consegue vencer a investida do império. Entretanto, no final da quinta atualização do jogo base, acontece um Golpe de Estado na cidade de *Ul'dahe* o personagem é traído por membros de um exército privado dos *Scions*. Exilado da região, o personagem é obrigado à fugir para uma nação vizinha que estava fechada para estrangeiros. Nesta nova região começa a primeira expansão, *Heavensward*.

Essa nação, *Ishgard*, possui uma longa história de conflitos com os dragões da região. A trama desta expansão direciona o personagem para ambos lados deste conflito, descobrindo novamente ações do inimigo recorrente, os *Ascians*. Mas a vitória sobre os *Ascians* não acabou com os problemas de *Ishgard*. Nas atualizações posteriores ao lançamento da expansão, o dragão primordial *Nidhogg* manteve sua declaração de guerra contra os habitantes de *Ishgard*, mesmo depois de tentativas de reconciliação.

Após a resolução dos conflitos de *Ishgard* e uma mudança de líder, a nação abriu suas portas para nações vizinhas e entrou para a aliança de *Eorzea*. Junto com tais mudanças, a reputação do personagem do jogador foi restaurada e ele foi permitido voltar para as nações do jogo base. Porém, novos conflitos estavam surgindo entre o povo de uma nação dominada pelo império, tal nação é conhecida como *Ala Mhigo*.

Este é o início da segunda expansão, *Stormblood*. Para libertar *Ala Mhigo*, o personagem viaja para uma nação distante e também no comando do império, *Doma*. Nesta região, o personagem contribui para uma rebelião contra o império para libertação da nação. Uma vez sucedidos, o personagem reúne seus aliados no confronto em *Ala Mhigo* para enfrentar o filho do imperador, que estava comandando a unidade residente daquela região.

As atualizações que vieram após o lançamento de *Stormblood*, exploram o destino de cada nação após a resolução do conflito com o império. Outro lado explorado inclui conflitos internos do próprio império; um lado com ideais pacíficos e outro com relações com os *Ascians*. Finalmente, essa história levará à terceira expansão, *Shadowbringers*, onde o foco da história envolverá a origem do Império de *Garlemald*.

Este foi um breve resumo da trama principal do jogo até o ponto de escrita deste trabalho. Existe uma grande quantidade de pequenas narrativas contidas no mundo, assim como sequências de missões com suas próprias histórias. Entretanto para o escopo deste trabalho, não serão exploradas todas narrativas, caracterizando esse resumo como suficiente para compreensão do trabalho, assim como adequado para o entendimento do gênero abordado na história do mundo.

4.3 Criação de Personagem

Final Fantasy XIV possui uma extensa e variada gama de opções para customização do personagem com o qual o jogador irá jogar. Essas variações permitem que o jogador se identifique melhor com seu avatar virtual e se sinta único no mundo com milhões de personagens. Essa identificação facilita a imersão no jogo e como consequência tende a aumentar o tempo dedicado ao jogo, caso o jogador se identifique com seu personagem.

Ao iniciar o processo de criação de personagem, o jogador pode escolher qual vestimenta o personagem irá usar durante a criação, com as opções: roupas clássicas da raça, armadura da classe e apenas com a roupa de baixo. Em seguida o jogador poderá escolher sua raça, clã e gênero. Gênero está dividido entre Masculino e Feminino. Clã compõem pequenas diferenças dentro da raça com cada uma tendo duas variações, como um exemplo, a raça *Hyr*, possui os clãs *Midlanders*, com alturas e composições físicas médias, e *Highlanders*, com maiores alturas e corpos mais musculosos. Essa raça é equivalente ao Humano, sendo a raça mais populosa do mundo de *Final Fantasy XIV* e regularmente usada como avatar do personagem principal nos materiais promocionais.

Semelhantemente à outros MMORPGs, existe uma raça equivalente à mitologia dos elfos, seres orgulhosos, altos e com orelhas pontudas, denominados *Elezen*. No mundo de *Eorzea* existem os clãs *Wildwood*, habitantes das cidades como *Gridania*, identificados por uma pele um pouco mais morena que o segundo clã. Este sendo os *Duskwight*, *Elezens* que fizeram de cavernas e ruínas suas casas e por consequência possuem uma coloração clara de pele.

Lalafells podem ser comparados com os Anões em mitologias fantasiosas, entretanto somente em tamanho. Os dois clãs possuem composições físicas bem similares, mas os habitantes do litoral *Plainsfolk* possuem orelhas mais pontudas que seus semelhantes, os *Dunesfolk*, que habitam os desertos. Sua baixa estatura e olhos grandes são relacionados à feições mais jovens.

A raça dos homens-gato, composta por caçadores nômades é uma das raças menos encontradas em *Eorzea*, de acordo com o enredo do jogo. Seus clãs se diferem pelo período em que caçam, assim como se são uma sociedade patriarcal, *Seekers of the Sun*,

ou matriarcal, *Keepers of the Moon*. Estes são diferenciados de seus semelhantes pelos seus pelos mais escuros, olhos mais redondos, orelhas, caninos e caudas maiores.

A raça em *Final Fantasy XIV* com maior estatura são os *Roegadyn* com silhuetas altas, musculosas e cujos atributos físicos são maiores que qualquer outra raça. Essa raça se dividiu entre o clã *Hellsguard* que decidiu residir nas montanhas, e acabaram desenvolvendo uma coloração de pele mais morena, e o clã *Sea Wolves* que decidiu habitar as cidades costeiras como pescadores, piratas e marinheiros, desenvolvendo no processo uma pele de coloração azulada.

A última raça a ser mencionada é a raça *Au Ra*, esta que foi somente introduzida na expansão *Heavensward* e necessita da mesma para ser usada. Essa raça é associada a dragões e possuem chifres laterais, rabo e escamas. Um atributo interessante desta raça é a diferença na altura entre seus membros, com *Au Ras* masculinos sendo muito maiores que *Au Ras* femininos. Membros do clã *Raen* adotaram estilos de vida tranquilos e abandonaram o estilo nômade de seus ancestrais, desenvolvendo escamas claras como consequência. Em contraste à esse estilo de vida, os membros do clã *Xaela* preferem a vida em constante peregrinação, caçando, coletando recursos e lutando. Esse estilo de vida levou-os a desenvolver escamas mais escuras.

Duas raças foram adicionadas na expansão *Shadowbringers*, uma delas exclusivas para personagens femininos, *Viera*, e outra exclusiva para personagens masculinos, *Hrothgar*. Ambas raças não serão detalhadas nesse trabalho pelo lançamento tardio no desenvolvimento deste trabalho.



Figura 4.1: Criação de Personagem

Após a escolha da raça, o jogador deverá escolher customizar a aparência do personagem. As opções são inúmeras, como tamanho e posição relativa de cada parte do rosto, composição física, cor e corte de cabelo, tatuagens, cicatrizes, cor da pele e voz. A Figura 4.1 mostra a tela neste passo da criação do personagem durante a escolha do corte do cabelo.

A próxima etapa consiste na escolha do aniversário do seu personagem e no deus padrinho do mesmo. Essas escolhas não afetam muito o *gameplay*, mas ajudam na imersão e identificação com o personagem. Em seguida, deverá ser escolhida a classe na qual o jogador iniciará o jogo, essa escolha impactará em qual cidade o jogador iniciará. Por fim, o último passo nesse processo é a escolha do *World*, ou servidor, onde o jogador irá jogar. Terminando essas seleções o jogo será iniciado com as customizações feitas.

4.4 Gameplay

Uma vez que personagem foi criado à imagem que o jogador desejou, começa a ação de jogar *Final Fantasy XIV*. Sendo um jogo complexo, com milhões de possibilidades e uma quantidade imensa de conteúdo, não será viável citar todos os elementos de *gameplay* dentro deste trabalho, então este relato será limitado aos principais componentes, assim como aqueles que serão relevantes para esta pesquisa.

Ainda relacionado com a aparência do personagem, existe um sistema chamado *glamour* que permite que um jogador possa equipar seus melhores equipamentos, mas customizar a aparência de cada parte do equipamento para que visualmente seja exibido seus equipamentos que mais o visualmente agradam.

O combate com inimigos constitui uma parte fundamental do jogo, com o personagem tomando parte nele utilizando uma combinação de ataques físicos especiais e habilidades mágicas. Tais ações são aprendidas pelo personagem através de uma progressão por níveis e execução de missões secundárias especiais. Outro meio de melhorar a eficácia das ações do personagem é o aprimoramento do equipamento usado, quanto melhor o equipamento, maiores são os status do personagem, estes que ditam a eficácia das habilidades e ataques do mesmo.

Esses combates podem ser realizados sozinhos ou com um grupo de personagens pertencentes a outros jogadores. Dependendo da atividade pode ser necessário tamanhos específicos de grupos. Para encontrar outros jogadores, existem mecanismos de correspondência automatizada que classifica os jogadores em grupos para o conteúdo instanciado selecionado, assim como serviços para jogadores poderem recrutar outros jogadores para qualquer tipo de conteúdo.

Dentro do jogo existem guildas dirigidas por jogadores, essas são chamadas de *Free Company*, ou Companhias Livres. Membros de uma dessas guildas podem ter acesso à um inventário compartilhado da guilda, uma canal de bate-papo privado e ações especiais que dão vantagens por pequenos intervalos de tempos aos membros, como recebimento de mais experiência que o normal. Os membros da *Free Company* também podem reunir seus recursos para comprar uma casa, ou comprá-la independente da *Free Company*, em um dos distritos residenciais. Além de decorar a casa, os jogadores podem usar os terrenos para cultivar itens únicos através do sistema de jardinagem, treinar seu companheiro animal, embarcar em expedições de aeronaves e comprar um quarto particular para uso pessoal.

Além dessas guildas de jogadores, existem três grandes facções no mundo de *Eorzea*: *Maelstrom*, *Immortal Flames* e *The Order of the Twin Adder*, pelas quais o jogador poderá jurar lealdade, participar de seus exércitos e avançar de patentes. *Maelstrom* fica localizada na cidade costeira de *Limsa Lominsa* e possui um esquadrão de navios à seu dispor, sendo bem caracterizado sob seu padrão militar e naval. Membros dos *Immortal Flames* estão situados na cidade mercantil de *Ul'Dah*, e por conta de seu foco econômico, possui muitos mercenários e guardas reais em seus números. Finalmente temos a facção *The Order of the Twin Adder* da cidade florestal de *Gridania*, uma cidade que procura paz e convivência com outros. Entretanto existem muitos conflitos com tribos bestiais, com propósito de manter a paz, essa facção foi criada.

Para interagir com todos esses elementos de *gameplay*, o jogador utiliza o controle de Playstation 4, caso esteja jogando no mesmo, ou um teclado e mouse, caso esteja jogando no computador, com a possibilidade de utilizar outros acessórios caso deseje. A figura 4.3 mostra a visão padrão de um jogador de Playstation 4, enquanto a figura 4.2 mostra a visão padrão do jogador de computador. As habilidades do personagem encontram-se na parte inferior central da tela, nelas é possível notar uma diferença entre os jeitos de exibição, isso acontece porque a *User Interface* do Playstation 4 é customizada para refletir o posicionamento dos botões no controle. Cada elemento da *User Interface* é ajustável em termos de tamanho, posição e exibição de acordo com o que o jogador desejar.

4.5 Estilos de Jogo

Os personagens criados podem desempenhar diversos papéis dentro do jogo, permitindo que o usuário jogue de diferentes formas dependendo de sua preferência. Tais papéis pertencem à variadas categorizações. Uma dessas categorizações são os Discípulos incluindo: Discípulos da Guerra, que contemplam aqueles que utilizam ataques físicos contra seus inimigos, Discípulos da Magia, que contemplam aqueles que fazem uso da magia para atingir seus objetivos, Discípulos da Terra, aqueles que reúnem e colhem os recursos na-



Figura 4.2: Configuração Padrão da HUD do PC

turais de mundo virtual, e finalmente, Discípulos da Mão, que usam esses recursos para fabricar mercadorias e obter lucro.

Final Fantasy XIV possui diversas classes, adequadas a todos os estilos de aventura. Cada classe é capaz de realizar ações exclusivas adquiridas através da experiência. Além disso, alcançar certo nível de domínio em uma determinada classe dos Discípulos da Guerra ou Magia irá desbloquear seu respectivo *job*, que apresenta uma especialização ainda maior. O jogador pode possuir todas as classes do jogo, nenhuma exclui a possibilidade de se obter outra.

Cada classe pode se transformar em um único *job*, exceto os *Arcanist* que podem se especializar em *Scholar* ou *Summoner*. As classes pertencentes aos Discípulos da Guerra ou da Magia são obrigatórias para a realização das missões da história principal, entretanto se o jogador procura atividades mais pacíficas, pode considerar classes pertencentes à Discípulos da Terra ou Discípulos da Mão.

Além dessas categorizações, cada classe pode pertence a uma *role*, ou função, estas sendo: *Tank*, *DPS*, *Healer*, *Gatherers* e *Crafter*. Todas classes estão listadas nas Tabelas 4.1 e 4.3, relacionando-as com suas armas, *jobs*, *Role*, qual Discípulos que tal classe pertence e quando essa classe foi disponibilizada. Existem alguns *jobs* que não são associados com classes, esses foram os lançados após o jogo base e são iniciados em níveis superiores à um.

O jogador pode mudar sua classe ou *job* quando quiser, se estiver fora de uma atividade em andamento. Essa mudança ocorre apenas mudando de arma equipada. Entretanto a classe deve ser previamente desbloqueada através de missões específicas.



Figura 4.3: Configuração Padrão da HUD do Console

Com o lançamento de novas expansões para o jogo, novas *jobs* foram adicionadas no jogo, essas não necessitam de uma classe para serem habilitadas, sendo apenas necessário a conquista de um certo nível em qualquer classe de Discípulo de Guerra ou Mágica. Essas *jobs* estão listadas com suas expansões de lançamento na Tabela 4.2.

4.5.1 Tank

Quando o jogador está jogando com classes e *jobs* da função *Tank*, o trabalho dele é ser a vanguarda de seus grupos, muito associado com o líder da equipe. Responsáveis por iniciar combates, controlar o posicionamento do inimigo e assegurar que nenhum outro membro do grupo está sendo atacado por inimigos, ou seja, responsáveis pelo controle e prevenção de perigos tanto para si quanto para seus companheiros. *Tanks* fazem esses trabalhos com habilidades que mitigam o dano recebido e habilidades que geram *enmity*, ou inimizade, quanto maior sua *enmity* maior foco de inimigos no seu personagem.

Identificados pela cor azul, essa função contempla as classes *Marauder* e *Gladiator*, assim como as *jobs*: *Dark Knight*, *Paladin* e *Warrior*. A maioria das armaduras dos *Tanks* é compartilhada entre a função, com algumas exceções, e essas são caracterizadas por designs que priorizarem a proteção no lugar da mobilidade. Entretanto cada *job* possui sua arma característica: *Dark Knights* possuem a grande espada, *Warriors* possuem o machado e os *Paladins* possuem a espada e o escudo, como é mostrado na Tabela 4.1.

Apesar de possuírem a mesma função, cada *job* executa-a de maneira diferente. *Paladins* possuem as melhores habilidades para mitigação de dano e algumas para recuperação de vida, entretanto seu potencial de dano é o menor dos *Tanks*. *Warriors* possuem o maior

Classe	Discipulos de	Role	Arma	Job
Gladiator	Guerra	Tank	Machado	Warrior
Marauder	Guerra	Tank	Espada e Escudo	Paladin
-	Guerra	Tank	Espada Grande	Dark Knight
Conjurer	Mágica	Healer	Cajados	White Mage
-	Mágica	Healer	Globos Estelares	Astrologian
Arcanist	Mágica	DPS/ Healer	Livros	Summoner/ Scholar
Pugilist	Guerra	DPS	Soqueira	Monk
Lancer	Guerra	DPS	Lança	Dragoon
Thaumaturge	Mágica	DPS	Cetros	Black Mage
Archer	Guerra	DPS	Arco e Flecha	Bardo
Rogue	Guerra	DPS	Adagas	Ninja
-	Guerra	DPS	Armas de Fogo	Machinist
-	Guerra	DPS	Katana	Samurai
-	Mágica	DPS	Florete	Red Mage

Tabela 4.1: Discípulos de Guerra e Mágica

Job	Disponibilidade
Astrologian	Heavensward
Dark Knight	Heavensward
Machinist	Heavensward
Samurai	Stormblood
Red Mage	Stormblood

Tabela 4.2: *Jobs* lançadas em Expansões

potencial de dano desta função e uma quantidade muito grande de vida, sua desvantagem está no seu conjunto de habilidades de mitigação de dano. Por fim, *Dark Knights* existem no entretermos das *Jobs* anteriores, possuindo boas habilidades de mitigação de dano e um bom potencial de dano.

4.5.2 *Healer*

Jogando com a função de *Healer*, o jogador deve ter a atenção redobrada com seus aliados, pois seu foco é proteção e recuperação de seus aliados, associado com um papel médico presente na retaguarda da equipe. Comumente irá cuidar para que o *Tank* não perca toda sua vida, entretanto quando outros personagens levem dano ou sejam infligidos com alguma status negativo, seu trabalho será recupera-los. Quando possível, o *Healer* também poderá auxiliar o grupo causando dano, causando status negativos no inimigo e distribuindo status positivos ao grupo. Resumindo, *Healers* são responsáveis pela recuperação e manutenção da boa condição do grupo.

Relacionados com a cor verde, existem três *jobs* de *Healers* atualmente no jogo, *Scholars*, *White Mages* e *Astrologians*. Peculiarmente, somente uma classe é associada a essa função, o *Conjurer*. Tal situação deriva dos fatos que o *job Astrologian* veio na expansão *Heavensward* e o *Scholar* é obtido da classe *Arcanist*, que é associada com a função de *DPS*. As roupas compartilhadas pelos *Healers*, costumam ser vestes, comumente relacionadas com mágica.

Utilizando diferentes armas, cada *Healer* realiza sua função de feitiços diferentes. *White Mages* usam cajados e possuem as melhores habilidades para recuperação de vida, focando na recuperação rápida pós dano. Por outro lado, os *Scholars* possuem habilidades que geram mitigação de dano nos membros da equipe além de algumas habilidades de cura, tendo que prever quando a equipe vai levar dano para mitiga-lo. Uma peculiaridade dos *Scholars* que é compartilhada com *Summoners*, por pertencerem à classe *Conjurer*, é a utilização de tomos para invocação de familiares que os auxiliam em suas funções, no caso do *Scholars*, eles invocam fadas para curar e retirar status negativos. Finalmente, o *Astrologian* possui habilidades para recuperação e mitigação de danos com uma mecânica de cartas randômicas que podem melhorar atributos de seus companheiros.

4.5.3 *DPS*

O foco do *DPS*, ou *Damage Per Second*, consiste na inflição de dano no oponente. Sua função principal é maximizar o dano causado no menor espaço de tempo, ou seja, personagens dessa função costumam focar mais na interação deles com o inimigo que com os companheiros de equipe. Esta função é associada com a cor vermelha dentro do jogo.

Entre as *jobs* desta categoria, existem três grupos que jogam de maneiras diferentes para atingir o mesmo objetivo.

O primeiro grupo, *Meele DPS*, executa suas ações em proximidade ao inimigo, muitas vezes tendo de se posicionar em relação à direção que o mesmo está voltado para maximizar o poder de suas habilidades. Pertencentes à essa subcategoria estão: *Dragoons*, *Monks*, *Samurais* e *Ninjas*. *Dragoons* são associados com técnicas para matar dragões e lanças, estas que incluem habilidades de pulo, onde o personagem cruza grandes distâncias para executar um ataque em curto tempo com um pulo, ele também possui habilidades para melhorar os atributos de outros personagens por curto tempo com intenção de aumentar o dano causado pela equipe. Similarmente, o *Ninja* utiliza técnicas de *Ninjutsu* para melhorar os atributos tanto de si mesmo quanto de sua equipe. Diferentemente dos *jobs* citados acima, *Samurais* e *Monks* focam na inflição de dano puro, através de técnicas marciais com katanas e punhos respectivamente, sem interação com outros membros da equipe, exibindo um dos maiores potenciais de dano do jogo, entretanto perdem para equipes bem coordenadas com *jobs* que se complementam.

O grupo seguinte utiliza habilidades a distância e foca na inflição de condições no inimigo que causam dano constante, assim como fornecer benefícios para a equipe. Este grupo é denominado *Physical Ranged DPS* e nele estão duas *jobs*: *Bards* e *Machinist*. Personagens equipados com a *job Bard* utilizam músicas para fornecer benefícios e flechas para atacar seus inimigos. *Machinist* utilizam armas e drones para controlar grupos de inimigos e causar dano em grandes inimigos.

O último grupo de *DPS* engloba as *jobs*: *Summoner*, *Black Mage* e *Red Mage*, configurando os *Magic Ranged DPS*. Esses *jobs* possuem um excelente potencial de dano, entretanto são limitados em seu movimento pelo tempo de lançamento dos feitiços, momento em que os personagens devem ficar parados até o tempo de *cast* acabar para enfim lançar seus feitiços. *Summoners* se especializam em causar danos contínuos em seus inimigos com auxílio de monstros invocados pelo o mesmo. Os *Black Mages* balanceiam seus feitiços elementais para causar uma grande quantia de dano de uma vez só. Finalmente, o *Red Mage* dispõem de uma gama de habilidades mágicas que variam de feitiços de cura, dano e até ataques físicos.

4.5.4 *Crafters*

Nota-se a partir da Tabela 4.3 que os Discípulos da Mão possuem uma quantidade numerosa de membros, dispondo de oito classes. Entretanto apesar de seus números, todas fazem basicamente a mesma atividade, criam itens utilizando materiais. Essas classes exercem sua função aplicando habilidades em conjunto com receitas e materiais, essas que aumentam a possibilidade do item criado ser de ótima qualidade e aumentam o progresso

Classe	Discípulos de	Arma
Carpenter	Mão	Serrote
Blacksmith	Mão	Martelo
Armorer	Mão	Martelo
Goldsmith	Mão	Malho
Leatherworker	Mão	Faca
Weaver	Mão	Agulha
Alchemist	Mão	Alambique
Culinarian	Mão	Panela
Botanist	Terra	Machadinha
Miner	Terra	Picareta
Fisher	Terra	Vara de Pescar

Tabela 4.3: Discípulos da Mão e Terra

da confecção do item. Se o progresso não for terminado dentro do limite de ações definido pela receita, a criação do item falha.

Carpinders criam itens com base em madeira, como lanças, arcos e alguns móveis. *Blacksmith* criam itens de ferreiro, com foco em armas, como espadas, escudos, machados e outros. *Armorers* focam nos itens de armaduras pesadas. *Goldsmiths* criam itens com jóias. *Leatherworker* se especializam em itens feitos de couro. *Weavers* utilizam de tecido para confeccionar seus itens. *Alchemist* criam itens com base na alquimia, comumente sendo poções. Finalmente, o *Culinarian* cria comidas que fornecem benefícios temporários a quem as comem, esses que variam de melhoras em habilidades de combate ou até de confecção ou coleta de itens para outras classes. Essa função costuma ser muito associada à geração de dinheiro, pois itens mais complexos feitos por *Crafters* avançados são vendidos por jogadores por preços altos.

4.5.5 *Gatherers*

A última função discutida possui apenas três classes, estes que são os *Gatherers*, presentes na Tabela 4.3 que referenciam os Discípulos da Terra e da Mão, relacionando-os com sua arma característica. Membros desta categoria tem como foco a coleta de materiais pelo mundo em locais específicos, utilizando habilidades para aumentar a probabilidade de coletar um item, assim como habilidades para que o mesmo seja de boa qualidade. Esta é uma função que não dependem de outras classes ou de outros personagens, ao contrário dos *Crafters* que dependem dos materiais. Quando o personagem esta com essas classes, ele procura evitar conflitos no mundo, pois esta é uma função pacífica e não possui habilidades de ataque, sendo assim existem habilidades com propósito de se mascararem de monstros fracos.

Botanists são responsáveis pela coleta de materiais derivados de plantas e árvores, seus métodos de coleta são semelhantes à classe *Miner* que executa as mesmas ações em pedras para coletar materiais derivados de minerais e pedras. Diferentemente das classes anteriores, que precisam apenas selecionar o item que querem coletar e aguardam a resposta para saber se o item foi coletado ou não, a classe dos *Fisher* deve prestar atenção na linha da vara de pescar para selecionar a ação de recolher a linha quando o peixe for fisdado. Os pontos de coleta das primeiras classes mencionadas são facilmente identificadas no mapa, entretanto *Fishers* devem tentar utilizar suas varas de pescar em locais com água para descobrir se podem pescar lá ou não.

No Patch 4.5, lançado no dia sete de Janeiro de 2019, foi introduzido no jogo um *limited job*, o *Blue Mage*. Entretanto esse novo *job* não fará parte desta análise. Assim sendo, ele não será discutido neste capítulo.

4.6 Atividades

As seções anteriores detalharam o mundo virtual, a criação do personagem e que tipo de ações ele pode fazer dentro desse mundo. Nesta seção serão detalhadas as diversas atividades que estão disponíveis para o jogador realizar dentro do jogo. Dentre tais atividades, uma das mais notáveis é a Missão Principal, ou *Main Quest*, missões impactantes no mundo e necessários para habilitar outras missões secundárias. Qualquer missão que não faça parte desses conjuntos de missões da Missão Principal é classificada como Missão Secundária, ou *Side Quest*, missões que não afetam tanto o mundo virtual e comumente são auto contidas em suas narrativas.

Após as execuções de determinadas missões, diversas atividades serão desbloqueadas, assim como exigidas do jogador para o progresso de algumas missões. Dentre as possíveis atividades, as quatro principais são realizadas em grupos e em instâncias separadas dos outros jogadores, isolando o grupo nelas. Tais atividades podem ser: *Dungeons*, aventuras contendo quatro chefões, *Trials*, uma intensa batalha contra um grande chefe, *Raids*, aventuras contendo quatro grande chefões, e Atividades *PvP*, que constituem cenários onde os jogadores enfrentam outros jogadores.

4.6.1 *Main Scenario Quests*

A trama principal do *Final Fantasy XIV*, discutida na Secção 2, é experienciada durante as missões presentes na *Main Scenario Quests*. Novas áreas e mecânicas costumam ser introduzidas por essas missões.

Na atualização do *Patch 4.5 Part 1*, introduzido em Janeiro de 2019, foram adicionadas cinco missões novas nesta categoria para a expansão atual na época, *Stormblood*,

totalizando 585 missões. Desse total, 187 foram introduzidas no jogo base. Nas atualizações seguintes, do 2.1 até o 2.5, mais 100 missões foram adicionadas. A primeira expansão, *Heavensward*, trouxe 94 novas missões, entretanto somente os jogadores que compraram essa expansão teriam acesso à esse conteúdo. Assim como no jogo base, as atualizações que vieram depois, do 3.1 até 3.5, trouxeram mais 44 missões para a trama principal. As missões seguintes fazem parte da terceira expansão, *Stormblood*, compondo as próximas 122 missões. Finalmente, as últimas 38 missões foram lançadas até agora em conjuntos nas atualizações do 4.1 até o 4.5 parte 1.

Nas primeiras 28 missões, existe uma divergência entre algumas classes, pois as classes *Gladiator*, *Thaumaturge* e *Pugilist* começam suas aventuras na cidade de *Ul'Dah*, enquanto *Conjurer*, *Lancer* e *Archer* iniciam a jornada em Gridania. As classes de batalha restantes tem seu início em *Limsa Lominsa*. Uma vez que o jogador inicia essas missões não será possível realizar-las novamente, nem trocar de cidade inicial. Entretanto, as animações presentes nessas missões podem ser revistas. Depois dessas missões iniciais, todas classes seguem o mesmo fluxo de missões.

Estas missões não dependem da conclusão de nenhum outro tipo de missão, algumas podem necessitar que o jogador atinja um determinado nível de qualquer classe de batalha ou melhores equipamentos, mas fora tal requisito, essas missões não são restringidas por outros fatores. Por outro lado, outros tipos de missões podem ser restringidas até que o jogador avance nestas missões.

Os objetivos presentes dentro destas missões variam muito. Algumas missões podem necessitar que jogador interaja com personagens de jogo, outras podem necessitar que o jogador complete um *Trial* ou *Dungeon*, algumas missões introduzem novos elementos de *gameplay*, como atirar em uma perspectiva de primeira pessoa. Algumas missões exigem que o jogador mate um determinado número de inimigos, outras podem necessitar que um tipo específico de inimigo seja derrotado e certas missões pedem que um alvo seja protegido contra inimigos.

4.6.2 *Side Quests*

Encontradas em muitos jogos do gênero *RPG* estão as missões secundárias, ou *Side Quests*. Dentro desta classificação estão todas missões dentro do jogo que não fazem parte da *Main Scenario*. Devido ao tamanho e dimensão de jogos MMORPGs, não será explorado todas variações e possibilidades de missões secundárias. Entretanto alguns tipos serão esclarecidos nesta subsecção por representarem partes importantes do jogo, logo de grande importância para análise de jogadores.

Todo conteúdo do jogo pode ser acessado por qualquer jogador, entretanto uma parte considerável do conteúdo está bloqueada por missões específicas. Assim sendo, o jogador

deve completa-las para ter acesso ao conteúdo. Essas missões variam muito de objetivos, desde uma simples conversa pra liberar tal conteúdo, até uma série de missões complexas que apenas no final será liberado algo. O conteúdo aqui referido, costuma ser novas classes, novas *Dungeons, Raids, Trial*, novas opções de PvP, novos mini-games ou até novas mecânicas. Essas missões são diferenciadas das demais por um símbolo do ponto de exclamação azul com um sinal de soma ao lado.

Uma vez habilitada, cada classe e *job* possuem habilidades que só podem ser aprendidas através da finalização de missões próprias das mesmas, estas que só serão disponíveis após certo nível de experiência com a *job* correspondente. Essas missões costumam envolver histórias mais complexas e engajantes que a maioria das missões secundárias, assim como também oferecem equipamentos exclusivos para seus *jobs*. Sua importância é muito grande, pois o jogo foi feito considerando que o personagem possui todas habilidades disponíveis ao aceitar entrar em uma atividade separada do mundo compartilhado.

Existem três séries de missões secundárias que também fornecem um bom discernimento sobre jogadores que as completam. A primeira envolve as tribos bestiais que costumam ser inimigas das principais cidades do mundo. Essa série de *side quests* oferece uma quantidade limitada diariamente de missões repetíveis. Cada vez que uma missão é completada, sua relação com aquela tribo específica aumenta, até um certo nível. As recompensas para atingir o nível máximo costumam ser armaduras visualmente atraentes, mas com pouco valor em combate e montarias que permitem que o jogador percorra mais rápido o terreno. Pelo limite diário de execução, essas missões costumam demorar meses para alcançar o nível máximo de cada tribo bestial. O jogo base possui cinco tribos bestiais com esse tipo de missão, e cada expansão trouxe mais três tribos, sendo que uma em cada expansão e uma no jogo base são exclusivas para classes discípulas da mão ou terra.

O segundo tipo de missão secundária que será mencionado são as missões relacionadas à arma mais forte e mais visualmente complexa de cada expansão e jogo base. Esse conjunto de missões envolve o jogador adquirindo um arma e evoluindo a cada missão cumprida. Essas são armas especiais identificadas por uma coloração roxa de raridade. Essas missões costumam ser as missões mais consumidoras de tempo do jogo, necessitando de muita dedicação para conseguir completa-las na atualização em que foram lançadas. O jogo base possui as *Zodiac Weapons*, *Heavensward* possui as *Anima Weapon* e a última expansão possui as *Eureka Weapons*.

A última série de missão que será mencionada é a linha de missões do Inspetor Hildibrand. As recompensas dessas missões incluem poucas roupas, umas animações, e alguns *minions* (bonecos digitais pequenos que seguem o jogador sem afetar gameplay em áreas públicas). Entretanto o foco dessas missões é a comédia. Diferente do tom da maioria

das histórias contadas no jogo, as situações encontradas nessas missões se assemelham à personagens caricaturados em desenhos animados, com circunstâncias cuja física comum do jogo não é respeitada. Essas missões servem como um contraste às missões normais, principalmente na *Main Scenario* que visa tratar de assuntos mais maduros.

4.6.3 *Dungeons*

Uma das principais atividades que jogadores podem realizar juntos são as *Dungeons*, também conhecidos como calabouços. Nessas atividades quatro jogadores são colocados em uma instância privada, com seus níveis sincronizados para balancear o desafio da atividade, onde devem percorrer um caminho repleto de inimigos e desafios. Atividades assim costumam agrupar um personagem da função *Tank*, um *Healer* e dois *DPS*, entretanto é possível realizar essas atividades com qualquer combinação. Dentro dessas instâncias, existem quatro inimigos maiores, ou chefes, que possuem muito mais vida que inimigos normais e limitam uma arena para enfrenta-los. Depois de derrotar cada chefe, os jogadores são recompensados com equipamentos superiores ao nível da atividade. Entretanto, os jogadores devem disputar esses itens entre si. Essa disputa é resolvida através de um número aleatório gerado para cada jogador quando ele expressa que deseja obter o item, quem tiver o maior número receberá o item. Como alguns equipamentos são exclusivos para certas classes e *jobs* os usuários que estão usando-as no momento dessa recompensa, recebem prioridade sobre os outros.

Se um jogador morre durante essa atividade, ele pode escolher reviver no início da *Dungeon* ou esperar para outro jogador o reviver. Entretanto se todos jogadores morrerem, todos iniciam no início, se isso acontecer durante uma batalha com um chefe, ele irá recuperar sua vida em total. Existe um tempo para finalização deste tipo de atividade, se ela finalizar, todos são expulsos da respectiva instância.

4.6.4 *Trials*

Trials são atividades de oito jogadores, compostas por dois *Tanks*, dois *Healers* e seis *DPS*. Entretanto, apesar de terem sido desenvolvidas para essa composição, é possível enfrentar o desafio com outras composições se todos os envolvidos aceitarem. Esta atividade envolve apenas um luta em uma arena.

Apesar da simples premissa, essa luta possui diferentes mecânicas que exigem coordenação e reflexo. Assim como *Dungeons*, *Trials* possuem um tempo limite para serem completados, caso todos jogadores pereçam, a batalha é reiniciada, mas o tempo continua correndo. Se um jogador morrer, ele só poderá voltar à luta se um jogador revive-lo.

Existem vinte e sete *Trials* no jogo base com diferentes dificuldades da mesma luta. Sete pertencentes à maior dificuldade *Extreme*, com essas mesmas batalhas possuindo uma versão na dificuldade *Hard*. Dessas sete, três possuem uma versão da dificuldade padrão. Todas outras *Trials* não possuem dificuldades diferentes. A primeira expansão possui catorze *Trials*, sendo compostas por sete batalhas com dificuldades *Hard* e *Extreme*. A segunda expansão possui uma batalha a mais com ambas dificuldades da primeira. Muitas dessas *Trials* foram lançadas em atualizações.

4.6.5 *Raids*

Esta atividade pode funcionar de forma muito similar à *Dungeons* ou à *Trials*. Existem dois tipos de *Raids*, uma com vinte e quatro jogadores juntos e outra com oito. A *Raid* de vinte e quatro jogadores funciona exatamente como uma *Dungeon* com quatro chefes, entretanto os jogadores são divididos em três alianças de oito membros cada, com cada aliança disputando recompensas entre si. A composição de cada aliança é igual à de *Trials*, entretanto um *Tank* é substituído por um *DPS*. Existem nove atividades assim no jogo, com três em cada expansão e três no jogo base, todas foram liberadas em atualizações pós lançamento.

As *Raids* de oito jogadores se assemelham à *Trials*, com foco em um inimigo de complexidade maior. Algumas podem envolver percursos antes do chefe ou até nenhum chefe, entretanto após a primeira expansão o padrão foi estabelecido com um chefe em cada atividade dessas. A composição de personagens é exatamente igual à *Trials*. Essas atividades foram lançadas em três grupos de quatro em atualizações intermitentes, totalizando doze em cada expansão e treze no jogo base, pois o jogo base continha uma fase a mais no primeiro grupo. O segundo grupo de *Raids* desse tipo do jogo base e todas *Raids* de oito das expansões continham uma atividade extremamente difícil chamada *Savage*, que são as atividades mais difíceis de todo o jogo em seus lançamentos, recompensando seus vencedores com os equipamentos mais fortes disponíveis.

4.6.6 *PvP*

Atividades onde jogadores jogam contra outros jogadores são conhecidas como atividades *Player vs Player* ou *PvP*. Tais dinâmicas variam entre mini-games com cartas, corridas de animais e Mahjong. Entretanto a atividade *PvP* mais proeminente são partidas onde jogadores utilizam habilidades de suas classes e *Jobs* para executar seus oponentes e atingir objetivos específicos em grupos. Tais objetivos variam dependendo do tipo de partida.

Recompensas para o sucesso em partidas assim são armaduras exclusivas, mascotes e montarias. Existe um sistema de ranqueamento entre jogadores que participam de

partidas ranqueadas. Uma posição alto pode também ser considerado uma recompensa por si próprio, para alguns jogadores com um perfil mais competitivo.

4.6.7 *High End*

Atividades desta categoria são destinadas aos jogadores com maior dedicação, habilidade e coordenação com suas equipes. Esta categoria pode inclui *Trails* da dificuldade *Extreme* e *Raids* da dificuldade *Savage* durante suas atualizações de lançamento. Essas atividades possuem os melhores equipamentos como recompensa e possuem a menor tolerância para falhas. Pela sua dificuldade, essas atividades servem como objetivo final para jogadores aspirarem quando lançadas.

Capítulo 5

Base de Dados

Este capítulo constitui uma descrição da Base de Dados que foi utilizada nesta pesquisa. É feita uma descrição da base, assim como do projeto que a disponibilizou. Sua estrutura, informações e limitações também serão descritas.

5.1 Descrição

O estudo comparativo de técnicas de clusterização depende do conjunto de dados no qual elas serão aplicadas. O qual deve ser capaz de fornecer um discernimento sobre cada jogador.

Algumas empresas desenvolvedoras de jogos fornecem um banco de dados oficial e/ou uma Interface de Programação de Aplicações (API) para acesso à um banco de dados privado. Um exemplo muito utilizado em pesquisas do gênero é o World of Warcraft, no qual a desenvolvedora, Blizzard, fornece o acesso à um banco de dados com uma gama ampla de dados.

Com o intuito de ampliar o domínio de jogos MMORPG pesquisados e melhorar a generalização de resultados de pesquisa, o jogo Final Fantasy XIV foi escolhido para este estudo. Entretanto, a Square Enix não fornece uma API ou um banco de dados para acesso à dados do jogo e jogadores.

Para superar esse empecilho, será usado uma base de dados coletada, tratada e disponibilizada pelo site FFXIV Census [27]. Este site coleta suas informações mensalmente da aplicação web Lodestone [131], site oficial da comunidade do Final Fantasy XIV administrado pela Square Enix. Dentro da Lodestone é possível pesquisar por todos os personagens presentes no jogo, utilizando seus nomes e servidores. O resultado da pesquisa revela diversos atributos do personagem como seu perfil, atributos, níveis de classes, montarias, mascotes e conquistas. As informações contidas nesses atributos, coletadas pelo FFXIV Census, são detalhadas na Seção 6.3.

A próxima seção detalhará os arquivos, de formato SQL, de cada mês, fornecidos pelo site. Entretanto, além de fornecer tais arquivos, o site [27] também disponibiliza estatísticas sobre essa base de dados. A Figura 5.1 mostra a tela inicial do site, esta que exibe as estatísticas do mês vigente. Na página de cada mês, as seguintes estatísticas são exibidas:

- Número de personagens presentes no jogo
- Quantos personagens em cada região: América, Japão, Europa.
- Distribuição de gênero e raça
- Distribuição de Classes
- Distribuição por Servidores
- Distribuição por *Grand Companies*
- Distribuição por tempo inscrito
- Distribuição por Nível superior à 4 das Missões de Tribos Bestiais
- Número de pré-ordem feitas para cada expansão e jogo base.
- Números de compras de edição de colecionador para PC e Playstation 4
- Número de compras de alguns itens físicos reais
- Número de jogadores que completaram diversas linhas de missões ou eventos, como casar online, presenciar um casamento online, receber 50 elogios, completar a linha de missões de Hildibrand e completar as charadas visuais do jogo base.

Tais estatísticas são o ponto de início desta pesquisa, que relacionou as diversas informações contidas na base de dados para melhor compreensão dos agrupamentos, ou clusters, de perfil de jogadores.

O site FFXIV Census [27] faz parte de um projeto maior, *Final Fantasy XIV Statistics Project*, cujos contribuidores são Matt Hillier, Jonathan Price e Peter Reid. Dentro deste projeto existem os programas *XIV Stats Gatherer (Java)*, *XIVStats* e *XIVStats-Gatherer-Ruby* (este que foi descontinuado). O projeto foi criado com intenção de auxiliar outros projetos pessoais que possam utilizar tais dados, um exemplo citado é a comparação de jogadores.

O programa mais relevante dentro desse projeto para esta pesquisa é o *XIV Stats Gatherer (Java)*. Este faz a coleta das informações de todos os personagens registrados no jogo e cria um banco de dados com a estrutura apresentada na Seção 6.2. Como mencionado anteriormente, essas informações são retiradas do Lodestone [131], utilizando

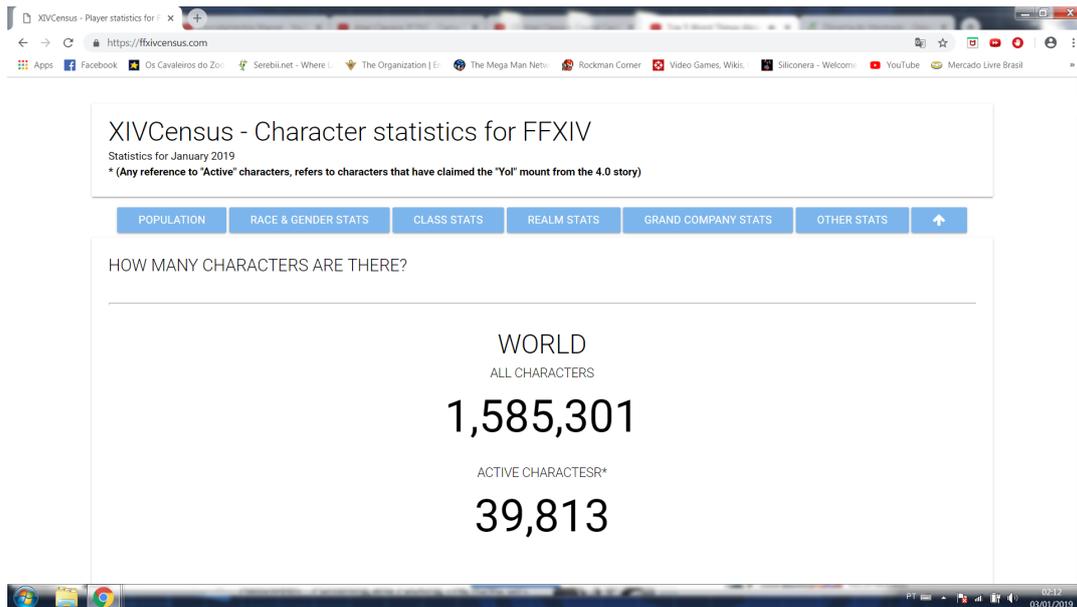


Figura 5.1: FFXIV Census

uma gama de ids de personagens, e recuperando todos personagens que possuem ids dentro desse conjunto.

5.2 Estrutura

O site [27] permite a importação dos dados coletados por intermédio de arquivos, comprimidos, do formato *.sql*. Tal arquivo deverá ser executado em um esquema de banco de dados relacional SQL, dentro de um servidor próprio do usuário. Uma vez executado, o código criará uma tabela, com nome de *tblplayers*, seguindo a estrutura apresentada na Tabela 5.1 e a populará com os registros dos personagens. A única tabela que estará presente no banco de dados possuirá setenta e sete colunas, ou seja, cada registro contido nela, possuirá 77 atributos relacionados à um único personagem registrado no Final Fantasy XIV. Em Novembro de 2018, o arquivo, referente à tal mês, possuía 13,519,949 registros de personagens.

5.3 Conteúdo

A pesquisa usou esta base de dados como fundação para um estudo comparativo de técnicas de clusterização em jogadores de MMORPG. Para realização deste trabalho, será necessária uma interpretação dos dados contidos nessa base. Utilizando uma descrição da base, fornecida pelos autores do site, e associando-a com o conhecimento dos autores

Nome do Atributo	Tipo de Dado
id	Inteiro
name	Texto
realm	Texto
race	Texto
gender	Texto
grand_company	Texto
free_company	Texto
level_gladiator	Inteiro
level_pugilist	Inteiro
level_marauder	Inteiro
level_lancer	Inteiro
level_archer	Inteiro
level_rogue	Inteiro
level_conjurer	Inteiro
level_thaumaturge	Inteiro
level_arcanist	Inteiro
level_darkknight	Inteiro
level_machinist	Inteiro
level_astrologian	Inteiro
level_scholar	Inteiro
level_redmage	Inteiro
level_samurai	Inteiro
level_carpenter	Inteiro
level_blacksmith	Inteiro
level_armorer	Inteiro
level_goldsmith	Inteiro
level_leatherworker	Inteiro
level_weaver	Inteiro
level_alchemist	Inteiro
level_culinarian	Inteiro
level_miner	Inteiro
level_botanist	Inteiro
level_fisher	Inteiro
p30days	Binário
p60days	Binário
p90days	Binário
p180days	Binário
p270days	Binário
p360days	Binário
p450days	Binário

Nome do Atributo	Tipo de Dado
p630days	Binário
p960days	Binário
prearr	Binário
prehw	Binário
presb	Binário
arrartbook	Binário
hwartbookone	Binário
hwartbooktw	Binário
hasencyclopedia	Binário
beforemeteor	Binário
beforethefall	Binário
soundtrack	Binário
saweternalbond	Binário
sightseeing	Binário
arr_25_complete	Binário
comm50	Binário
moogleplush	Binário
topazcarubuncleplush	Binário
emeraldcarbuncleplush	Binário
hildibrand	Binário
ps4collectors	Binário
dideternalbond	Binário
arrcollector	Binário
kobold	Binário
sahagin	Binário
amaljaa	Binário
sylph	Binário
moogle	Binário
vanuvanu	Binário
vath	Binário
hw_complete	Binário
hw_31_complete	Binário
hw_33_complete	Binário
legacy_player	Binário
mounts	Texto
minions	Texto
date_active	Data
is_active	Binário

Tabela 5.1: Estrutura da Tabela *tblplayers*

decorrente de mais de cem horas de experiência com o jogo, foram derivadas as informações discutidas nas seguintes subseções.

5.3.1 Informações Contidas

Nesta subseção será descrito cada atributo presente na Tabela 5.1.

O atributo *id* é número inteiro utilizado como chave primário para o uso administrativo do sistema de gerenciamento de Banco de Dados, não relaciona nenhuma informação do personagem.

O atributo *name* está relacionado ao nome do personagem em questão, escolhido pelo jogador. Utilizando o atributo *realm* é possível identificar uma possível localização do jogador. Este atributo mostra o nome do servidor desse personagem, e, utilizando a relação do nome do servidor com a localização de servidores, disponibilizada publicamente, é feita uma aproximação de onde o jogador pode estar jogando. Essa aproximação é feita com base na hipótese de que o jogador irá escolher servidores mais próximos do seu local para diminuir possíveis problemas de conexão.

Como mencionado no Capítulo 5, o jogador pode escolher entre seis raças para seu personagem, tal escolha é armazenado em *race*. O sexo do mesmo é registrado como um texto no atributo *gender*. O nome de uma das três uniões do jogo a qual o personagem pertence é armazenado na propriedade *grand_company*, caso ele pertença a uma delas. Semelhante ao nome do clã de jogadores o qual pode pertencer, esse ficará registrado em *free_company*.

Os níveis de experiência de cada um dos vinte e seis *job* do personagem são armazenados nos atributos *level_X* onde *X* é o nome do *job* em questão. O nível é um número inteiro.

Analisando os itens que o personagem possui, foi possível determinar quantas mensalidades o jogador já pagou para jogar. Essas informações são representadas pelos valores binários dos atributos *pXdays*, onde *X* é o número de dias pago. Tal fato advém do fato de que se o jogador pagar uma determinada quantidade de mensalidade, o jogo o recompensará com certos itens exclusivos. Logo, se o jogador possui o referido item, pagou o equivalente em mensalidades. Os possíveis valores para os dias pagos são: trinta, sessenta, noventa, cento e oitenta, duzentos e setenta, trezentos e sessenta, quatrocentos e cinquenta, seiscentos e trinta e novecentos e sessenta.

Os atributos *prerr*, *prehw* e *presb* registram valores binários que identificam se o jogador realizou a pré-compra do jogo base, expansão *Heavensward* ou expansão *Stormblood*, respectivamente. Essa informação também é adquirida realizando uma verificação por itens exclusivos no inventário do personagem.

Informações sobre itens físicos comprados fora do jogo também podem ser adquiridas analisando itens digitais exclusivos, estes são adquiridos como bônus juntos com o item físico. Na Tabela 5.2 estão as relações entre os atributos com os itens físicos comprados.

Além de itens físicos, é possível identificar se o jogador comprou edições de colecionador do jogo base, atributo *arrcollector*, e especificamente a versão de Playstation 4, atributo *ps4collector*.

Algumas propriedades relatam eventos ou ações que os jogadores realizaram ou participaram, assim como quais missões o jogador fez e até quanto progresso o jogador fez na história principal. Um evento importante dentro do jogo é o *Eternal Bond*, evento que espelha o casamento no mundo real. O atributo *swaeternalbond* informa quais personagens já foram convidados para um evento desses e o atributo *dideternalbond* informa quais personagens fizeram essa união com outra personagem.

Uma das mecânicas do jogo inclui o registro de um elogio à outro jogador com quem fez alguma atividade. No momento em que um personagem acumula cinquenta elogios, o mesmo é recompensado com um item exclusivo, identificado com o atributo *comm50*. Outra mecânica do jogo abrange a solução de charadas, que envolvem o personagem indo a certos locais virtuais em certas horas específicas para solucionar-las, caso o personagem solucione oitenta dessas charadas, o item exclusivo recebido é relacionado com o atributo *sightseeing*.

Dentre as missões possíveis para os personagens, existem algumas específicas de raças inimigas, ou Tribos Bestiais, que os jogadores podem fazer para melhorar a relação entre as uniões principais e algumas tribos de raças inimigas. Quanto mais missões o personagem completar, maior será o ranking do personagem com aquela determinada raça. As seguintes raças são representadas pelos correspondentes atributos de mesmo nome: *Kobold*, *Sahagin*, *Amalj'aa*, *Sylph*, *Moogles*, *Vanu Vanu* e *Vath*.

Existe um conjunto de missões cujo principal tema é de cunho mais cômico. Tais missões visam situações absurdas e fora do clima normal do jogo. Completando essa linha de missões presentes no jogo base, o personagem recebe itens exclusivos, permitindo a identificação dos mesmos através do atributo *hildibrand*.

Mencionado anteriormente, alguns atributos podem identificar se o personagem já alcançou certo ponto da história principal do jogo. O atributo *arr_25_complete* identifica se o personagem já exauriu toda história presente no jogo base incluindo todas atualizações antes do lançamento da primeira expansão. Os atributos *hw_complete*, *hw_31_complete* e *hw_33_complete* identificam que o personagem completou a história da primeira expansão antes de qualquer atualização, após completou a primeira atualização da expansão inicial e por fim completou a história da terceira atualização da expansão inicial, respectivamente. Por fim, o atributo *legacy_player* identifica aqueles personagens os quais foram criados

Atributo	Item Físico	Descrição
arrartbook	FINAL FANTASY®XIV: A REALM REBORN™THE ART OF EORZEA -ANOTHER DAWN	Um livro contendo artes conceituais pertencentes ao jogo base.
hwartbookone	FINAL FANTASY®XIV: HEAVENSWARD™THE ART OF ISHGARD -STONE AND STEEL	Um livro contendo artes conceituais pertencentes à primeira expansão do jogo, Heavensward.
hwartbooktw	FINAL FANTASY®XIV HEAVENSWARD™THE ART OF ISHGARD -THE SCARS OF WAR	Um livro contendo artes conceituais pertencentes à primeira expansão do jogo, Heavensward.
hasencyclopedia	ENCYCLOPAEDIA EORZEA - THE WORLD OF FINAL FANTASY XIV	Uma enciclopedia, com termos e histórias sobre o jogo base.
beforemeteor	BEFORE METEOR FINAL FANTASY®XIV ORIGINAL SOUNDTRACK	Um CD com as músicas referentes ao jogo anterior ao jogo base.
beforethefall	BEFORE THE FALL: FINAL FANTASY®XIV ORIGINAL SOUNDTRACK	Um CD com as músicas referentes ao jogo base.
soundtrack	FINAL FANTASY®XIV: A REALM REBORN™ORIGINAL SOUNDTRACK	Um CD com as músicas referentes ao jogo base.
moogleplush	FINAL FANTASY®XIV PLUSH KUPLU KOPO (MOOGLE)	Uma pelúcia de um mascote do jogo.
topazcaruncleplush	FINAL FANTASY®XIV: HEAVENSWARD™TOPAZ CARUNCLE	Uma pelúcia de um dos monstros parceiros do jogador.
emeraldcaruncleplush	FINAL FANTASY®XIV: HEAVENSWARD™EMERALD CARUNCLE	Uma pelúcia de um dos monstros parceiros do jogador.

Tabela 5.2: Relação de Atributos e Itens Físicos

na primeira versão do jogo, lançada em 2010. Detalhes maiores sobre essa primeira versão foram apresentados no Capítulo 5.

Os atributos *mounts* e *minions* revelam quais montarias e quais *minions* o personagem possui.

Finalmente, o atributo *date_active* revela a ultima data que o personagem entrou no jogo. Apesar do fato de que a maioria dos atributos são uteis para análise do perfil dos jogadores, o ultimo atributo, *is_active*, usa critérios definidos pelos criadores do XIVCensus [27] que não correspondem exatamente ao número real de jogadores ativos.

5.3.2 Meses Disponíveis

Como mencionado na Seção 1 deste capítulo, a coleta dos dados é feita mensalmente, assim sendo, cada arquivo disponibilizado para importação é referente à um mês do ano. As informações contidas representam o estado de cada personagem no dia em que os dados foram coletados.

A seguir estão listados os meses que estão disponíveis para importação:

- **2016:** Julho, Agosto, Setembro, Outubro, Novembro, Dezembro.
- **2017:** Janeiro, Fevereiro, Março, Maio, Junho, Setembro, Outubro, Novembro, Dezembro.
- **2018:** Janeiro, Fevereiro, Março, Abril, Maio, Junho, Julho, Agosto, Setembro, Outubro, Novembro.

Qualquer mês anterior à Junho de 2016 não está disponível para importação pois a coleta apenas começou no referido mês. Os meses Abril, Julho e Agosto do ano 2017 não possuem arquivos para importação por razões desconhecidas.

5.3.3 Limitações

Apesar da grande quantidade de registros e atributos coletados, a base de dados utilizada ainda possui muitas limitações se comparadas às outras bases semelhantes de outros jogos.

Primeiramente, existe a possibilidade de jogadores possuírem mais de um personagem em sua conta, ou até um jogador possuir mais de uma conta. Isso impossibilita a suposição de que o número de personagens equivale ao número de jogadores, e por consequência, impedindo que esta pesquisa saiba quantas pessoas efetivamente estão jogando Final Fantasy XIV.

Uma informação altamente relevante para análise de perfil dos jogadores, que não está presente na base estudada, é o tempo da sessão de jogo. O tempo de sessão mede quanto

tempo o jogador passou interagindo com o jogo, desde o momento que realizou o login até a hora que encerra suas atividades. Este tempo é muito utilizado para diferenciar aqueles jogadores considerados *hardcore*, que dedicam muito tempo ao jogo e possuem muita experiência no mesmo, dos jogadores casuais, que não jogam com tanta frequência ou por muito tempo.

Outra informação não contida nesta base que auxiliaria muito nesta pesquisa é o *item level* ou equipamento. Este atributo está relacionado ao *endgame*, mencionado no Capítulo 5 como conteúdo destinado à jogadores que chegaram no nível máximo e procuram os maiores desafios que o jogo pode trazer. Tal conteúdo não é aproveitado pela maioria dos jogadores em seu momento de lançamento, logo essa diferença de jogadores que buscam esse conteúdo forneceria uma riqueza maior à análise.

Finalmente, a falta de um histórico de atividades para cada personagem impede o estudo detalhada do engajamento com o conteúdo desenvolvido, assim como impede a determinação de perfis baseados nas atividades e nas correlações de conteúdo e classes. O estudo sobre o *engame* também é bastante prejudicado, pois sem o acesso ao *item level* e armaduras, este seria um dos poucos meios de avaliar esse tipo de conteúdo.

Capítulo 6

Proposta e Metodologia

Este capítulo iniciará com a proposta desta investigação e, em seguida, serão descritos os procedimentos necessários para execução desta pesquisa, bem como os métodos de avaliação e pré-processamento. O algoritmo desenvolvido nesta pesquisa e mencionado neste capítulo está disponível em um repositório online¹.

6.1 Proposta

O Capítulo 3 evidenciou o interesse da comunidade científica em investigar jogadores, assim como a ampla gama de trabalhos que podem ser realizadas dentro do campo de pesquisa de *Game Analytics*. O gênero MMO possui um notável potencial para fornecimento de dados para tais estudos, este que está ativamente sendo explorado pelas pesquisas relatadas. Entretanto, uma oportunidade foi identificada ao compilar esses artigos. Pesquisas sobre um dos MMORPG mais populares atualmente são poucas. O jogo *Final Fantasy XIV* possui muitas peculiaridades que o tornam um objeto de estudo interessante. Lançado inicialmente em 2010, depois uma nova versão lançada em 2013, é considerado um MMO novo. Atualmente está em expansão, com constantes atualizações e com o número de jogadores crescentes, chegando a mais de dezesseis milhões em Agosto de 2018 [27]. Foi lançado tanto para consoles quanto para computadores e seus jogadores jogam em um mundo compartilhado e possui um modelo de monetização baseado em mensalidades. Finalmente, suas mecânicas permitem que o jogador utilize apenas um personagem para uma série de estilos de jogo diferentes. Apesar dessas singularidades, o MMO *Final Fantasy XIV*) possui poucas pesquisas e nenhuma fez uso de um banco de dados extensivo.

Tal lacuna pode ser justificada pela falta de disponibilidade completa de dados do jogo. A empresa Square Enix (Desenvolvedora do *Final Fantasy XIV*) não fornece um

¹<https://github.com/LucasVF/ClusterAnalysisFFXIV>

acesso a um banco de dados oficial, nem uma aplicação de acesso, dificultando assim a análise do jogo. Entretanto, ela oferece um site que disponibiliza informações atuais sobre todos os personagens do jogo [131]. Utilizando requisições regulares a esse site, um grupo de indivíduos compôs um Banco de Dados e disponibilizou dados mensais sobre todos os personagens do jogo [27]. Logo, permitiu-se que muitas pesquisas pudessem analisar os jogadores de *Final Fantasy XIV*, desde que fosse superadas as limitações de não possuírem todos os atributos normalmente presentes em dados semelhantes e dispuserem apenas de informações coletadas em um momento de cada mês e não de todos os meses.

Supondo que o pesquisador ainda tenha interesse em realizar sua pesquisa com os dados deste jogo, ele terá o desafio de escolher qual técnica usará para analisar os dados que possui. Como visto no Capítulo 2, uma das técnicas populares em Mineração de Dados é a Clusterização, ou Agrupamento. Após esse agrupamento, cada grupo, ou *cluster*, é analisado para que conclusões, tiradas sobre o grupo, sejam generalizadas para seus membros. Entretanto, apesar de sua popularidade, não existe um consenso sobre o melhor método de clusterização, fato derivado da diversidade de abordagens que existem.

Visando contribuir para o aumento do domínio dos estudos em *Game Analytics*, assim como auxiliar na superação de desafios relacionados a limitações de dados e aplicação de algoritmos de clusterização, este trabalho propõe um estudo comparativo entre diferentes abordagens de clusterização para analisar os dados derivados do jogo *Final Fantasy XIV*. Tal comparação será feita utilizando métodos de avaliação de cluster presentes na literatura. Interpretações sobre resultados desses agrupamentos serão levadas em consideração conjuntamente.

Após a finalização deste trabalho, três principais contribuições serão feitas. A primeira envolve a introdução da Base de Dados disponíveis do *Final Fantasy XIV* para a comunidade científica e um incentivo para que mais pesquisas sejam realizadas, aumentando o horizonte explorado por cientistas. A segunda contribuição consiste no fornecimento de um embasamento para futuras pesquisas que irão aplicar técnicas de clusterização em dados de MMORPG com dados ausentes. A última principal contribuição do trabalho abrange o conhecimento que será adquirido dos jogadores de *Final Fantasy XIV* mediante os resultados dos algoritmos de clusterização.

6.2 Recursos

A seguir são elencados os diversos softwares, linguagem e bibliotecas utilizados nesta pesquisa:

- Software: MySQL Workbench
- Software: MySQL Server

- Linguagem: SQL
- Linguagem: Python 3
- Sistema Operacional: Windows 10 PRO
- Biblioteca: Scikit-learn [54]
- Biblioteca: Matplotlib [132]
- Biblioteca: seaborn: statistical data visualization [133]
- Biblioteca: pandas: powerful Python data analysis toolkit [134]
- Biblioteca: PyClustering [61]

A lista abaixo apresenta as configurações físicas do computador utilizado para realizar as análises deste trabalho:

- Processadores: 8 Núcleos x1200WB V6 INTEL XEON E3-1230 V6 3.5 GHZ 8 MB
- Memória RAM: 64GB DDR4
- Memória HD: SSD 120GB + 4TB SATA3
- Placa de Vídeo: Quadro Pascal p400 2GB 256 CUDA

6.3 Coleta de Dados

Os dados analisados nesta pesquisa são provenientes do jogo *Final Fantasy XIV*, agrupados mensalmente pelo *FFXIV Census* [27]. Neste trabalho serão utilizados como parâmetro os seguintes meses:

- 2017: Junho, Outubro
- 2018: Fevereiro, Junho, Outubro
- 2019: Fevereiro, Junho

A existência de um período padrão entre os meses permitiu uma organização do resultado. A escolha pela utilização dos meses listados foi feita buscando a menor cadência possível entre os meses disponíveis dentro de uma expansão, resultando em uma diferença de quatro meses entre um mês analisado e o próximo. Os meses originalmente disponíveis para análise eram:

- 2016: Julho - Dezembro

- 2017: Janeiro - Março, Maio - Junho, Setembro - Dezembro
- 2018: Janeiro - Novembro
- 2019: Fevereiro - Junho

Os arquivos *SQL* dos meses relacionados serão baixados e executados com o *MySQL Workbench* para criação do Banco de Dados com uma tabela de personagens para cada mês.

Com a Utilização da linguagem *Python* será desenvolvido um código que, quando executado, realizará a conexão com o Banco de Dados resultante e coletará os dados de cada personagem no mês requerido para aplicação da clusterização.

6.4 Pré-Processamento

Nesta seção serão abordados os processos e métodos executados para lidar com o pré-processamento de dados. Todos os tratamentos listados serão executados uma vez para cada mês analisado. Cada algoritmo de clusterização será executado utilizando o resultado desta etapa.

6.4.1 Valores Ausentes

Classificado como um processo de limpeza de dados, a verificação e correção de valores ausentes analisa cada objeto de dados para qualquer atributo que não esteja gravado.

Um algoritmo em Python será desenvolvido para se conectar com o Banco de Dados estudado e executará uma pesquisa em busca de atributos vazios ou nulos. O algoritmo finalizará informando se houve algum atributo vazio.

Em caso positivo, os atributos ausentes serão preenchidos manualmente no Banco de Dados com o valor mínimo relacionado ao domínio do atributo.

6.4.2 Conversão de Atributos

Processos que executam comparações de atributos possuem uma eficiência maior ao utilizar valores numéricos pois uma comparação dos mesmos é mais simples que a comparação de valores não-numéricos. Será realizada uma conversão de cada valor não-numérico contido no Banco de Dados coletados para análise. Apenas o nome do personagem não será convertido em um atributo numérico.

Os atributos originalmente binários serão convertidos no valores 0 para *false* e 1 para *true*. O atributo de gênero será convertido nessa mesma representação binária pois no jogo possuem apenas dois gêneros. O atributo de raça e *Grand Company* serão convertidos

utilizando a estrutura de dados contida em Python, dicionário. A partir de tal estrutura cada opção é convertido em um valor numérico. A representação do jogador membro de um clã de jogadores será convertido em uma representação binária (0,1) utilizando o atributo que referencia o nome do clã.

Tais conversões serão realizadas por um algoritmo que será desenvolvido neste trabalho, cuja execução ocorrerá após a coleta dos dados.

6.4.3 Agregação de Atributos

A agregação de atributos é feita para melhorar a eficiência do processo de mineração de dados, assim como tornar o resultado mais interpretável. Neste processo um grupo de atributos dos personagens resgatados do Banco de Dados são compilados em atributos mais gerais.

As Tabelas 6.2 e 6.1 mostram quais atributos foram compilados em quais atributos agregados. A conversão dos atributos relacionados aos levels de cada *class* e *job* em atributos agregados relacionados à experiência foram feitas através da soma da experiência necessária para atingir cada level informado. A conversão em experiência foi necessária pois a quantidade de experiência necessária depende do level atual e aumenta exponencialmente. Entretanto é importante ressaltar que novos *jobs*, lançados em expansões, são habilitados em níveis superiores a trinta, logo não foi necessária a mesma dedicação para chegar no mesmo nível que um *job* do jogo base. Mas a natureza exponencial da experiência necessária para avançar de nível configura essas experiências não conquistadas como menos de oito por cento da experiência total para o nível máximo na expansão, então um tratamento diferenciado não foi realizado.

Os atributos de *class* e *job* foram agregados de acordo com suas *roles*, devido ao fato de que cada *role* caracteriza um estilo de jogo diferente. Esse agrupamento de estilos de jogo é feito de acordo com os agrupamentos determinados pela Square Enix, desenvolvedora do jogo.

O atributo agregado *minimalCertainPaidTime* configura o limite inferior de mensalidades pagas pelo jogador, identificadas através da análise do banco de dados. O cálculo deste atributo agregado é feito analisando os atributos relacionados aos dias pagos, identificando a maior quantidade e dividindo por trinta, resultando em um número de mensalidades pagas. O atributo *PhysicalObject* visa identificar quantos itens físicos no mundo real o jogador comprou, somando cada atributo que identifica que o jogador adquiriu algum item.

Relacionando quantas tribos bestiais o jogador já atingiu o nível máximo de cooperação, o atributo *BeastTribe* soma a quantidade de atributos de tribos bestiais que identificam se ele alcançou ou não o nível máximo. Finalmente, o atributo *Marriage* rea-

Atributo	Atributo Agregado	Tratamento
kobold	BeastTribe	Somatório
sahagin		
amaljaa		
sylph		
moogle		
vanuvanu		
vath		
dideternalbond	Marriage	Atribui VERDADEIRO, caso um dos atributos seja VERDADEIRO
saweternalbond		

Tabela 6.1: Agregação de Atributos

liza a configuração na hipótese do personagem fazer parte de uma cerimônia de casamento virtual, seja como convidado ou participante.

Tais agregações serão realizadas pelo algoritmo em Python que será implementado neste trabalho. Esta execução será feita após a conversão de atributos.

6.4.4 Normalização

Técnicas de normalização são utilizadas para evitar a dependência da escolha das unidades de medidas. Essas técnicas são necessárias para casos nos quais as unidades de medida dos atributos afetem o resultado da análise [25].

A normalização envolve a transformação dos dados numéricos originais em uma faixa menor e comum para todos atributos, comumente sendo entre [0,1] ou [-1,1] [25].

A técnica aplicada neste trabalho é a *Min-Max Normalization*, técnica esta que preserve os relacionamentos entre os valores originais dos dados [25].

Tal técnica realiza uma transformação linear dos dados originais. Supondo que min_A e max_A são os valores máximos e mínimos de um atributo, A com n valores observados, $v_1, v_2, v_3 \dots v_n$. Então, o algoritmo irá mapear um valor, v_i de A , para um valor v'_i pertencente ao intervalo $[novo_min_A, novo_max_A]$. O calculo do valor de v'_i pode ser representado como [25]:

$$v'_i = \frac{v_i - min_A}{max_A - min_A} (novo_max_A - novo_min_A) + novo_min_A. \quad (6.1)$$

A implementação usada neste trabalho é a presente na biblioteca *Scikit-Learn*, com o nome de *MinMaxScaler* [54]. O intervalo que será usado será [0,1].

Atributo	Atributo Agregado	Tratamento
level_gladiator	TankExp	Conversão para Pontos de Experiência, seguida por somatório com membros do mesmo estilo de jogo.
level_marauder		
level_darkknight		
level_pugilist	PhysicalMeeleExp	
level_lancer		
level_rogue		
level_samurai		
level_archer	PhysicalRangerExp	
level_machinist		
level_astrologian	HealerExp	
level_conjurer		
level_thaumaturge	MagicRangerExp	
level_redmage		
level_carpenter	CraftExp	
level_blacksmith		
level_armorer		
level_goldsmith		
level_leatherworker		
level_weaver		
level_alchemist		
level_culinarian		
level_miner		GatherExp
level_botanist		
level_fisher		
p30days	minimalCertainPaidTime	Escolha do maior número de dias pagos e, sem seguida, realização da divisão por 30
p60days		
p90days		
p180days		
p270days		
p360days		
p450days		
p630days		
p960days		
arrartbook	PhysicalObject	Somatório de itens físicos
hwartbookone		
hwartbooktw		
hasencyclopedia		
beforemeteor		
beforethefall		
soundtrack		
moogleplush		
topazcarubuncleplush		
emeraldcarbuncleplush		

Tabela 6.2: Cont: Agregação de Atributos

6.4.5 Diminuição de Dimensão

A estrutura de um banco de dados real pode possuir dezenas de atributos, dentre eles a possibilidade de existir atributos irrelevantes ou redundantes é alta. Manter esses atributos irrelevantes pode ser prejudicial à análise, causando confusão para o algoritmo de mineração empregado. Uma solução para esse dilema é a remoção de tais atributos, entretanto essa tarefa realizada com cuidado pois a remoção de atributos relevantes pode resultar em análises de baixa qualidade [25].

O auxílio de um especialista no domínio pode ser bastante útil na escolha dos subconjuntos de atributos. Entretanto, tal atividade pode ser complexa e demorada, principalmente quando o comportamento dos dados for desconhecido [25].

A solução utilizada neste trabalho envolve uma matriz de correlação utilizando a Correlação de Pearson com os atributos resultantes dos outros processos de pre-processamento [135].

O mapa de correlação resultante irá relacionar todas variáveis entre si, calculando o coeficiente de correlação para cada par. O valor deste coeficiente varia no intervalo $[-1,1]$, onde [135]:

- Valores próximos de 1, possuem alta correlação positiva.
- Valores próximos de 0, possuem baixa correlação.
- Valores próximos de -1, possuem alta correlação negativa.

Este coeficiente pode ser calculado através da formula a seguir, considerando que X e Y , são os conjuntos de valores de dois atributos, $cov(X, Y)$ é a covariância entre esse dois conjuntos e a variância populacional desses conjuntos são representados por $var(Y)$ e $var(X)$:

$$p = \frac{cov(X, Y)}{\sqrt{var(X) * var(Y)}} \quad (6.2)$$

Os atributos que possuírem baixa correlação com todas outras variáveis serão retirados da análise de clusterização. Ou seja, aqueles atributos que não possuírem correlações fora do intervalo de $[-0.5,0.5]$ serão ignorados. Este método de filtragem foi escolhido pela simples implementação que permite o tratamento de muitos atributos sem maiores custos e pela sua boa performance com Análises de Dados Exploratória [135].

A implementação que será utilizada fez uso das bibliotecas em Python de *pandas*, *matplotlib* [132] e *seaborn*.

6.4.6 Diminuição do Tamanho do Domínio

A análise de quantidades expressivas de dados pode gerar dificuldades para o *hardware* que for realizar esse experimento. Tais dificuldades derivam da necessidade de grandes recursos

computacionais para realizar as operações com o volume massivo de dados. Entretanto os recursos necessários para a realização de todos algoritmos de clusterização na totalidade de personagens contidos na base de cada mês analisado podem não estar disponíveis para este trabalho. Logo a análise deverá ser feita com uma amostra do domínio estudado, caso os recursos computacionais disponíveis não sejam capazes de analisar o conjunto completo de personagens coletados.

A seleção dos membros da amostra será feita randomicamente utilizando de um comando presente no *SQL* para recuperar a amostra de forma arbitrária. Essa escolha é feita para evitar um possível resultado tendencioso, conseqüentemente invalidando a generalização das conclusões atingidas. Decorrente dessa seleção, os personagens analisados em um determinado mês provavelmente não irão coincidir com os personagens analisados em outro mês, inviabilizando uma análise temporal de personagens neste estudo.

Para determinar o número de personagens presentes na amostra foi utilizada uma ferramenta para cálculo de proporção de população, [136]. Essa ferramenta, presente no site *Select Statiscal Services* [136], utiliza a seguinte formula para o calculo do tamanho da amostra:

$$n = \frac{N * X}{X + N - 1} \quad (6.3)$$

, onde

$$X = \frac{Z_{a/2}^2 * p * (1 - p)}{e^2} \quad (6.4)$$

As variáveis presentes nas formulas acima possuem as seguintes definições [137]: n representa o tamanho da amostra, N é o tamanho da população, e é a margem de erro, p é a proporção da amostra, $Z_{a/2}$ é o valor crítico da distribuição Normal em $a/2$ e a é o valor de 1 menos o nível de confiança.

Neste trabalho os valores de cada variável foram escolhidos como:

- $N = 16000000$
- $a = 1 - 99\%$
- $e = 1\%$
- $p = 0,5$

O número escolhido para representar a população é 16000000 pois esse é o menor valor arredondado que é superior à população de qualquer mês estudado. A proporção foi estimada como 50% pela falta de conhecimento sobre o resultado. Finalmente, a margem de erro e níveis de confiança foram escolhidos como 1% e 99% para produção de resultados mais confiáveis.

Utilizando esses valores, o resultado deste cálculo é 16571. Conseqüentemente, o tamanho da amostra foi escolhido como 17000, valor este que é superior ao resultado obtido.

6.4.7 Escolhendo o número de Clusters

Alguns algoritmos podem necessitar o fornecimento do número de clusters resultantes antes da execução do algoritmo, o *k-means* é um exemplo desse algoritmos. Entretanto uma escolha errônea pode ocasionar em um resultado tendencioso, que conseqüentemente gerará resultados não confiáveis.

Existem muitas formas na literatura para o alcance do número "correto" de clusters, e, como clusterização, não existe o melhor meio para tal decisão [25]. A abordagem escolhida para este trabalho é simples, popular e eficaz, denominada como o Método do Cotovelo, ou *Elbow Method*.

O Método do Cotovelo sugere que o número ideal de clusters é o maior número possível antes que o aumento de agrupamentos não provoque um redução significativa na soma da variância dentro de cada cluster. Essa sugestão é fundamentada no fato de que apesar de mais clusters permitirem a captura de grupos mais refinados de objetos de dados, que são semelhantes entre si, o efeito desse aumento de clusters na soma das variâncias será cada vez menos significativo. Esse decréscimo em verdade pode ser observado ao se dividir um agrupamento coeso em dois e ambos resultados serão semelhantes entre si devido a coesividade do cluster original [25].

Resumindo, a heurística deste método para selecionar o número certo de clusters é usar o ponto de viragem na curva da soma das variâncias dentro do cluster em relação ao número de clusters [25].

A implementação escolhida exibe um gráfico com a distorção causada pelo método de clusterização *K-means* utilizando diferentes valores de k , sendo k o número de clusters. O valor da soma da variância de clusters é representado pela distorção causada pelo método [138]. A Figura 6.1 exibe um gráfico resultante da implementação descrita. Nesse gráfico o valor sugerido para clusterização é três, pois qualquer número acima provoca mudanças pouco significativas.

6.5 Validação de Clusters

Esta seção tratará sobre os métodos que serão aplicados para avaliação e comparação dos resultados da clusterização. Esses métodos são: Estatística Hopkins, Coeficiente de Silhueta, Índice de Calinski-Harabasz e Índice de Davies-Bouldin.

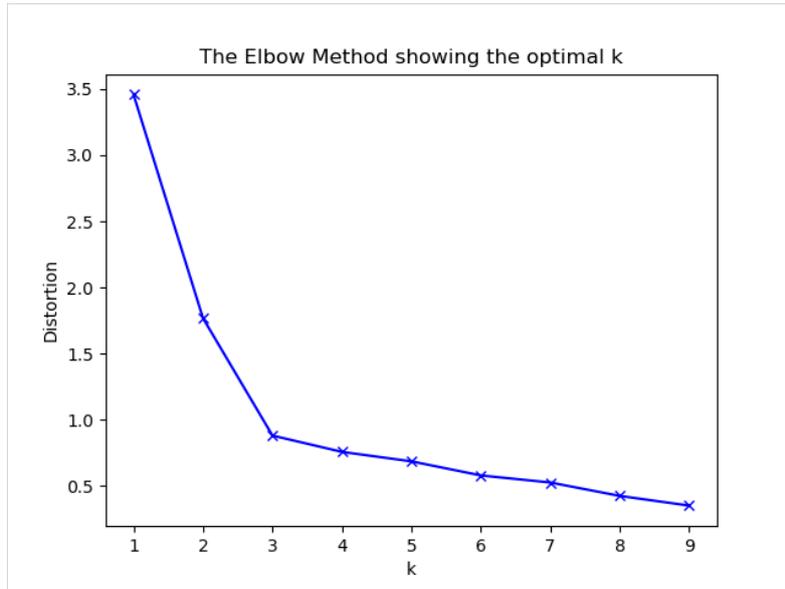


Figura 6.1: Método do Cotovelo [138]

6.5.1 Estatística Hopkins

Considerando o não conhecimento prévio sobre a distribuição dos dados na amostra estudada, existe a possibilidade de uma distribuição homogênea. Tais distribuições invalidam qualquer resultado obtido por uma eventual clusterização. Essa invalidação ocorre pois em uma distribuição randômica, qualquer informação deduzida do resultado obtido não permitirá a indução de novas informações confiáveis. Enquanto que, na distribuição homogênea, as divisões criadas pelos clusters serão irrelevantes.

Existe na literatura um método que permite avaliar a tendência para clusterização, sendo este a Estatística de Hopkins [25]. Tal estatística estatística especial avalia o quão especialmente randômico uma variável distribuída em um espaço é [25]. Dado um conjunto de dados $X = \{x_i | i = 1 \text{ até } n\}$ e seja $Y = \{y_i | i = 1 \text{ até } m\}$ um conjunto de m amostras coletadas randomicamente com o mesmo domínio, sendo $m \ll n$. O cálculo desta estatística é feito seguindo os seguintes passos [139]:

1. Defina u_j como sendo a menor distância entre y_j e seu vizinho mais próximo em X .
2. Defina w_j como sendo a menor distância entre um elemento randomicamente selecionado de X e seu vizinho mais próximo.
3. Calcule a Estatística de Hopkins usando a seguinte formula:

$$H = \frac{\sum_{j=1}^m u_j}{\sum_{j=1}^m u_j + \sum_{j=1}^m w_j} \quad (6.5)$$

O valor resultante será contido no intervalo de 0 até 1. Baseada na hipótese nula, o resultado 0 e valores próximos indicam uma distribuição uniforme dos dados. Valores próximos à 0.5 sugerem uma distribuição randômica, enquanto valores próximos de 1 apontam uma alta tendência à presença de clusters.

A implementação que será usada neste trabalho foi retirada da internet². Tal implementação será aplicada em cada mês analisado.

6.5.2 Coeficiente de Silhueta

A ausência de informações externas para validação de resultados da clusterização previne a utilização de meios extrínsecos de validação, restando apenas métodos intrínsecos. Métodos Intrínsecos são utilizados para avaliar o quão separados são os clusters entre si e o quão próximos são os elementos dentro do mesmo cluster.

A medida que será utilizada para comparação neste trabalho é o Coeficiente da Silhueta [25]. Para um conjunto de dados, D , com n objetos, suponha que o mesmo está dividido em k clusters, $C_1 \dots C_k$. Para cada objeto de dado $o \in D$, $a(o)$ é definido como a distância média entre o e todos outros objetos pertencentes ao cluster ao qual o pertence, este valor representa a compacidade do cluster. Quanto menor o valor de $a(o)$, mais compacto será o cluster. Outra importante definição é $b(o)$, média da distância mínima de o para todos clusters que não possuem o , esse valor representa o quão separado o objeto está dos clusters. Quanto maior $b(o)$, maior será a separação entre clusters [25].

Com tais definições, o Coeficiente de Silhueta do objeto de dados o é calculado como:

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}}, \quad (6.6)$$

onde

$$a(o) = \frac{\sum_{o' \in C_i, o' \neq o} \text{dist}\{o, o'\}}{|C_i| - 1} \quad (6.7)$$

e

$$b(o) = \min_{C_j: 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} \text{dist}\{o, o'\}}{|C_j|} \right\} \quad (6.8)$$

Consequentemente os valores possíveis para este coeficiente são entre -1 e 1. Valores próximos ao 1, evidenciam que o cluster contendo o é compacto e separado dos demais, caracterizando o melhor dos casos, ao contrário do que é evidenciado se o valor resultante for negativo, onde o objeto o é mais próximos de outros clusters que objetos no próprio cluster [25].

²<https://matevzkunaver.wordpress.com/2017/06/20/hopkins-test-for-cluster-tendency/>

O Coeficiente de Silhueta usado para comparação de métodos de clusterização é calculado usando a média dos coeficientes de cada cluster. O coeficiente de cada cluster, por sua vez, é calculado pela média dos Coeficientes de Silhueta de cada um de seus membros [25].

A implementação que será utilizada para o cálculo desta medida neste trabalho está contida na biblioteca de Aprendizado de Máquina Scikit-learn [54].

6.5.3 Índice de Calinski-Harabasz

Além do Coeficiente de Silhueta, este estudo aplicará o Índice de Calinski-Harabasz [140] para realização da avaliação intrínseca dos clusters resultantes. Onde um valor superior deste indicador indica clusters bem definidos [54].

O cálculo deste índice é realizado utilizando as definições de soma dos quadrados entre clusters, B_c , e dentro deles W_c . Ambas somas são calculadas utilizando as seguintes fórmulas [140], onde K é o número de clusters resultantes, x são os elementos de dados, C são os clusters e o valor de $w_{k,i}$ é 1 quando $x_i \in C_k$ e 0 caso contrário:

$$B_c = \sum_{k=1}^K |C_k| |\overline{C_k} - \bar{x}|^2 \quad (6.9)$$

e

$$W_c = \sum_{k=1}^K \sum_{i=1}^N w_{k,i} |x_i - \overline{C_k}|^2 \quad (6.10)$$

Finalizando, ao utilizar ambas definições acima na fórmula abaixo [140], obtêm-se o valor do índice de Calinski-Harabasz, cuja implementação está contida em uma das bibliotecas que serão utilizadas [54]:

$$CH(k) = \frac{B_c(k)}{(k-1)} / \frac{W_c(k)}{(n-1)} \quad (6.11)$$

6.5.4 Índice de Davies-Bouldin

Outro índice disponível para avaliação do modelo de clusters resultantes, sem a utilização de uma verdade fundamental, é o Índice de Davies-Bouldin [141, 54]. Este índice, contido na biblioteca utilizada [54], é definido como a média de similaridade entre cada cluster C_i , para $i = 1, \dots, k$ e seu cluster mais similar C_j .

A medida de similaridade utilizada, R_{ij} , pode ser definida como:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}, \quad (6.12)$$

onde s_i é a distância média entre cada ponto do cluster i do centro daquele cluster e d_{ij} é a distância entre os centros dos clusters i e j .

Finalmente, utilizando a medida acima, este índice é calculado através da fórmula [141, 54]:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (6.13)$$

Clusters bem separados resultam em valores próximos de zero. Quanto mais próximo do valor zero, melhor será a separação entre clusters, indicando assim uma melhor partição de dados [54].

6.6 Atribuição de Perfil

Após a aplicação do algoritmo de clusterização e os agrupamentos resultantes avaliados, resta apenas uma etapa para finalização do processo de mineração de dados, a inferência de informações do resultado. Se clusters bem definidos foram gerados e, todavia, não foi possível inferir novas informações do resultado sequer confirmar hipóteses, então a prática pode ter sido realizada em vão. Contudo tal fracasso não invalida o resultado das avaliações das clusterizações. A união de outras informações ou a interpretação de outros indivíduos podem obter o sucesso desejado.

Este trabalho pretende analisar as propriedades dos clusters resultantes a fim de se buscar informações sobre os perfis de jogador que existem em *Final Fantasy XIV*. Entretanto, se faz necessário ressaltar que, como mencionado anteriormente, se tais perfis não forem definidos para um determinado algoritmo de clusterização, tal fato não invalida o algoritmo.

As propriedades listadas e consideradas para atribuição de perfis para cada cluster resultante são compostas pelas seguintes estatísticas descritivas:

- Quantidade de Membros;
- Média de valores;
- Desvio Padrão;
- Valor Mínimo Encontrado;
- 1º, 2º e 3º Quartil;
- Valor Máximo Encontrado.

Capítulo 7

Resultados

Este capítulo consistirá no relato e descrição dos resultados obtidos após a realização de todos os passos detalhados no capítulo anterior.

7.1 Pré-Processamento

Esta seção destacará as etapas realizadas durante o pré-processamento dos dados provenientes do *Final Fantasy XIV*, com foco nos procedimentos cujos resultados provocaram, ou poderiam provocar, decisões sobre o prosseguimento do estudo. Por exemplo, a verificação de dados ausentes poderia ocasionar em decisões sobre como tratar o dado ausente e a diminuição de dimensão gerará uma decisão sobre quais atributos manter no domínio.

O seguintes processos ocorreram como esperado e não necessitaram de deliberações:

- Conversão de Atributos
- Agregação de Atributos
- Normalização

7.1.1 Diminuição do Tamanho do Domínio

O algoritmo desenvolvido para realização de todas etapas do pré-processamento foi executado no domínio total de personagens contidos no mês menos populoso. Entretanto, o programa foi incapaz de finalizar pois a quantidade de memória necessária não estava disponível. Logo este trabalho realizou a clusterização na totalidade de dezessete mil personagens escolhidos aleatoriamente em cada mês analisado. A escolha dos personagens é feita uma vez para cada mês, logo a repetição de personagens é possível, mas improvável pela extensa quantidade disponível.

7.1.2 Verificação de Dados Ausentes

A primeira etapa realizada consiste na verificação de atributos ausentes contidos no banco de dados de cada mês. Um comando *SQL* foi executado para cada mês procurando a quantidade de personagens que possuíam valores nulos ou vazios em algum atributo. A Figura 7.1 exibe o retorno desta etapa executada sobre os meses de 2018, tal retorno foi o mesmo para os meses de 2017 e 2019.

```
=====Análise para Mes: 2 Ano: 18
Hora de Inicio = 17:03:29
Conexão MySQL esta aberta
=====Verificando existência de Valores Ausentes
Hora da Finalização da Checagem = 17:04:01
Resultado: Não foi encontrado Valores Ausentes
=====Valores Ausentes Verificados
=====Análise para Mes: 6 Ano: 18
Hora de Inicio = 16:57:08
Conexão MySQL esta aberta
=====Verificando existência de Valores Ausentes
Hora da Finalização da Checagem = 16:57:39
Resultado: Não foi encontrado Valores Ausentes
=====Valores Ausentes Verificados
=====Análise para Mes: 10 Ano: 18
Hora de Inicio = 17:01:54
Conexão MySQL esta aberta
=====Verificando existência de Valores Ausentes
Hora da Finalização da Checagem = 17:02:25
Resultado: Não foi encontrado Valores Ausentes
=====Valores Ausentes Verificados
```

Figura 7.1: Execução da verificação por valores ausentes nos meses analisados de 2018

7.1.3 Diminuição de Dimensão

A execução desta etapa produziu um mapa de calor com as correlações entre atributos para cada mês. As Figuras 7.2, 7.3 e 7.4 exibem os mapas de calor resultantes deste processo. Ao observar o resultado, os seguintes atributos foram retirados da análise devido à pouca correlação com outras variáveis:

- *race*: atributo que define a raça do jogador;
- *gender*: atributo que define o gênero do jogador;
- *realm*: atributo que define o local do servidor onde o jogador joga;
- *grand_company*: atributo que define a *Grand Company* na qual o jogador faz parte;

- *legacy_player*: atributo que define se o jogador jogou a versão lançada em 2010;
- *PreArr*: atributo que define se o jogador fez a compra do jogo prévia ao lançamento da segunda versão do jogo;
- *PObject*: atributo que define se o jogador comprou um item relacionado ao jogo no mundo real.

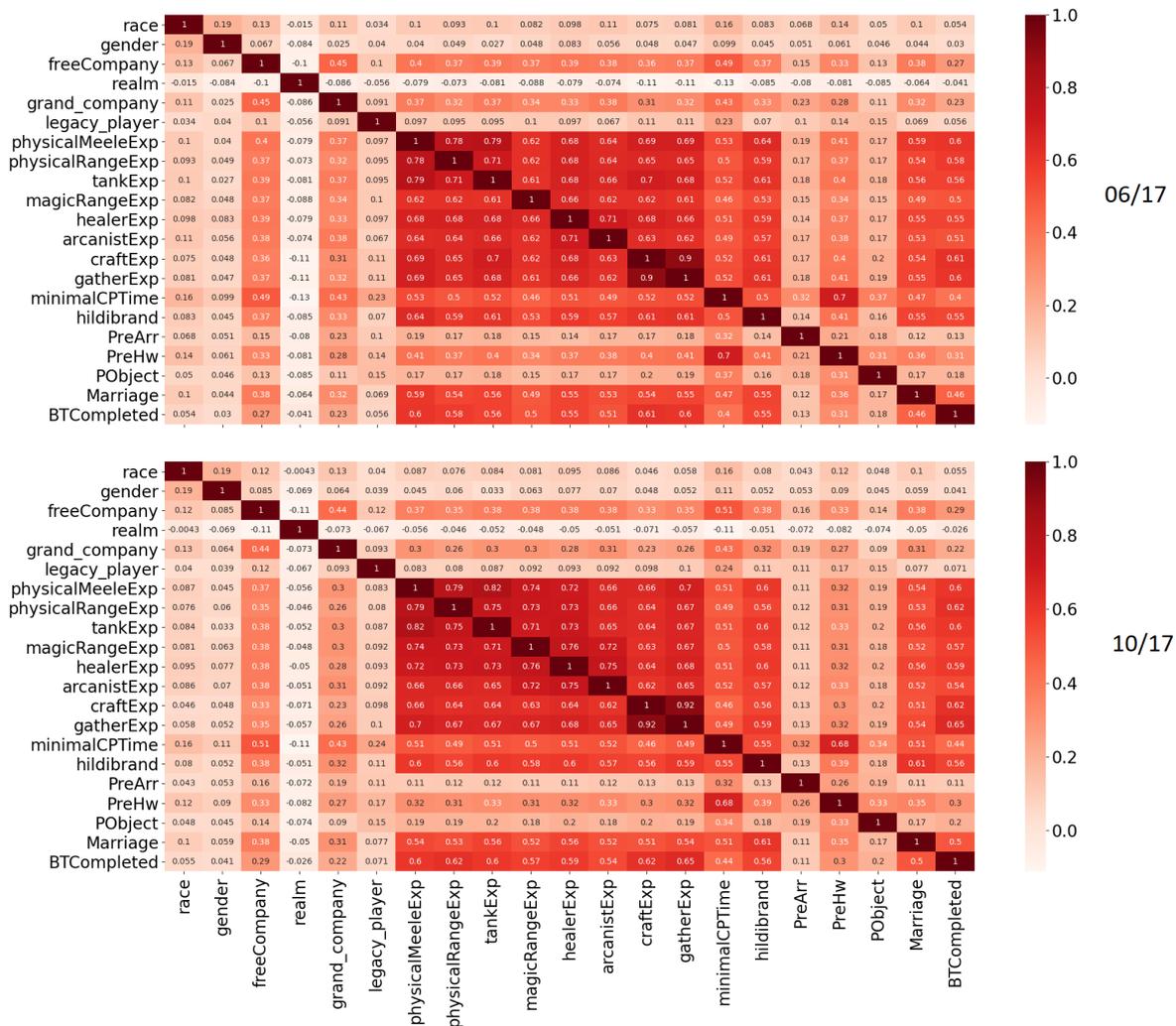


Figura 7.2: Mapa de Calor com as correlações entre atributos nos meses analisados de 2017

Após a remoção dos atributos acima, o jogador será representado pelos atributos presentes na Tabela 7.1.

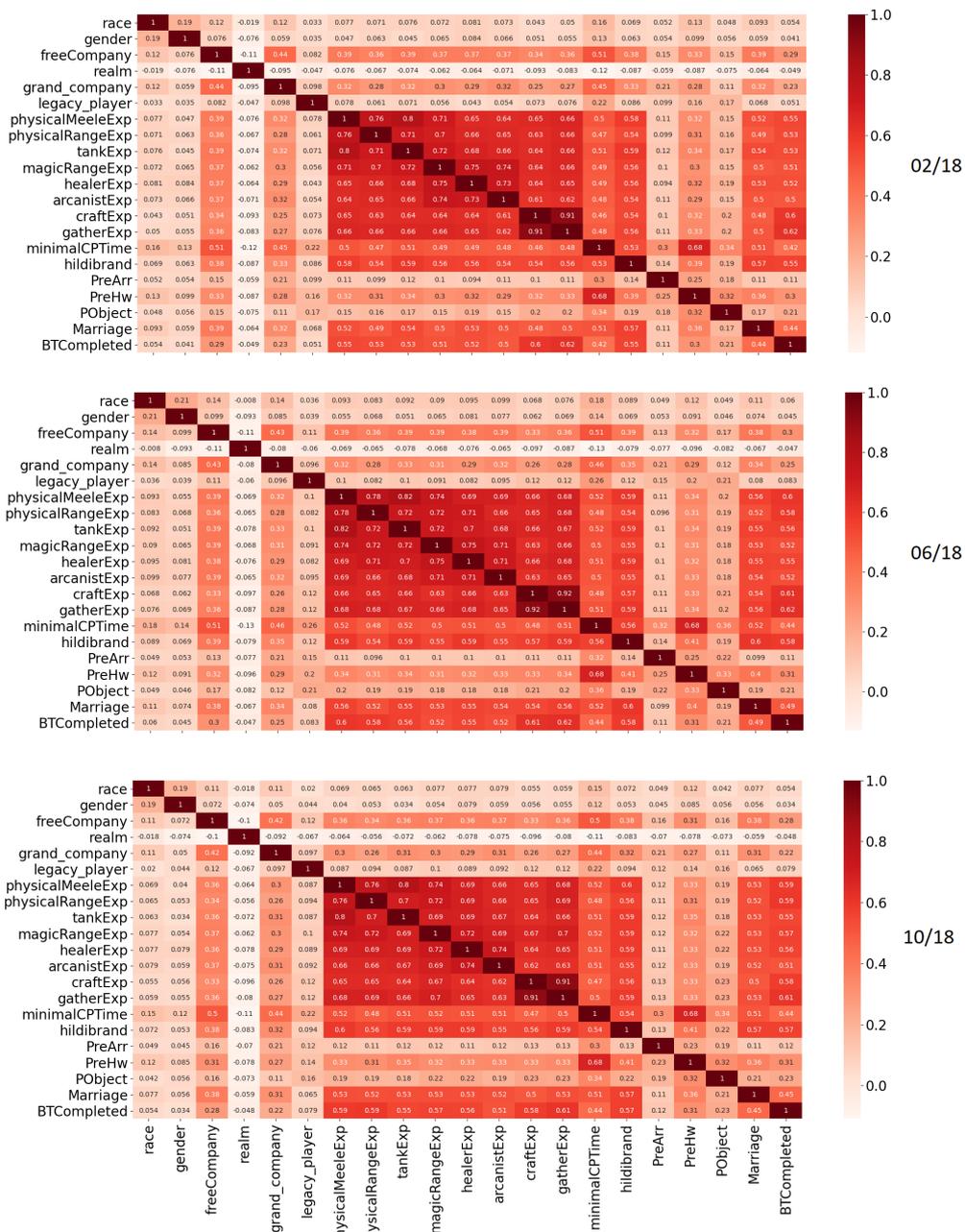


Figura 7.3: Mapa de Calor com as correlações entre atributos nos meses analisados de 2018

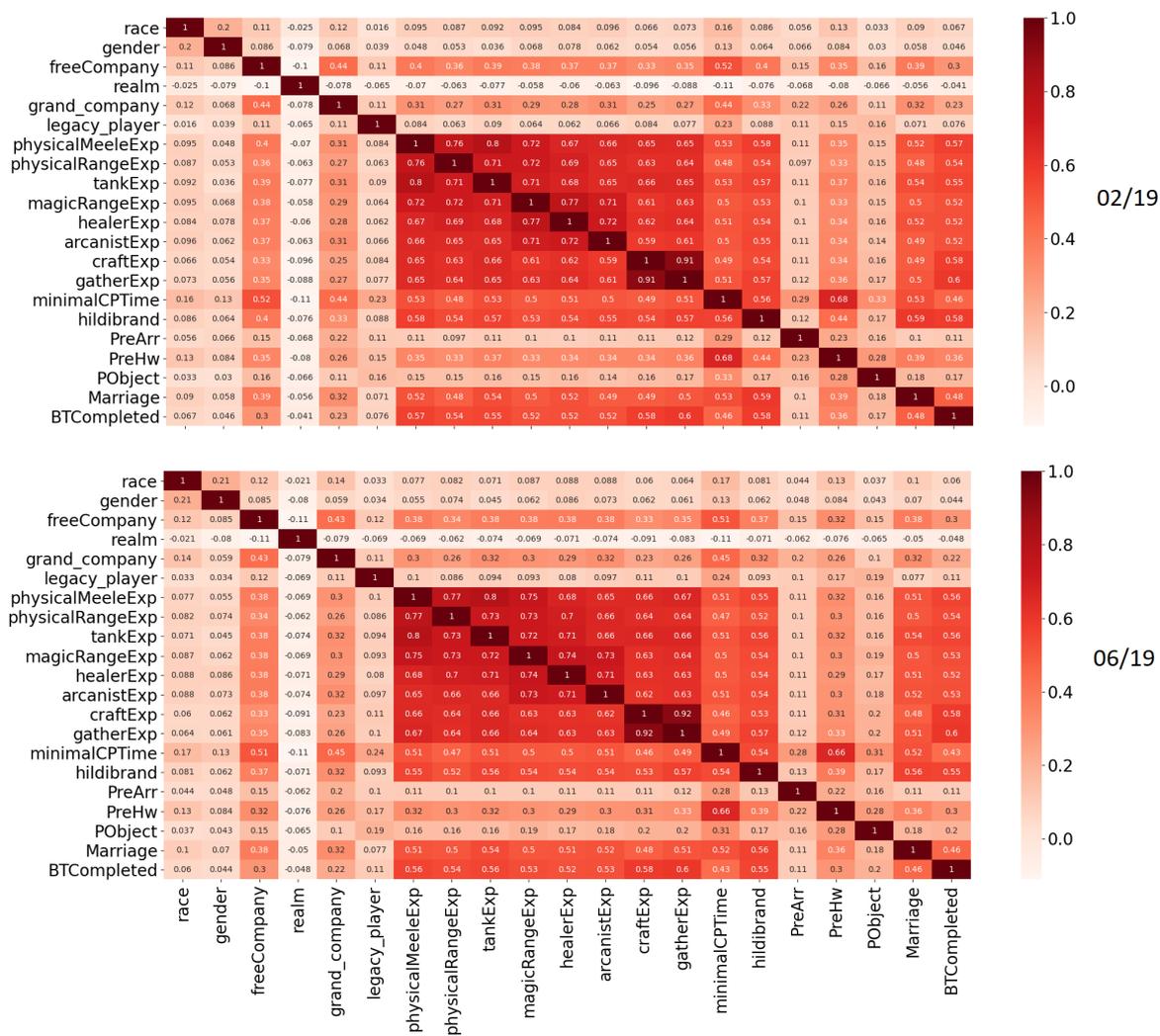


Figura 7.4: Mapa de Calor com as correlações entre atributos nos meses analisados de 2019

Atributo	Descrição
Free Company	Identificador se o jogador pertence a um grupo de jogadores
Physical Melee	Soma das experiências de jobs que atacam fisicamente corpo a corpo
Physical Ranged	Soma das experiências de jobs que atacam fisicamente a distância
Tank	Soma das experiências de jobs que focam em defender o time
Magic Ranged	Soma das experiências de jobs que atacam com magia a distância
Healer	Soma das experiências de jobs que focam em curar aliados
Arcanist	Soma das experiências do job Arcanist que pode curar ou atacar com magia a distância
Crafter	Soma das experiências dos jobs que focam na confecção de itens
Gather	Soma das experiências dos jobs que focam na coleta de materiais
Limite Inferior de Meses Pagos	Mínimo valor de mensalidades que o jogador, com certeza, pagou
Pre Ordem Heavensward	Identificador se o jogador fez a compra prévia da expansão Heavensward antes de seu lançamento
Hildibrand	Identificador se o jogador fez todas missões contidas no jogo base da história cômica de Hildibrand
Casamento	Identificador se o jogador casou ou fez parte de um casamento virtual
Tribos Bestiais Completas	Identificador de quantas tribos bestiais o jogador conseguiu alcançar o nível máximo

Tabela 7.1: Lista dos atributos do jogador a serem considerados na clusterização

7.1.4 Número de Clusters

A aplicação do método do cotovelo fez uso dos gráficos exibidos nas Figuras 7.5, 7.6 e 7.7. A conclusão alcançada ao se analisar tais gráficos foi a utilização de três clusters como valor de entrada para os algoritmos que requerem tal dado.

7.2 Resultados dos Clusters

Após a execução do processo de pré-processamento descrito anteriormente, foi realizado a aplicação dos algoritmos de clusterização.

Esta seção exibirá a descrição dos clusters resultantes com base nos valores dos atributos de seus membros. Para fins de comparação, a figura que exhibe a descrição do domínio clusterizado de cada um dos sete meses esta presente no Anexo I.

7.2.1 K-means

O primeiro algoritmo aplicado no domínio foi o *K-means*. Como definido pelo valor de entrada, cada aplicação gerou três clusters resultantes. Cada cluster resultante está

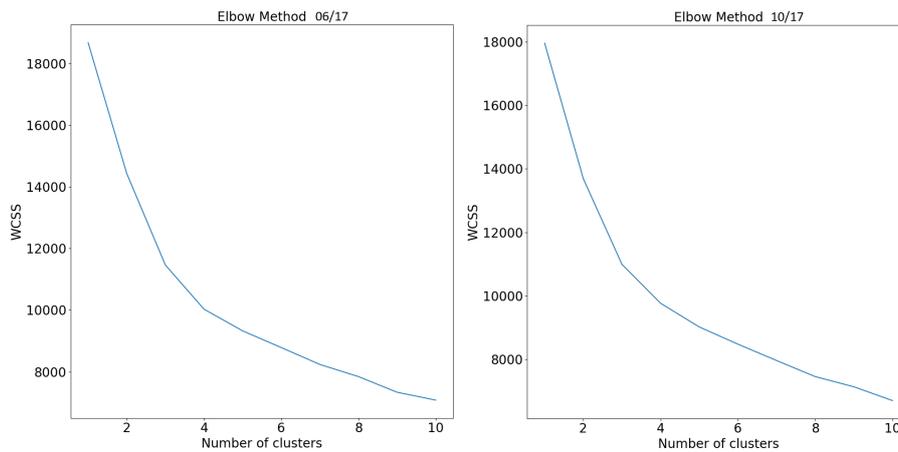


Figura 7.5: Método do Cotovelo nos meses analisados de 2017

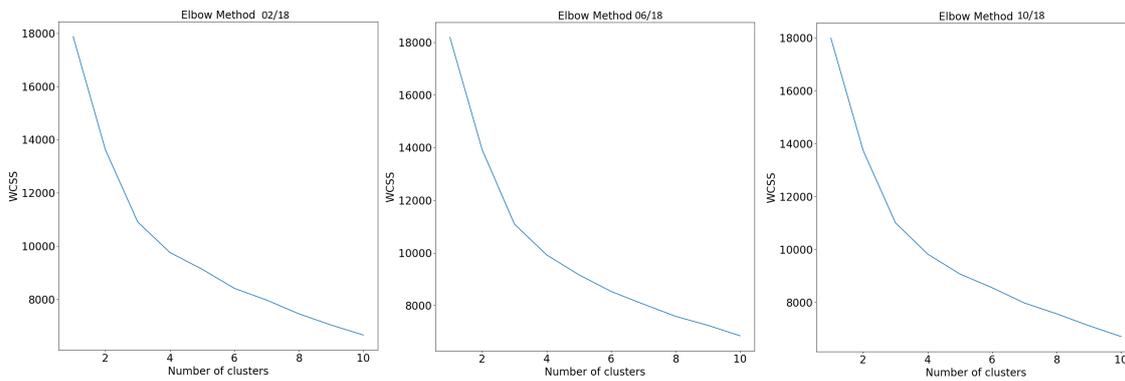


Figura 7.6: Método do Cotovelo nos meses analisados de 2018

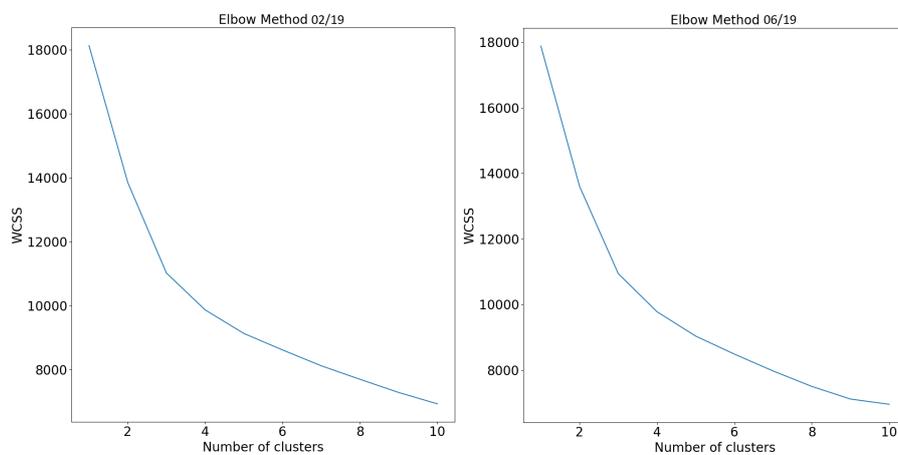


Figura 7.7: Método do Cotovelo nos meses analisados de 2019

descrito nas figuras presentes no Anexo I.

Médias

A semelhança das médias dos atributos dos clusters resultantes permite a identificação entre resultados. Para exemplificar essa afirmação, observe o Cluster 1 exibido na Figura I.3, este que possui médias e população semelhante ao Cluster 1 da Figura I.6 e do Cluster 2 da Figura I.5.

Esse resultado evidencia que o algoritmo produziu os mesmo clusters em todas suas aplicações, entretanto a ordem não foi mantida. Caso os clusters resultantes não fossem tão diferentes, a tarefa de identificação seria mais complexa.

Utilizando a primeira execução do algoritmo no mês de Junho de 2017, Figura I.2, como base, foi criado um gráfico para ilustrar as médias de cada atributo, relacionando-as com seus respectivos clusters. Tal gráfico está exibido na Figura 7.8. Para melhor compreensão, cada média foi normalizada para uma porcentagem em relação ao seu valor máximo possível. O gráfico evidencia a tendência de que as maiores médias dos atributos estudados estão presentes no clusters com menor representatividade em relação à população total.

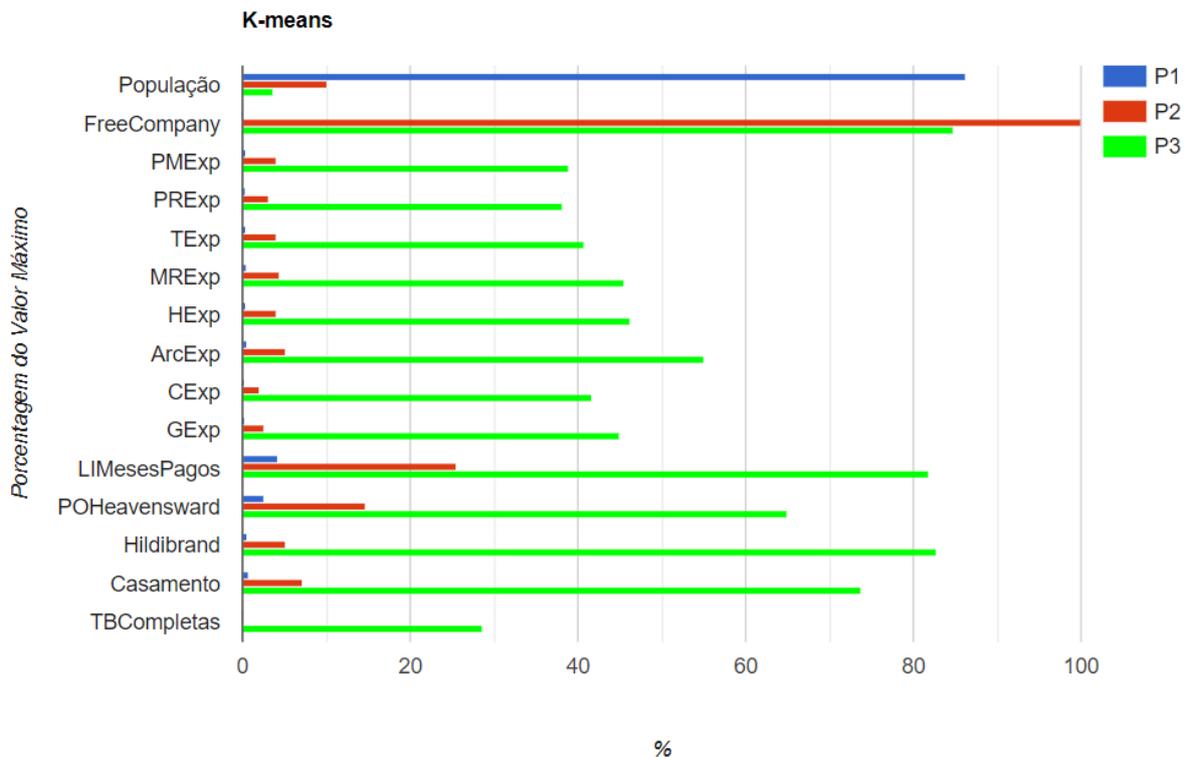


Figura 7.8: Porcentagens média de cada Cluster Resultante do K-means

7.2.2 WARD

O segundo algoritmo aplicado foi o *WARD*. Igualmente ao algoritmo anterior, a execução gerou três clusters para cada mês. A descrição estatística de cada cluster gerado está presentes nas figuras presentes no Anexo I.

Médias

A identificação de clusters realizada no algoritmo anterior pode ser replicada neste resultado. Entretanto, o Cluster 1 identificado na aplicação do mês Junho de 2017, na Figura I.9, permanece como Cluster 1 em cada aplicação subsequente, característica observada nos Clusters 2 e 3 do mesmo modo.

Esse caráter consistente de geração de cluster pode ser aproveitado em análises sequenciais cuja distância temporal de bases sejam distantes a ponto de dificultar a identificação de cada cluster formado em relação a um estado passado.

Assim como realizado com o algoritmo *K-means*, a Figura 7.9 mostra as porcentagens relativas as médias dos clusters identificados no resultado do algoritmo *WARD*. Apesar da semelhança entre médias, o valores dos clusters resultantes não são iguais, os valores de avaliação da clusterização identificarão quais clusters resultantes são mais próximos da realidade.

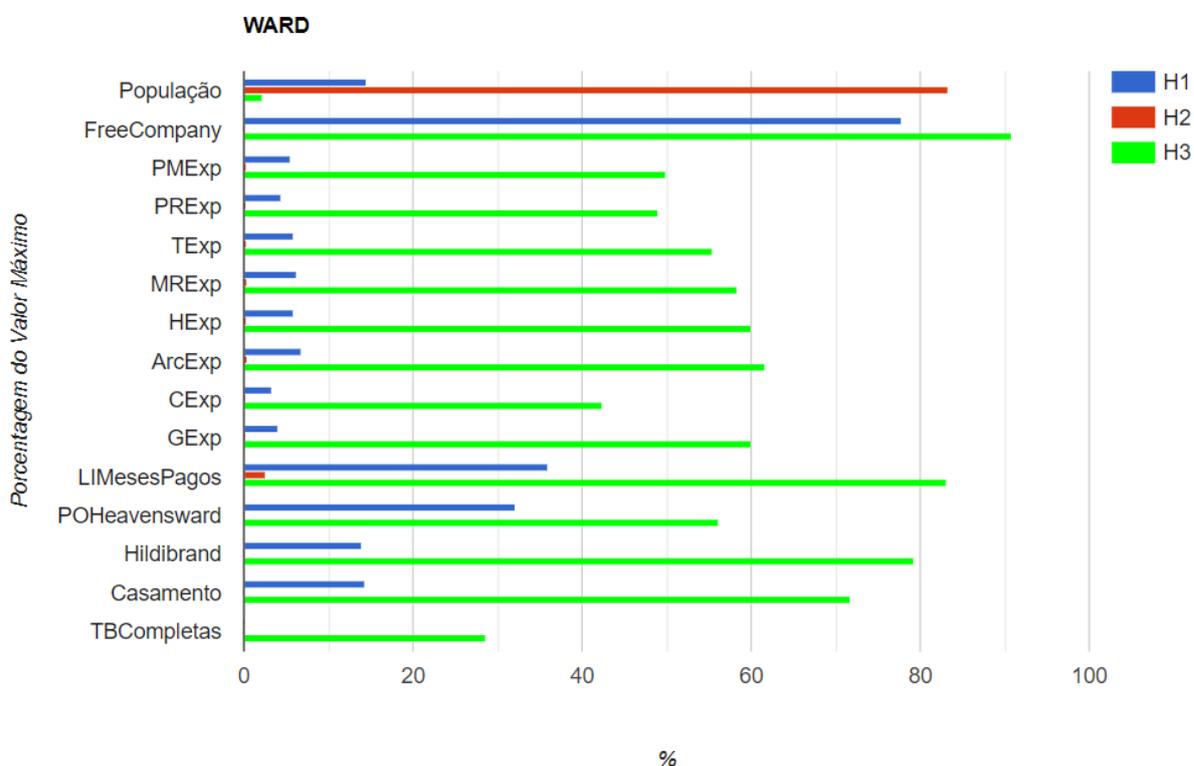


Figura 7.9: Porcentagens média de cada Cluster Resultante do WARD

7.2.3 DBSCAN

Após a execução da clusterização hierárquica, o método aplicado em sequência foi o baseado em densidade. Diferentemente dos algoritmos anteriores, o *DBSCAN* não requer um número de clusters como entrada. Este foi o algoritmo que gerou mais clusters para cada uma de suas aplicações. Os clusters gerados estão descritos nas figuras presentes no Anexo I.

Médias

Como o algoritmo não é tendencioso quanto ao número de clusters que serão formados em seu resultado esse número variou ao longo das aplicações. Devido à grande quantidade de clusters formados, a identificação de um cluster resultante de um mês em outro mês foi dificultada. Apenas os Clusters 1, 2, 3, 4 e 7 do mês Junho de 2017 foram relacionados em todos os meses. Fato que dificultou ainda mais a análise dos clusters foi que a característica determinística do algoritmo em conjunto com a seleção randômica de personagens não possibilitou que o algoritmo gerasse os clusters na mesma ordem.

Focando nos Clusters que foram identificados em todos os meses, a Figura 7.10 exibe a porcentagem das médias dos atributos desses clusters. Apesar de não ser consistente em seu resultado, esse algoritmo identificou, com frequência, clusters não identificados por outros algoritmos até agora.

7.2.4 BANG

O algoritmo *BANG* foi executado depois da aplicação do algoritmo baseado em densidade. Semelhantemente ao algoritmo anterior, a definição de um número de cluster não é requisitada. Entretanto, ao contrário do resultado obtido com o *DBSCAN*, a finalização deste algoritmo não resultou em múltiplos clusters. O algoritmo não conseguiu identificar diferentes clusters em nenhum dos meses aplicados. Uma hipótese para tal resultado é a dificuldade em gerar grades para domínios com muitas dimensões. O resultado está evidenciado na figura presentes no Anexo I.

7.2.5 SOM

A clusterização aplicada em seguida representou a metodologia *Model-Based*. O algoritmo *SOM* resultou nos clusters descritos nas figuras presentes no Anexo I. Assim como outros algoritmos, o *SOM* necessitou de um número de clusters como entrada, logo sua execução criou três clusters.

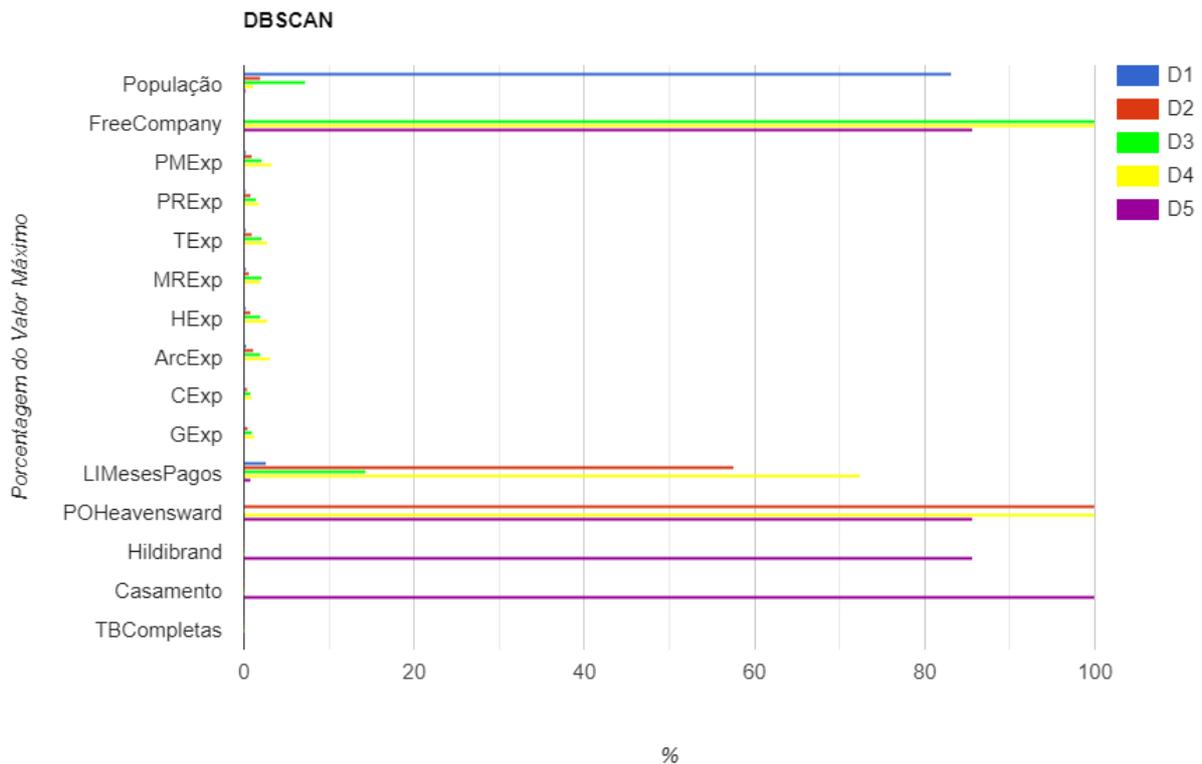


Figura 7.10: Porcentagens média de cada Cluster Resultante do DBSCAN

Médias

Ao observar os clusters resultantes, é notável que as médias do primeiro cluster formado no primeiro mês são muito semelhantes às médias do primeiro cluster de cada mês. Essa consistência é persistida em todos clusters formados por esse algoritmo.

Seguindo o processo executado com os algoritmos anteriores, o gráfico contido na Figura 7.11 descreve as porcentagens das médias dos atributos de cada cluster identificado no resultado.

Se compararmos o resultado deste algoritmo com o resultado do algoritmo *K-means*, a diferença entre resultados é a menor entre qualquer outra comparação de clusters resultantes. Ou seja, ambas clusterizações identificaram os mesmos clusters, mas este algoritmo produziu resultados com maior consistência.

7.2.6 Spectral

A execução do algoritmo *Spectral* foi realizada após a metodologia baseada em modelo. Os clusters resultantes estão descritos nas figuras presentes no Anexo I. Este algoritmo faz parte do conjunto de algoritmos que foram iniciados com o número de cluster pré-definidos. Conseqüentemente, o resultado de cada mês contém três clusters.

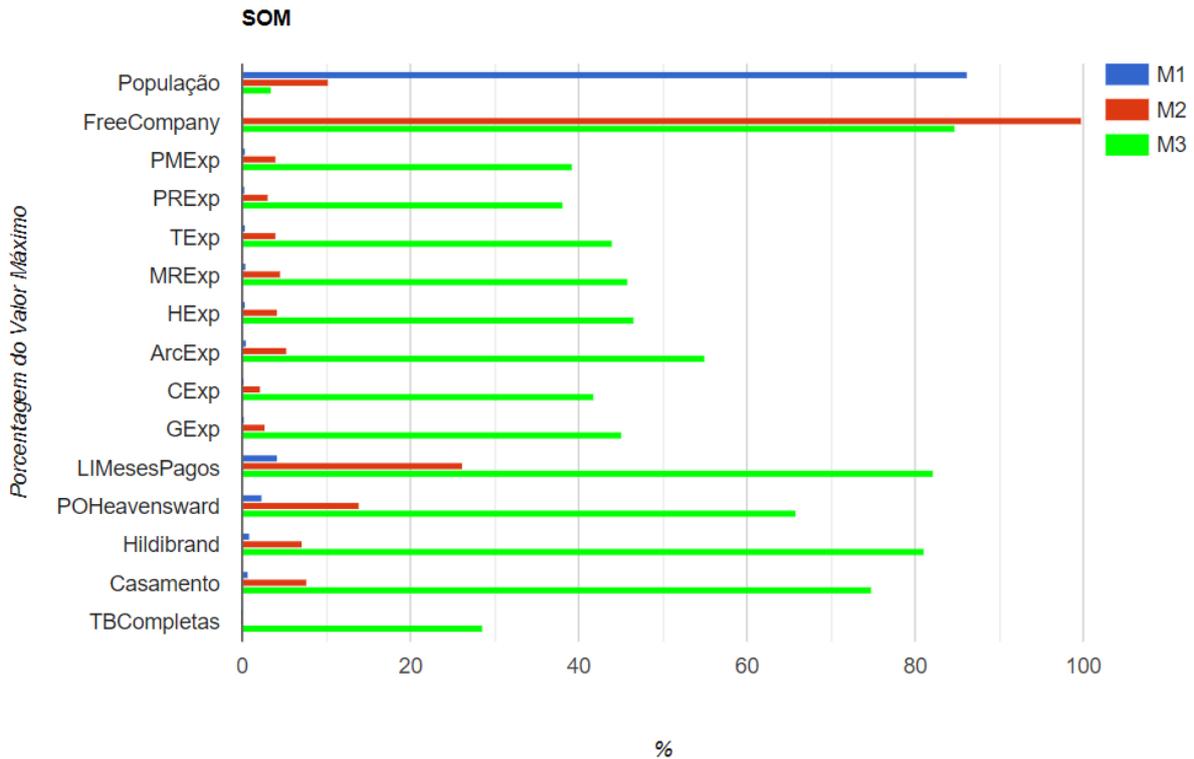


Figura 7.11: Porcentagens média de cada Cluster Resultante do SOM

Médias

O gráfico resultante das médias deste algoritmo está exibido na Figura 7.12. Analisando os resultados, foi possível identificar a existência consistente de dois clusters, o primeiro e terceiro cluster criados no mês Junho de 2017. Estes foram encontrados no meses seguintes, enquanto o Cluster 2 desse mesmo mês não foi identificado nos meses seguintes sendo substituído por outro cluster com médias semelhantes, mas diferentes para serem diferenciados. Uma possível explicação é que as porcentagens do mês de Junho de 2017 foram retiradas no final da segunda expansão do jogo, logo elas podem representar o potencial final do Cluster 2 no meses subsequentes. A consistência desse algoritmo foi quebrada no mês de Junho de 2018, ao identificar o Cluster 3 de meses anteriores como Cluster 2.

7.2.7 Fuzzy C-means

O último algoritmo a ser executado foi o representante da metodologia Difusa, *Fuzzy C-means*. Como o algoritmo anterior, o *Fuzzy C-means* resultou em três clusters devido ao valor de entrada. As figuras presentes no Anexo mostram as descrições dos clusters resultantes.

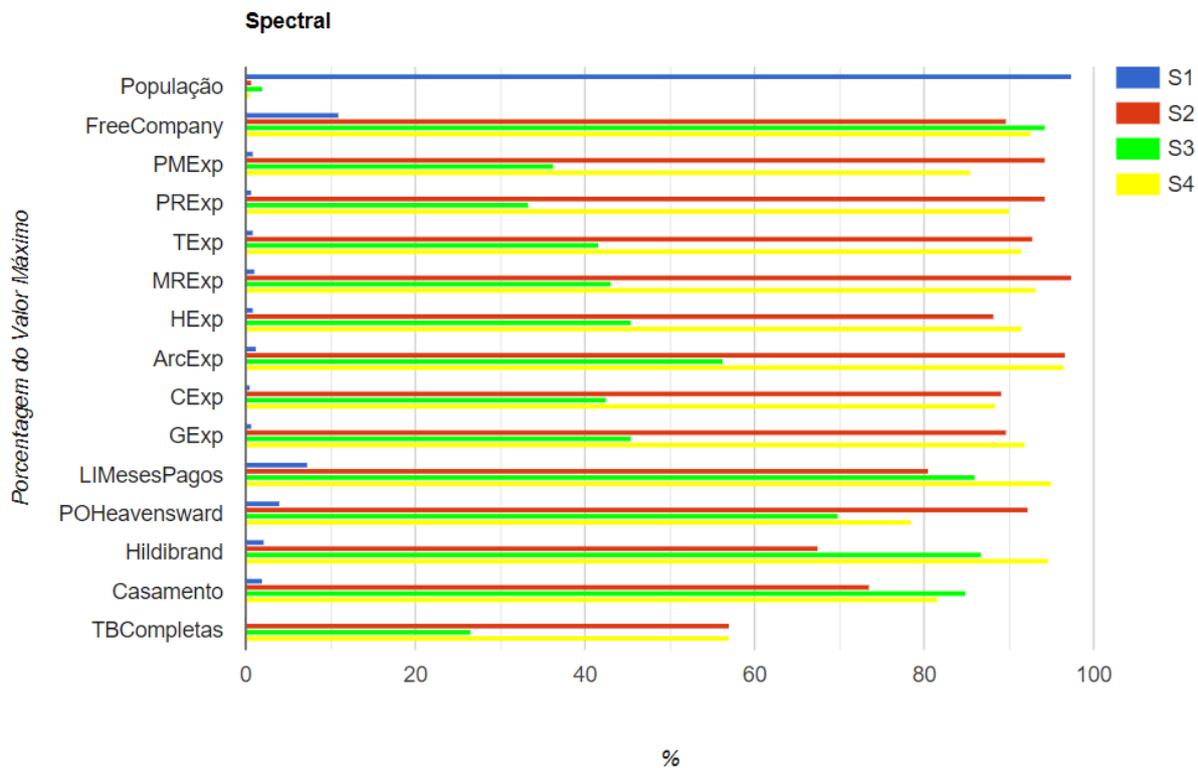


Figura 7.12: Porcentagens média de cada Cluster Resultante do Spectral

Médias

Apesar da característica de que um elemento de dado possa pertencer a mais de um cluster, o resultado deste algoritmo não apresentou nenhum personagem em mais de um cluster. Fato evidenciado pela soma dos membros de todos clusters em cada mês ser igual ao tamanho do domínio.

Observando o gráfico gerado pela porcentagem das médias dos clusters identificados exibido na Figura 7.13, a semelhança com os clusters formados pelos algoritmos *SOM* e *K-means* é evidente. Outra peculiaridade importante a se ressaltar é a consistência presente nos clusters resultantes. Como em outros algoritmos anteriores, o Cluster 1 do mês de Junho de 2017 é o mesmo identificado nos Clusters 1 de cada mês subsequente. Ocorrência que se manteve para os Clusters 2 e 3.

7.3 Avaliações

O foco desta seção inclui as avaliações realizadas sobre o resultado das clusterizações e da base de dados. Conjuntamente, uma comparação de performance de cada algoritmo e o esforço de atribuição de perfis a clusters descobertos são relatados nesta seção.

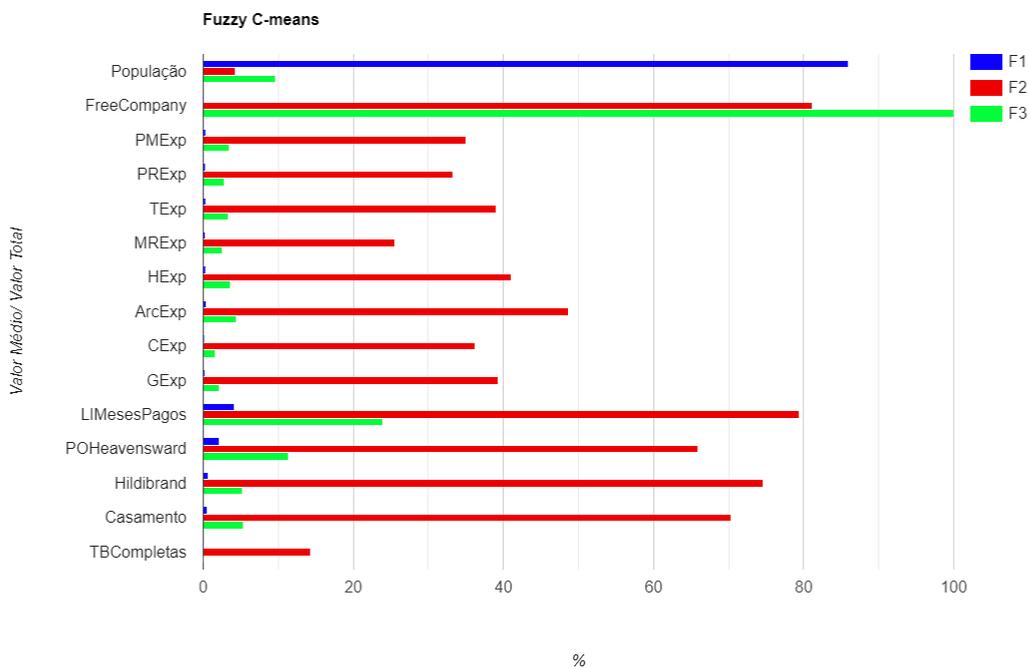


Figura 7.13: Porcentagens média de cada Cluster Resultante do C-means

7.3.1 Estatística Hopkins

A aplicação da Estatística de Hopkins foi realizada em cada um dos meses para validar a distribuição dos objetos de dados. O resultado desta aplicação está exposto no gráfico da Figura 7.14. O resultado da estatística foi constantemente relatado como superior à 0,965. Esses valores evidenciam a alta tendência à clusterização do domínio explorado. Logo, tais resultados confirmam a hipótese de que a base de dados do *Final Fantasy XIV* contém o potencial para análises de agrupamentos.

7.3.2 Avaliação dos Clusters Resultantes

Após a validação do domínio, utilizando a estatística de Hopkins, e a aplicação dos algoritmos de clusterização, o resultado do processo deve ser avaliado. Essa avaliação foi feita de acordo com os métodos relacionados na Metodologia. O resultado médio do Coeficiente de Silhueta para cada algoritmo está disposto na Figura 7.15. Observando os valores finais, nota-se que todos algoritmos alcançaram bons números, com o *DBSCAN* alcançando o maior coeficiente. Entretanto a diferença entre os resultados foi mínima, indicando uma equivalência quando observando apenas o Coeficiente de Silhueta.

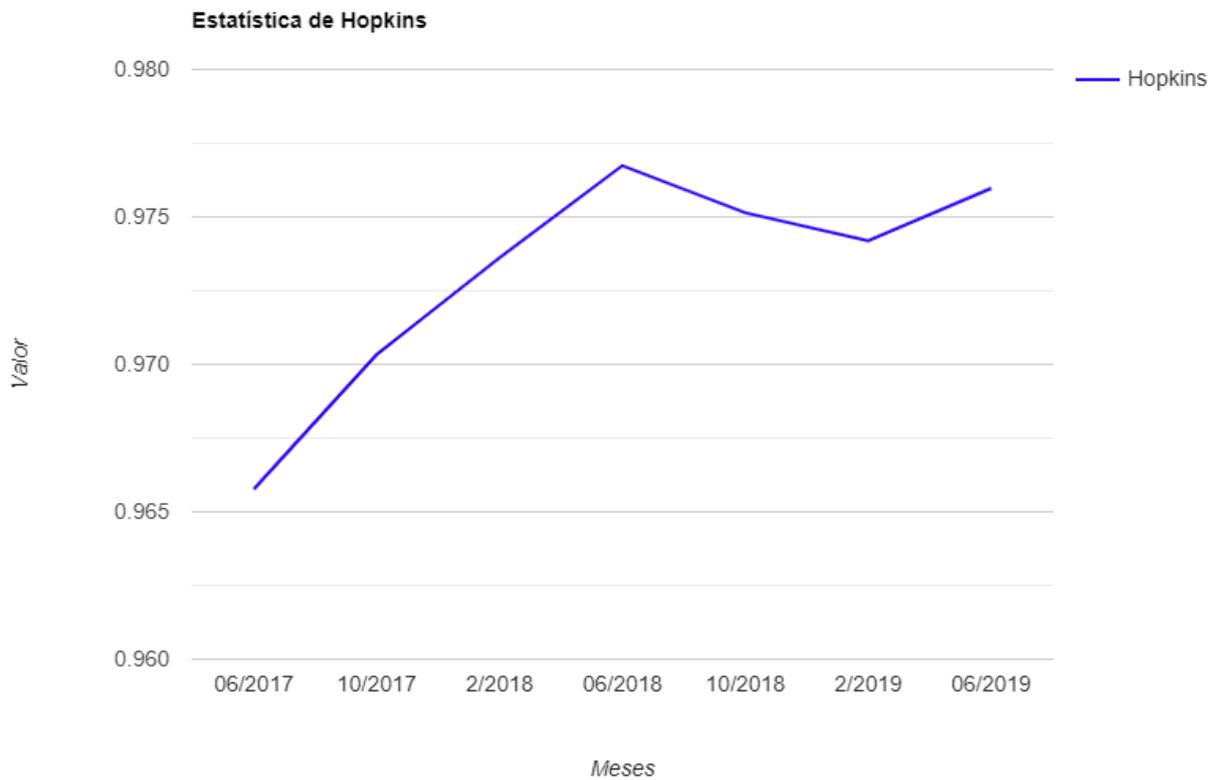


Figura 7.14: Estatística Hopkins dos meses analisados

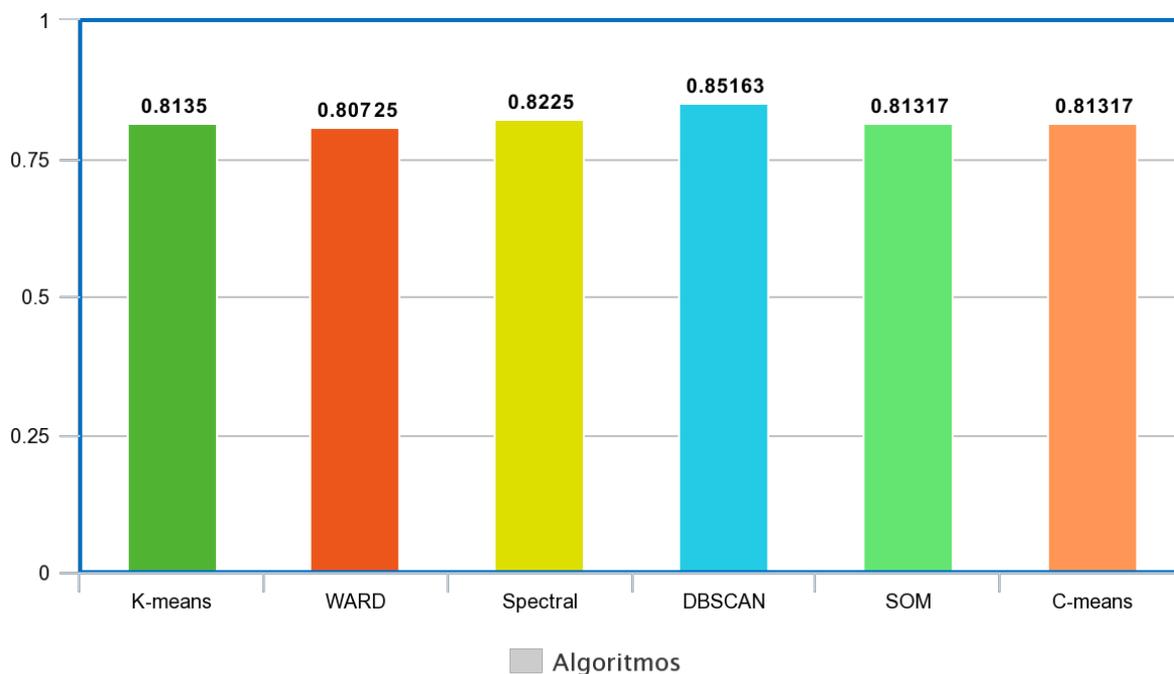
Ao examinar o resultado do Índice de Calinski-Harabasz projetado na Figura 7.16, uma disparidade maior é notada entre os algoritmos. Os algoritmos *K-means*, *SOM* e *C-means* relataram os melhores números, com o algoritmo *WARD* em quarto lugar, com números aproximadamente 20% menores que o índice do melhor algoritmo desta avaliação, *K-means*. E finalmente, os algoritmos *Spectral* e *DBSCAN* resultaram nos piores índices.

A última técnica de avaliação de clusters aplicada neste trabalho foi o Índice de Davies-Bouldin, cujo resultado está apresentado na Figura 7.17. Neste gráfico todos os algoritmos alcançaram resultados razoáveis, entretanto a diferença entre o melhor e pior algoritmo, *K-means* e *Spectral* respectivamente, é consideravelmente maior que a diferença no Coeficiente de Silhueta.

7.3.3 Performance

Além dos métodos intrínsecos de avaliação de clusters, a performance de cada algoritmo foi avaliada em termos do tempo médio de execução. Gráfico da Figura 7.18 exibe o tempo médio de cada algoritmo calculado a partir da aplicação nos sete meses. As conclusões imediatas retiradas do gráfico são:

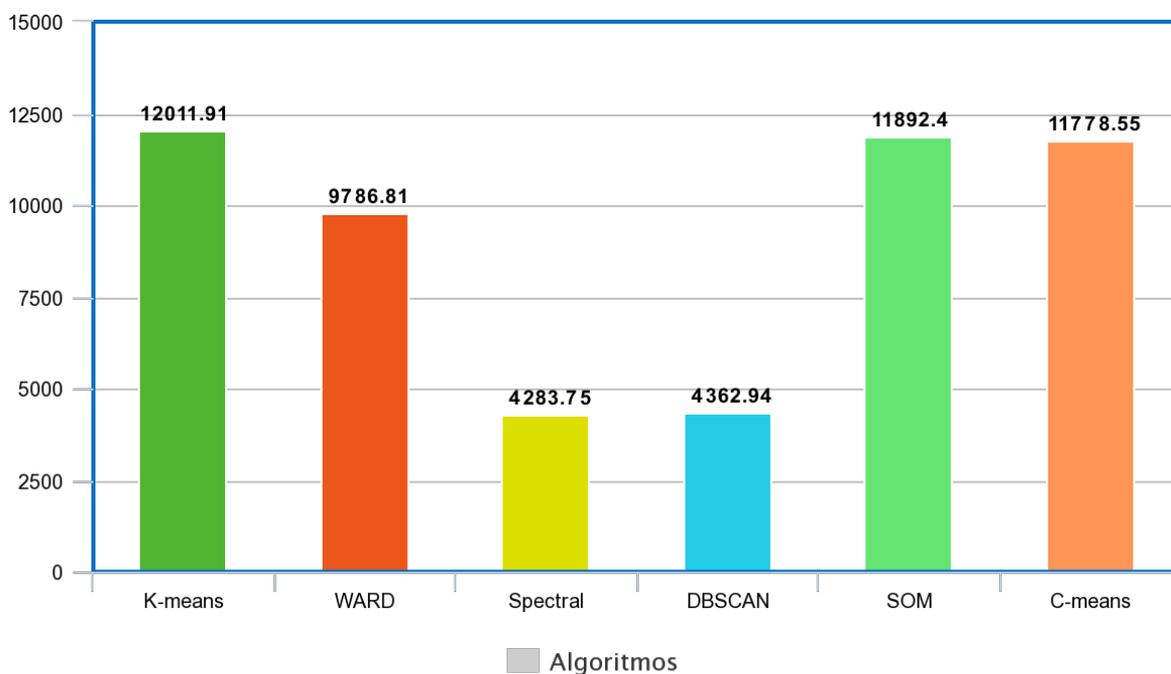
Coeficiente de Silhueta



meta-chart.com

Figura 7.15: Média dos Coeficientes de Silhueta

Índices de Calinski-Harabasz



meta-chart.com

Figura 7.16: Média dos índices de Calinski-Harabasz

Índices de Davies-Bouldin

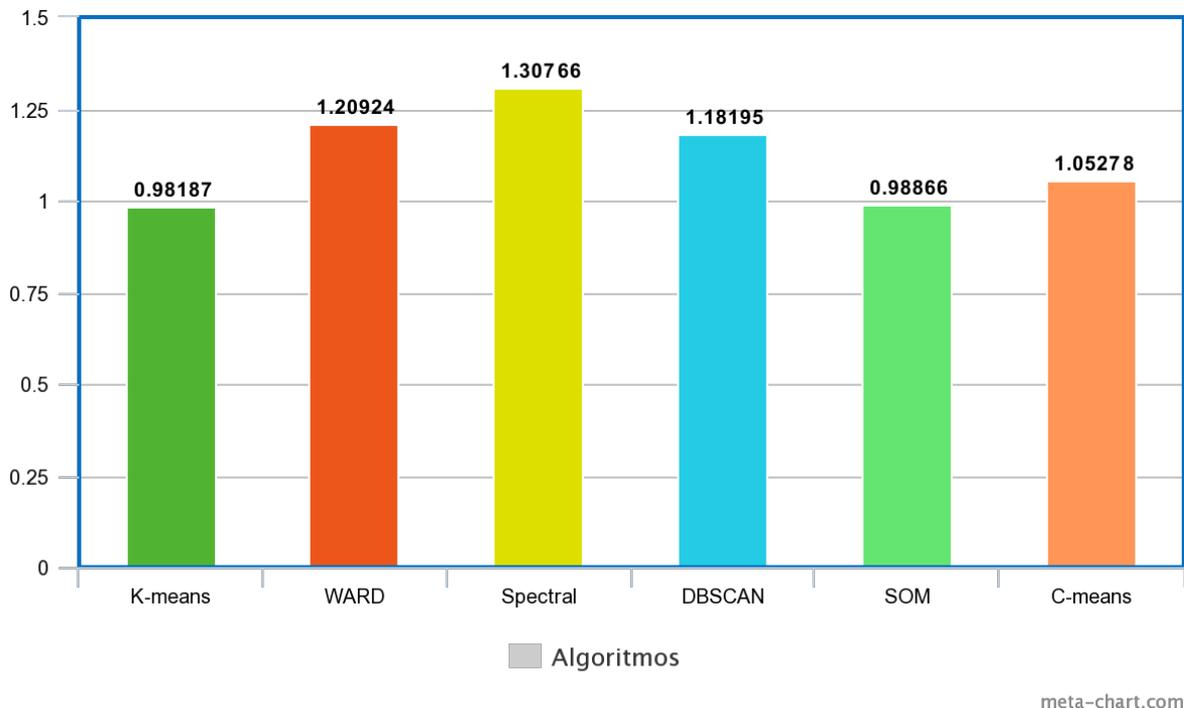


Figura 7.17: Média dos índices de Davies-Bouldin

- Os algoritmos *Spectral* e *BANG* possuem pior performance, demorando mais de trinta segundos que os algoritmos restantes.
- O algoritmo que alcançou a melhor performance foi o *K-means*.
- Os algoritmos *SOM* e *Fuzzy C-means* obtiveram performances semelhantes.

7.4 Atribuição de Perfis

A atribuição de perfis aos clusters resultantes não é englobada pelos métodos de avaliação de clusterização deste trabalho. Entretanto, com intuito de aumentar a contribuição desta pesquisa para melhor compreensão dos jogadores de *Final Fantasy XIV*, uma análise sobre os valores de cada atributo dos clusters produzidos foi usada como base para proposta de perfis. Essa atribuição foi feita agrupando os clusters resultantes de acordo com seus atributos e, posteriormente, analisando as margens de seus valores para determinação de um perfil condizente.

A população de cada agrupamento resultante foi o indicador usado para atribuição de um cluster à um grupo específico. Em seguida, as médias de cada atributo foram analisadas para criação de margens descritivas daquele grupo. Diferentemente dos ou-

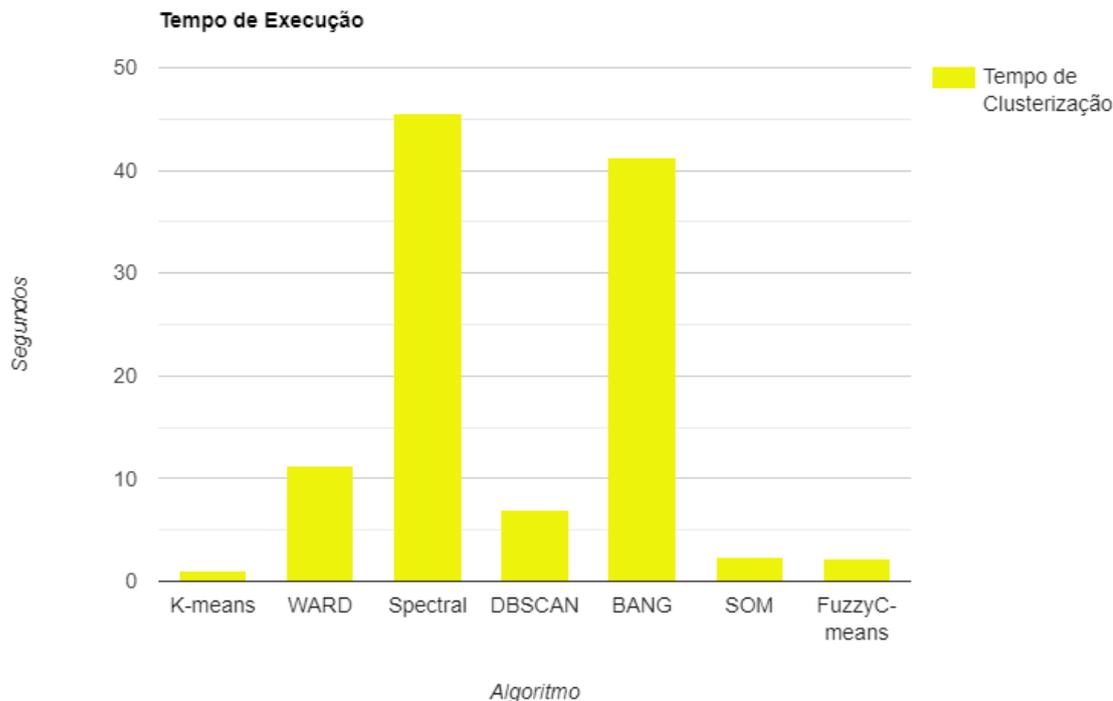


Figura 7.18: Média do tempo de execução dos algoritmos de Clusterização

tros algoritmos, o *DBSCAN* identificou três clusters com população semelhante à outros clusters, mas possuindo atributos com valores distantes, logo evidenciando um grupo de clusters diferente.

Analisando os limites identificados de cada atributo dos grupos de clusters, foi feita uma proposta de perfil de jogador com base na porcentagem de conteúdo explorado, principalmente conteúdo relacionado com os níveis de *jobs* e *class*. Essa proposta faz uso do conhecimento de mecânicas sobre progressão de personagem presente no Capítulo 4. Um exemplo desta distinção é a identificação de jogadores que exploram missões secundárias em contrapartida a jogadores que focam apenas em conteúdos que melhorem seu desempenho. A Tabela 7.2 relaciona os clusters resultantes com os perfis propostos.

As porcentagens foram geradas ao se dividir o valor médio do atributo pelo máximo possível referente aquele atributo. As Figuras 7.19 até 7.22 demonstram as margens descritivas de cada atributo referente ao específico perfil.

O perfil Iniciante, identificado por todos algoritmos com exceção do *BANG*, contempla a maior porcentagem dos personagens presentes na base de dados, contando com números superiores a 80% da população. A característica mais marcante deste perfil é o fato de ser o ponto inicial para todos jogadores. Devido a essa característica, a maioria dos atributos

	K-means	WARD	Spectral	SOM	C-means	DBSCAN
Iniciante	P1	H2	S1	M1	F1	D1
Casual	P2	H1		M2	F3	D3
Dedicado	P3	H3	S3	M3	F2	
Intermitente						D2,D4,D5
Hardcore			S2,S4			

Tabela 7.2: Tabela de Identificação de Perfil por Algoritmo

dos membros deste perfil possuem os menores valores médios de qualquer perfil. Esse perfil, geralmente, não explorou ainda mais de 1% de qualquer estilo de jogo, com pouca participação nas atividades de Tribos Bestiais, Casamento ou Hildibrand. Menos de 10% se comprometeram com a compra da primeira expansão. Aproximadamente 15% dos personagens fazem parte de um clã de jogadores e a média de mensalidades pagas está em torno dos três meses. A soma de todos jogadores que iniciaram o jogo recentemente e que desistiram de jogar após pouca experiência compõem este grupo. Um dos desafios mais importantes de desenvolvedores de MMOs é incentivar os jogadores a migrarem deste perfil para perfis mais dedicados.

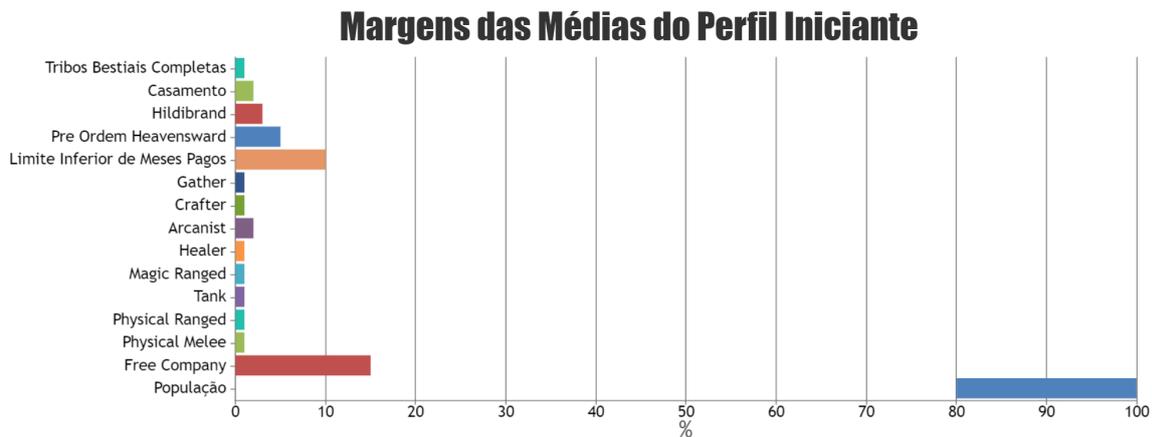


Figura 7.19: Porcentagem Média dos Atributos do Perfil Iniciante

O próximo perfil identificado, denominado Casual, reúne jogadores com uma maior experiência no jogo do que o grupo anterior, possuindo uma média entre 5 e 12 mensalidades pagas. Estes jogadores superaram os desafios do início da jornada e começaram um engajamento com jogo. Suas médias estão presentes na Figura 7.20. Este perfil engloba uma população média menor que 15%, mas maior que 5%, do domínio. Suas experiências com cada meio de se jogar superando os valores mínimos dos Iniciantes, mas ainda não constituem nem 10% do possível em qualquer atributo deste tipo. Jogadores casuais, em geral, são denominados como tal devido ao ato pouco frequente de jogar. Seu foco costuma ser o entretenimento momentâneo, evitando investir o tempo necessário para otimizar suas

ações dentro de qualquer jogo. Apesar de não serem jogadores ávidos, eles ainda investem no jogo e podem voltar com facilidade quando alguma atualização os atraia.

Um terço dos jogadores casuais realizaram a compra prematura da primeira expansão, enquanto uma média de 15% participaram de um casamento e completaram as missões de Hildibrand. As missões de Tribos Bestiais não foram exploradas por este perfil. Uma característica interessante é a alta porcentagem de membros pertencentes a clãs, superando 75% dos membros. Este aspecto é altamente vantajoso para o retorno em massa de jogadores, pois jogadores casuais com um forte aspecto social podem voltar a investir tempo no jogo se outros se unirem a ele. Esse fenômeno pode ser a justificativa deste perfil possuir um elevado número de meses pagos e baixa experiência, pois muitos voltam a jogar com facilidade, mas não dedicam tempo suficiente para avançar em seus *jobs*.

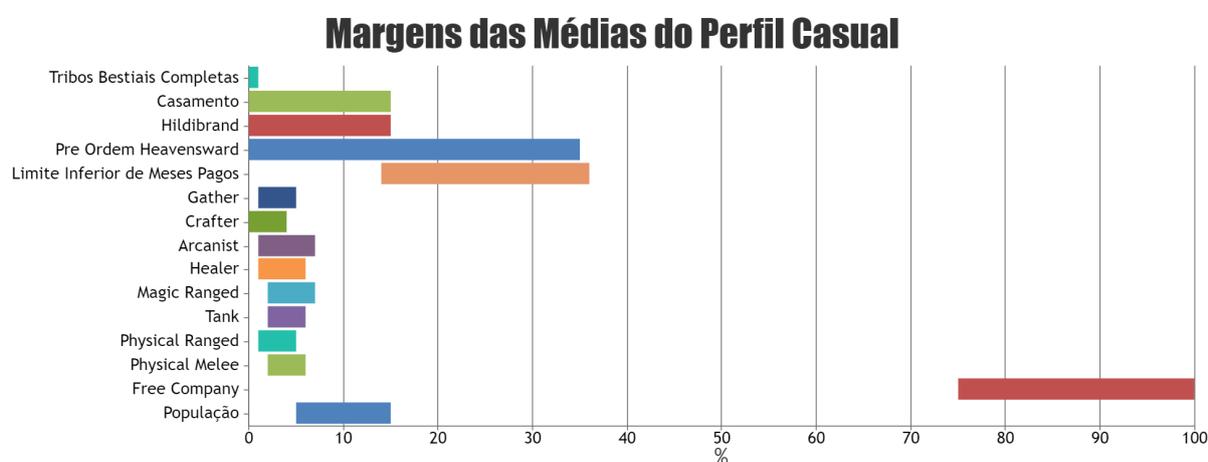


Figura 7.20: Porcentagem Média dos Atributos do Perfil Casual

Além de Iniciantes e Casuais, outro grande grupo de jogadores presentes em qualquer jogo são aqueles que investem uma quantidade consideravelmente maior de seu tempo no jogo. Esse grupo é representado pelo perfil Dedicado, cujas margens de porcentagens estão representadas na Figura 7.21. O alto engajamento deste perfil é evidenciado pelo aproveitamento de 40% em alguma ou maioria dos meios de jogo disponíveis, valor muito superior aos perfis anteriores. Esses valores exibem uma alta probabilidade do jogador possuir *jobs* no nível máximo nos diferentes meios de jogo, revelando uma característica de exploração e otimização das atividades. Tais características também são visualizadas nos altos valores dos atributos de Tribos Bestiais, Casamentos e missões de Hildibrand, configurando um considerável interesse deste perfil em experiências fora da missão principal.

Além do aproveitamento expressivo destes jogadores, o compromisso financeiro é notadamente maior que outros perfis. Uma margem de 55% a 79% dos jogadores deste perfil fizeram a compra prévia da primeira expansão e o atributo de mensalidades pagas é o

máximo visto neste banco de dados. A dedicação dos jogadores deste perfil é rara, constituindo apenas de 5% a 2% dos personagens presentes no jogo. Entretanto sua existência influencia e inspira outros perfis a dedicarem mais ao jogo para poder chegar onde este perfil chegou, auxiliando assim o crescimento do jogo.

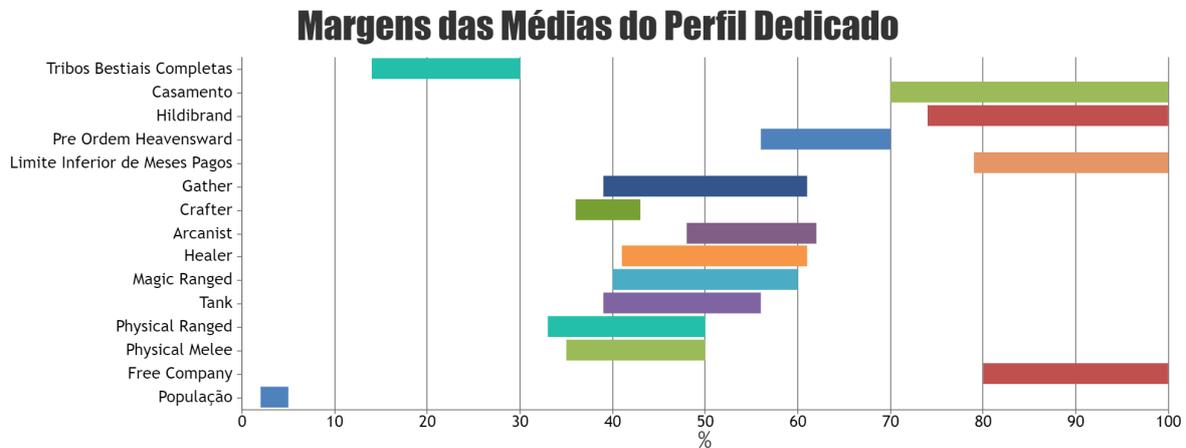


Figura 7.21: Porcentagem Média dos Atributos do Perfil Dedicado

Os algoritmos *DBSCAN* e *Spectral* identificaram clusters com porcentagens de população inferiores ao perfil Dedicado identificado pelos outros algoritmos. Possivelmente ignorados pela limitação do número de clusters, os dois próximos perfis estão contidos nos agrupamentos dos algoritmos que não os identificaram.

O perfil Hardcore pode ser visto como um extremo do Perfil Dedicado, o ápice dos jogadores. Jogadores deste perfil buscam explorar e otimizar o máximo de cada possibilidade oferecida pelo jogo. A Figura 7.22 mostra as maiores porcentagens observadas neste trabalho, com todos atributos possuindo uma margem que engloba o aproveitamento máximo. Mencionado no Capítulo 4, o conteúdo do *End Game* é o objetivo desses jogadores, eles buscam melhorar seu personagem e equipamento ao máximo para poder conquistar os desafios mais complexos e difíceis deste jogo. A recompensa dessas conquistas é comumente associada à um elemento visual, permitindo que esses jogadores destaquem-se entre outros. Esse destaque gera orgulho nos jogadores *Hardcore*, aumentando ainda mais seu engajamento, e incentiva outros jogadores a dedicarem-se mais ao jogo para alcançar tal regozijo. Em contra partida, a população deste perfil é menor que 1% do domínio indicando a dificuldade para alcançar tamanho resultado.

Finalmente, o último perfil identificado é o perfil Intermitente, composto por clusters que não foram atribuídos a nenhum outro perfil já citado. Apenas o algoritmo *DBSCAN* foi capaz de identificá-lo, dentro das configurações desta pesquisa. Os três clusters atribuídos a esse perfil possuem médias semelhantes em quase todos atributos. Uma peculiaridade é a concordância entre esses clusters quando relatam uma alta porcentagem

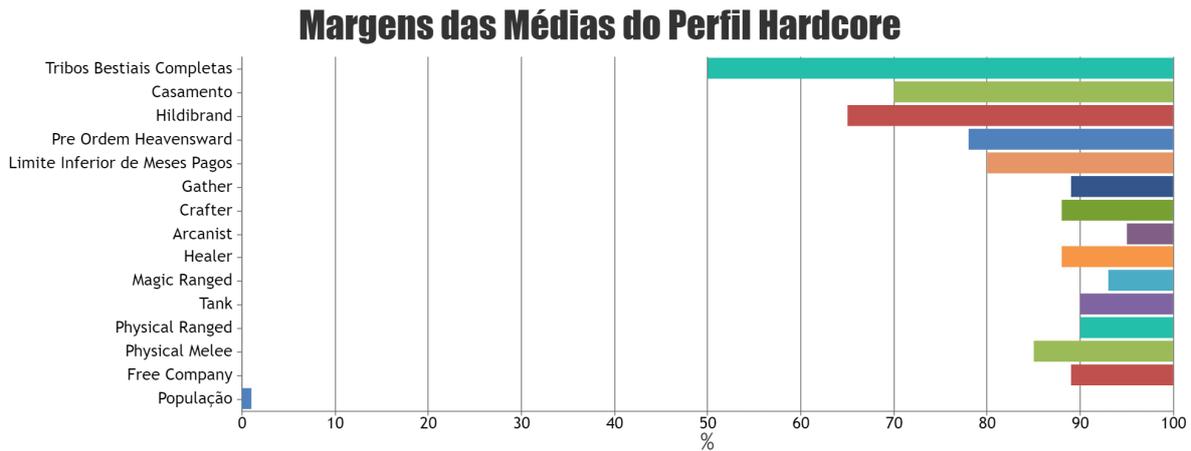


Figura 7.22: Porcentagem Média dos Atributos do Perfil Hardcore

de jogadores que fizeram a compra prévia da expansão *Heavensward*. Entretanto existem quatro atributos onde os clusters deste perfil diferem, *FreeCompany*, Mensalidades, Hildibrand e Casamento. Nesses atributos, a diferença entre porcentagens chega a ser superior a 50%. Apesar de um dos cluster possui uma alta média de mensalidades paga, suas experiências estão menores que 10% do máximo.

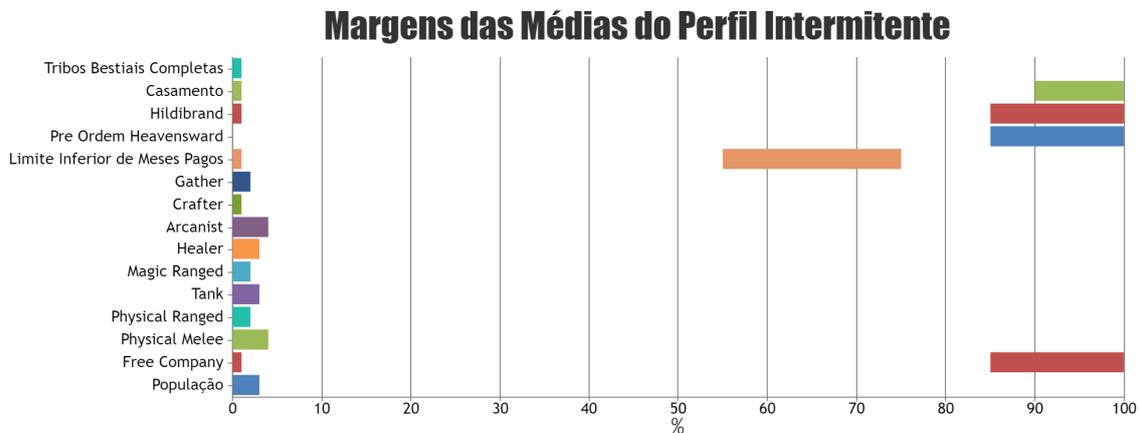


Figura 7.23: Porcentagem Média dos Atributos do Perfil Intermitente

Uma hipótese para este comportamento é a criação de personagens em prol de outro personagem. Por exemplo, criar um personagem extra para casar com o personagem principal e possibilitar as vantagens do casamento para o personagem principal sem depender de outro jogador, ou até para aquisição de itens únicos com o segundo personagem e vendê-los para o lucro do personagem principal sem que ele perca o item único dele. Outra hipótese sobre os personagens deste perfil inclui jogadores novos que conheciam outro jogador já pertencente a uma *Free Company*, assim o personagem entra em um clã

logo no início do jogo, realiza a compra da primeira expansão antes do lançamento e logo após não dá continuidade às sessões de jogo.

7.5 Perfis Não Identificados

Apesar da limitação dos dados disponíveis não ter impossibilitado a análise de cluster, a ausência de atributos importantes para o estudo de MMORPGs dificultou a identificação de perfis derivados dos perfis identificados, ou classificou personagens que não condizem com o perfil do jogador.]==][p]

O acesso às informações sobre as sessões de jogo do personagem permitiriam a classificação de perfis *Hardcore* Especialistas. Um perfil de jogador que optou por se especializar em apenas um *job*, jogando quantidades concentradas e prolongadas de tempo, provavelmente seria enquadrado como Dedicado ou Casual nos algoritmos aplicados neste trabalho, apesar de suas características serem mais semelhantes ao *Hardcore*.

Outro perfil que seria identificado com essas informações ausentes é o *Hardcore* Temporário. Diferentemente do *Hardcore*, este perfil focou em evoluir os *jobs* disponíveis para o máximo e parou de jogar, não investiu mais tempo no jogo e não explorou conteúdo *End Game*. Até que uma nova expansão seja lançada e aumente o nível máximo, esses jogadores serão classificados como *Hardcore* pelos algoritmos que identificaram este perfil. Entretanto, a dedicação e engajamento associados ao perfil *Hardcore* não estão presentes no perfil *Hardcore* Temporário na mesma proporção.

Esses perfis englobam tipos de sessão de jogo diferentes, que não são identificáveis por esse banco. A frequência e volume de sessões de jogo, usadas em outros trabalhos [13], não pode ser estudada neste trabalho devido a limitação da base de dados, dificultando a identificação desses perfis. Para contornar tal adversidade, uma avaliação sobre os equipamentos e itens do jogador permitiria identificar o tipo de conteúdo explorado pelo jogador assim como a frequência com que um novo item é adquirido. Infelizmente tal informação também está fora do domínio disponível.

Por fim, dados peculiares que forneceriam informações interessantes sobre os jogadores de *Final Fantasy XIV* seriam a identificação do jogador dono daquele personagem e de quais atividades foram feitas por ele. Além de melhorar a classificação de personagens na clusterização, conhecer as atividades do jogador permite a investigação do fenômeno *Roleplay*. Esse fenômeno envolve a simulação de uma vida real dentro do jogo, onde o jogador visa cumprir com as responsabilidades e atividades que uma pessoa do mundo real possui, porém no mundo virtual. A identificação deste perfil não foi possível com os dados disponíveis.

A identificação de jogador possibilitaria o estudo sobre jogadores que possuem mais de um personagem. Independentemente do fato de um personagem possuir a capacidade de participar e executar qualquer conteúdo do jogo, existe a possibilidade de pagar para possuir mais de um personagem. O estudo do perfil desses jogadores possui grande relevância na compreensão do sucesso de jogos MMORPGs.

Capítulo 8

Considerações Finais

Este capítulo dedica-se à finalização desta dissertação. Para encerrá-la, serão feitas sumarizações e reflexões sobre cada uma das contribuições discutidas na proposta. Por fim, serão apresentadas algumas possíveis vertentes para futuros trabalhos a partir desta pesquisa.

8.1 Nova Base de Dados

A primeira contribuição esperada deste trabalho é a introdução da Base de Dados do *Final Fantasy XIV* para comunidade científica, ampliando assim a diversidade nos estudos acerca dos jogos MMO.

Entretanto, no caso do *Final Fantasy XIV* especificamente, a simples sugestão de utilização de um banco de dados diferente não basta para impulsionar pesquisas neste jogo. Para corroborar com o objetivo desta pesquisa, a aplicação da estatística de Hopkins revelou o alto potencial de agrupamento contido nesta base de dados. Outra contribuição foi feita utilizando o Mapa de Calor com Correlações de Pearson, mapa este que indicou a alta correlação entre variáveis independentes. A restrição contida nessas bases, tanto de frequência quanto de atributos comumente associados às análises de MMO, não foram prejudiciais o suficiente para impossibilitar a sua análise. Finalmente, esta base possui um aspecto almejado em análises de dados, a alta dimensionalidade e tamanho do domínio, aspectos estes que promovem diversas combinações de análises, permitindo a superação das limitações da base de dados. A proposta deste trabalho de estudar a base de dados com os jogadores do *Final Fantasy XIV* foi corroborada pelos resultados obtidos, agora espera-se um maior engajamento da comunidade científica com essa base.

8.2 Comparação dos Algoritmos

A contribuição central deste trabalho consiste no resultado da comparação das diferentes técnicas de clusterização. Esta pesquisa aplicou sete algoritmos de agrupamento, cada um pertencente a uma metodologia diferente, em sete bancos de dados referentes aos jogadores de *Final Fantasy XIV*, no qual cada banco corresponde a um mês específico. Após essa aplicação, os métodos de avaliação de clusterização disponíveis foram empregados no resultado. Embasando-se nessas avaliações e resultados, múltiplas conclusões foram feitas.

Não foi observado no resultado das clusterizações algum impacto negativo da ausência de mais atributos ou da disponibilidade de dados com maior frequência. Logo a restrição inerente desta base de dados não prejudicou a análise de clusters.

Em concordância com a ausência de um melhor algoritmo de clusterização para qualquer caso, não foi possível identificar o melhor algoritmo para os personagens de *Final Fantasy XIV*. Entretanto, o algoritmo *BANG* não conseguiu clusterizar em nenhuma de suas aplicações, configurando o pior resultado além de possuir o segundo pior desempenho.

O algoritmo mais clássico de clusterização, *K-means*, alcançou o melhor desempenho de todos algoritmos aplicados, caracterizando-o como escolha primária caso o pesquisador possua limites em termos de hardware. Com clusters altamente semelhantes, o algoritmo *SOM* obteve resultados quase idênticos ao *K-means*, entretanto, a vantagem do algoritmo *SOM* é identificada na consistência dos clusters formados, ou seja, o primeiro cluster em um determinado mês é o mesmo primeiro formado em qualquer mês subsequente. Essa consistência configura o algoritmo *SOM* como uma boa escolha ao analisar bases de dados com uma grande diferença temporal entre si, facilitando a identificação de possíveis perfis. Peculiarmente, o algoritmo *Fuzzy C-means* alcançou as mesmas conquistas que o *SOM*, logo suas aplicações neste cenário são equivalentes.

O algoritmo *Spectral* obteve baixas avaliações, possuindo o pior Índice de Davies-Bouldin e o pior Índice de Calinsk-Harabasz. Além desses resultados, o desempenho deste algoritmo foi o pior observado, dificultando a recomendação de utilizar este algoritmo. Entretanto, uma das vantagens deste algoritmo foi a identificação do Perfil Hardcore, algo exclusivo em relação aos outros algoritmos, logo sua aplicação pode ser feita para complementar a aplicação de outro algoritmo.

O algoritmo *WARD* não conquistou melhor resultado em nenhuma avaliação e teve o terceiro pior desempenho. Apesar de ter identificado clusters com consistência, não foi o único a conseguir tal feito. Conseqüentemente, baseando-se nesses resultados, sua aplicação é aconselhada apenas em casos de validação de resultados de outras aplicações.

Finalmente, o último algoritmo estudado neste trabalho, *DBSCAN*, obteve avaliações inconsistentes, possuindo o melhor Coeficiente de Silhueta e um dos piores Índices de Calinski-Harabasz. Em contra partida ao *Spectral*, seu desempenho foi razoável, sendo

o quarto melhor e pior desempenho. A peculiaridade deste algoritmo envolve a falta de limitação quanto ao número de clusters, logo o número final de clusters não é conhecido a priori. Essa característica impacta na consistência do resultado, de que apenas cinco clusters foram relacionados com persistência. Outra fraqueza desse algoritmo é a categorização de objetos de dados como ruído, aumentando o erro na conclusão final sobre a amostra. Apesar de suas fraquezas, o uso deste algoritmo é vantajoso para descobrir novos clusters e perfis.

8.3 Perfis de jogadores de *Final Fantasy XIV*

Uma das intenções deste trabalho é analisar o jogador de *Final Fantasy XIV*, compreender a forma como o jogador consome essa mídia. Nesse sentido, a terceira contribuição pretendida deste trabalho é feita mediante a relação dos clusters resultantes com possíveis perfis de jogadores. A descrição de cada perfil foi feita manipulando as médias dos atributos de cada cluster, manipulação essa que atribuiu uma margem de porcentagens para cada atributo. Deste modo, cinco perfis foram identificados e apontados: Iniciante, Casual, Dedicado, Hardcore e Intermitente.

A predominância do perfil Iniciante tanto em seu reconhecimento quanto em sua presença pode ser justificada pela facilidade que existe na criação de um personagem novo, podendo ser realizada até antes de comprar o jogo. Jogadores deste perfil não chegaram a vivenciar 1% das diversas experiências que o jogo fornece. Enquanto poucos personagens deste perfil possuem clãs de jogadores, o perfil Casual, que engloba um domínio de 5% a 15% dos personagens, possui por volta de três quartos de sua população, membros de uma *Free Company*. Personagens do perfil Casual experimentam os diferentes meios de jogo, apesar de não superar 10% do que é oferecido. Alguns já participaram de casamento e outros já realizaram as missões de Hildibrand. Esses jogadores já pagaram por volta de 4 até 12 mensalidades.

O terceiro perfil amplamente identificado pelos algoritmos é o perfil Dedicado. Este perfil reúne cerca de 2% até 5% do domínio, onde a maioria possui clãs de jogadores, participou de casamentos, realizou as missões de Hildibrand e fez a compra prévia da expansão *Heavensward*. Jogadores deste perfil tendem a explorar a maioria dos meios de jogo, com aproveitamento de cada meio sempre com médias superiores a 30%. Personagens que alcançaram este perfil tendem a necessitar de muito dedicação e tempo, conseqüentemente, a média mensalidades pagas tende a ser maior que vinte e cinco meses.

O perfil Hardcore caracteriza um extremo do perfil Dedicado, onde todos meios de jogar possuem mais de 85% de experiência. Entretanto, a média de aproveitamento de missões de Hildibrand é menor que a do perfil Dedicado, possivelmente justificado pelo

pouco incentivo para realizá-las. Um fato peculiar é o pouco aproveitamento observado das missões de Tribos Bestiais, o perfil mais dedicado, Hardcore, só alcançou uma média de 50% de aproveitamento.

O quinto perfil identificado, Intermitente, ocupa um espaço de especulação. Não foi possível aferição sobre o comportamento dos jogadores deste perfil baseando-se apenas em suas médias. Compondo cerca de 3% da população, este perfil necessita de uma maior investigação para sua compreensão.

A restrição presente nesta base de dados, em termos de atributos de tempo de jogo e de atividades, provavelmente dificultou a descrição do perfil Intermitente. Entretanto, apesar desse obstáculo, um ponto de partida para qualquer pesquisa que vá investigar os jogadores de *Final Fantasy XIV* foi delineado através da identificação desses perfis de jogadores.

8.4 Perspectiva

Analisando o processo realizado neste trabalho, certos aspectos foram identificados, os quais se beneficiariam de um estudo mais aprofundado. Um desses aspectos envolve uma reflexão mais detalhada sobre cada metodologia específica de clusterização. Cada metodologia de clusterização possui uma vasta gama de algoritmos, a exploração das diferenças entre essas soluções e a extração de conclusões generalizadas para o problema em questão, partindo das conclusões deste trabalho, exhibe áreas de contribuição para futuras pesquisas. O resultado dessa contribuição é observado no momento de escolha do algoritmo para aplicação em análises de clusters.

Outro ponto focal para os próximos estudos envolve a investigação temporal do comportamento dos clusters encontrados nesta base, direcionando-a para uma análise preditiva do costume dos jogadores de *Final Fantasy XIV*. Diferentemente da escolha randômica utilizada nesta pesquisa, a observação do comportamento de personagens específicos em suas transições entre clusters deve produzir resultados interessantes à desenvolvedores, auxiliando-os em projetar mecânicas e interações que facilitem a evolução dos personagens entre perfis. A análise preditiva permitirá um planejamento de conteúdo entre atualizações para condizer com a quantidade de personagens estimados que poderão usufruir do conteúdo desenvolvido.

Além dessas vertentes, existe outro aspecto dos clusters resultantes que pode ser aprofundado. A confirmação da existência dos perfis identificados de jogadores permitirá o direcionamento de ações de desenvolvedores para agradar os jogadores de um determinado perfil. Por outro lado, a correlação com outros perfis encontrados em outros jogos possui um amplo valor para a comunidade científica, possibilitando a criação de um perfil gené-

rico, que independe do jogo estudado. A validação dos perfis identificados é vantajoso para desenvolvedores, enquanto que a criação deste perfil possui um potencial extraordinário para compreensão do ser humano como jogador.

Referências

- [1] Ell, Kellie: *Video game industry is booming with continued revenue.* <https://www.cnn.com/2018/07/18/video-game-industry-is-booming-with-continued-revenue.html>. Acessado em: 2019-11-27. 1
- [2] Wright, Steven T.: *There are too many video games. what now?* <https://www.polygon.com/2018/9/28/17911372/there-are-too-many-video-games-what-now-indieapocalypse>. Acessado em: 2019-11-27. 1
- [3] Takahashi, Dean: *Newzoo: Games market expected to hit \$180.1 billion in revenues in 2021.* <https://venturebeat.com/2018/04/30/newzoo-global-games-expected-to-hit-180-1-billion-in-revenues-2021/>. Acessado em: 2019-11-27. 1
- [4] Florio, Philip: *The 15 biggest failures in mmo history.* <https://www.thegamer.com/the-15-biggest-failures-in-mmo-history/>. Acessado em: 2019-11-27. 1
- [5] Drachen, Anders, Magy Seif El-Nasr e Alessandro Canossa: *Game Analytics: Maximizing the Value of Player Data.* Springer, 2013. 1, 3, 7, 8, 9, 34, 35, 40
- [6] Fernandes, Lucas V, Carla D Castanho e Ricardo P Jacobi: *A survey on game analytics in massive multiplayer online games.* Em *2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, páginas 21–2109. IEEE, 2018. 1, 34, 36, 38, 40
- [7] Bauckhage, Christian, Anders Drachen e Rafet Sifa: *Clustering game behavior data.* IEEE Transactions on Computational Intelligence and AI in Games, 7(3):266–278, 2014. 2, 18, 43
- [8] Chen, Chih Wei e Taoi Hsu: *Game development data analysis visualized with virtual reality.* Em *2018 IEEE International Conference on Applied System Invention (ICASI)*, páginas 682–685. IEEE, 2018. 2, 18, 41
- [9] Dheandhanoo, Tuang, Sittichai Theppaitoon e Pisal Setthawong: *Game play analytics to measure the effect of marketing on mobile free-to-play games.* Em *2016 2nd International Conference on Science in Information Technology (ICSITech)*, páginas 125–130. IEEE, 2016. 2, 18, 41

- [10] Benmakrelouf, Souhila, Neila Mezghani e Nadjia Kara: *Towards the identification of players' profiles using game's data analysis based on regression model and clustering*. Em *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, páginas 1403–1410. ACM, 2015. 2, 18, 41
- [11] Yang, Wanshan, Gemeng Yang, Ting Huang, Lijun Chen e Youjian Eugene Liu: *Whales, dolphins, or minnows? towards the player clustering in free online games based on purchasing behavior via data mining technique*. Em *2018 IEEE International Conference on Big Data (Big Data)*, páginas 4101–4108. IEEE, 2018. 2, 18, 41
- [12] Drachen, Anders, Rafet Sifa, Christian Bauckhage e Christian Thurau: *Guns, swords and data: Clustering of player behavior in computer games in the wild*. Em *2012 IEEE conference on Computational Intelligence and Games (CIG)*, páginas 163–170. IEEE, 2012. 2, 18, 41
- [13] Siqueira, Elton Sarmanho, Carla Denise Castanho, Genaina Nunes Rodrigues e Ricardo Pezzoul Jacobi: *A data analysis of player in world of warcraft using game data mining*. Em *2017 16th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, páginas 1–9. IEEE, 2017. 2, 18, 38, 40, 41, 109
- [14] Sifa, Rafet, Eric Pawlakos, Kevin Zhai, Sai Haran, Rohan Jha, Diego Klabjan e Anders Drachen: *Controlling the crucible: A novel pvp recommender systems framework for destiny*. Em *Proceedings of the Australasian Computer Science Week Multiconference*, página 39. ACM, 2018. 2, 18, 40, 41
- [15] Rodrigues, Luiz AL e Jacques D Brancher: *Improving players' profiles clustering from game data through feature extraction*. Em *2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, páginas 177–17709. IEEE, 2018. 2, 18, 19, 21, 41, 42
- [16] Kwon, Hyoungjin, Wooyoung Jeong, Dae Wook Kim e Seong Il Yang: *Clustering player behavioral data and improving performance of churn prediction from mobile game*. Em *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, páginas 1252–1254. IEEE, 2018. 2, 18, 23, 41, 42
- [17] Saas, Alain, Anna Guitart e Africa Perriñez: *Discovering playing patterns: Time series clustering of free-to-play game data*. Em *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, páginas 1–8. IEEE, 2016. 2, 19, 21, 42
- [18] Hsu, Sheng Yi, Chia Lin Hsu, Shing Yun Jung e Chuen Tsai Sun: *Indicator products for observing market conditions and game trends in mmog*. FDG '17, páginas 54:1–54:5. ACM, 2017. 2, 19, 21, 38, 40, 42
- [19] Anagnostou, Kostas e Manolis Maragoudakis: *Data mining for player modeling in videogames*. Em *2009 13th Panhellenic Conference on Informatics*, páginas 30–34. IEEE, 2009. 2, 19, 42

- [20] Jiang, Jehn Ruey, Ching Chuan Huang e Chung Hsien Tsai: *Avatar path clustering in networked virtual environments*. Em *2010 IEEE 16th International Conference on Parallel and Distributed Systems*, páginas 845–850. IEEE, 2010. 2, 42
- [21] Bell, Jonathan, Swapneel Sheth e Gail Kaiser: *A large-scale, longitudinal study of user profiles in world of warcraft*. Em *Proceedings of the 22nd International Conference on World Wide Web*, páginas 1175–1184. ACM, 2013. 2, 36, 38, 42
- [22] Ahmad, Muhammad, David Huffaker, Jing Wang, Jeffrey William Treem, Dinesh Kumar, Marshall Poole e Jaideep Srivastava: *The many faces of mentoring in an mmorpg*. Em *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, páginas 270–275, agosto 2010. 2, 40, 42
- [23] Rodrigues, Lia C, Clodoaldo AM Lima e Pollyana N Mustaro: *Clustering online game communities through som*. Em *2009 International Joint Conference on Neural Networks*, páginas 2699–2702. IEEE, 2009. 2, 27, 38, 42
- [24] Rodrigues, Lia C, Clodoaldo AM Lima, Pedro PB de Oliveira e Pollyana N Mustaro: *Clusterization of an online game community through self-organizing maps and an evolved fuzzy system*. Em *2008 Fourth International Conference on Natural Computation*, volume 2, páginas 330–334. IEEE, 2008. 2, 27, 41, 42
- [25] Han, Jiawei, Jian Pei e Micheline Kamber: *Data mining: concepts and techniques*. Elsevier, 2011. 2, 7, 8, 13, 14, 15, 16, 17, 18, 19, 22, 23, 24, 25, 26, 29, 30, 31, 32, 33, 78, 80, 82, 83, 84, 85
- [26] Aggarwal, Charu C. e Chandan K. Reddy (editores): *Data Clustering: Algorithms and Applications*. CRC Press, 2014. 2, 14, 15, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33
- [27] *Final fantasy xiv census*. <https://ffxivcensus.com/>. Acessado em: 2018-06-26. 2, 3, 64, 65, 66, 71, 73, 74, 75
- [28] Indvik, Lauren: *The fascinating history of online role-playing games*. <https://mashable.com/2012/11/14/mmorpgs-history/>. Acessado em: 2019-11-27. 2
- [29] Metacritic: *First-person games by metacritic*. <https://www.metacritic.com/browse/games/genre/metacritic/first-person/all?view=condensed>. Acessado em: 2019-11-27. 4
- [30] Ohlheiser, Abby: *Activision's destiny ships \$500 million in one day*. <https://www.washingtonpost.com/news/technology/wp/2014/09/11/activisions-destiny-ships-500-million-in-one-day/?noredirect=on>. Acessado em: 2019-11-27. 4
- [31] Tassi, Paul: *'diablo immortal' blizzcon backlash leads to activision stock slide*. <https://www.forbes.com/sites/insertcoin/2018/11/06/diablo-immortal-blizzcon-backlash-leads-to-activision-stock-slide/#2c625c181b88>. Acessado em: 2019-11-27. 4

- [32] Medler, Ben: *Generations of game analytics, achievements and high scores*. Eudamos. *Journal for Computer Game Culture*, 3(2):177–194, 2009. 8, 36
- [33] Berry, Michael e Gordon Linoff: *Mastering data mining: The art and science of customer relationship management*. John Wiley & Sons, Inc., 1999. 8
- [34] Harris, Donna: *A comparative study of the effect of collaborative problem-solving in a Massively Multiplayer Online Game (MMOG) on individual achievement*. University of San Francisco, 2008. 10, 37, 40
- [35] Steinkuehler, Constance A: *Learning in massively multiplayer online games*. Em *Proceedings of the 6th international conference on Learning sciences*, páginas 521–528. International Society of the Learning Sciences, 2004. 10, 36
- [36] Magnussen, Rikke e Morten Misfeldt: *Player transformation of educational multiplayer games*. *Proceedings of other players*, 2004. 10
- [37] Ducheneaut, Nicolas e Robert J Moore: *The social side of gaming: a study of interaction patterns in a massively multiplayer online game*. Em *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, páginas 360–369. ACM, 2004. 10
- [38] Huffaker, D., J. Wang, J. Treem, M. A. Ahmad, L. Fullerton, D. Williams, M. S. Poole e N. Contractor: *The social behaviors of experts in massive multiplayer online role-playing games*. Em *2009 International Conference on Computational Science and Engineering*, volume 4, páginas 326–331, 2009. 10, 39
- [39] Tychsen, Anders: *Role playing games: comparative analysis across two media platforms*. Em *Proceedings of the 3rd Australasian conference on Interactive entertainment*, páginas 75–82. Murdoch University, 2006. 10, 11
- [40] Cover, Jennifer Grouling: *The creation of narrative in tabletop role-playing games*. McFarland, 2014. 10
- [41] Tychsen, Anders: *Tales for the many: Process and authorial control in multi-player role-playing games*. Em *Joint International Conference on Interactive Digital Storytelling*, páginas 309–320. Springer, 2008. 11
- [42] Crawford, Chris: *Chris Crawford on game design*. New Riders, 2003. 11
- [43] Sicart, Miguel: *Defining game mechanics*. *Game Studies*, 8(2):n, 2008. 11
- [44] Fabricatore, Carlo: *Gameplay and game mechanics: a key to quality in videogames*. Em *Expert Meeting on Videogames and Education*, 2007. 11
- [45] Technopedia: *What is a patch?* <https://www.techopedia.com/definition/24537/patch>. Acessado em: 2019-09-05. 12
- [46] Enix, Square: *Patch 5.0 notes*. <https://na.finalfantasyxiv.com/lodestone/topics/detail/330f2b280067d69d85b17831c66712a499e97484>. Acessado em: 2019-09-05. 13

- [47] Paternò, Fabio e G Faconti: *On the use of lotos to describe graphical interaction*. People and computers, páginas 155–155, 1992. 13
- [48] Wilson, Greg: *Off with their huds!: Rethinking the heads-up display in console game design*. https://www.gamasutra.com/view/feature/130948/off_with_their_huds_rethinking_.php. Acessado em: 2019-09-05. 13
- [49] Kaur, Puneet Jai *et al.*: *A survey of clustering techniques and algorithms*. Em *2015 2nd international conference on computing for sustainable global development (INDIACom)*, páginas 304–307. IEEE, 2015. 14, 17, 18, 19, 20, 22, 23
- [50] Madhulatha, T Soni: *An overview on clustering methods*. IOSR Journal of Engineering, 2012. 14, 21, 22, 23, 25, 27
- [51] Nerurkar, Pranav, Archana Shirke, Madhav Chandane e Sunil Bhirud: *Empirical analysis of data clustering algorithms*. Procedia Computer Science, 125:770–779, 2018. 14, 17, 19, 22, 23, 27, 30
- [52] Rokach, Lior e Oded Maimon: *Clustering methods*. Em *Data mining and knowledge discovery handbook*, páginas 321–352. Springer, 2005. 17, 19, 20, 21, 23, 26, 27
- [53] Priy, Surya: *Clustering in machine learning*. <https://www.geeksforgeeks.org/clustering-in-machine-learning/>. Acessado em: 2019-09-12. 18
- [54] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot e E. Duchesnay: *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12:2825–2830, 2011. 18, 19, 21, 22, 24, 29, 75, 78, 85, 86
- [55] Elavarasi, S Anitha, J Akilandeswari e B Sathiyabhama: *A survey on partition clustering algorithms*. International Journal of Enterprise Computing and Business Systems, 1(1), 2011. 19, 22, 23, 24, 25, 26, 28, 29
- [56] Bock, Tim: *What is hierarchical clustering?* <https://www.displayr.com/what-is-hierarchical-clustering/>. Acessado em: 2019-09-17. 20
- [57] Jain, Anil K, M Narasimha Murty e Patrick J Flynn: *Data clustering: a review*. ACM computing surveys (CSUR), 31(3):264–323, 1999. 21, 27, 29, 30
- [58] Kassambara, Alboukadel: *Practical guide to cluster analysis in r*. <https://www.datanovia.com/en/product/practical-guide-to-cluster-analysis-in-r/>. Acessado em: 2019-09-17. 23
- [59] Sander, Jörg, Martin Ester, Hans Peter Kriegel e Xiaowei Xu: *Density-based clustering in spatial databases: The algorithm gbscan and its applications*. Data mining and knowledge discovery, 2(2):169–194, 1998. 24
- [60] Nykjaer, Kunuk: *Clustering – grid cluster with c and html-canvas example (part 1)*. <https://kunuk.wordpress.com/2011/09/15/clustering-grid-cluster/>. Acessado em: 2019-09-17. 26

- [61] Novikov, Andrei: *PyClustering: Data mining library*. Journal of Open Source Software, 4(36):1230, apr 2019. <https://doi.org/10.21105/joss.01230>. 26, 27, 28, 30, 75
- [62] MathWorks: *Cluster with self-organizing map neural network*. <https://www.mathworks.com/help/deeplearning/ug/cluster-with-self-organizing-map-neural-network.html>. Acessado em: 2019-09-26. 28
- [63] Aoullay, Amine: *Spectral clustering for beginners*. <https://towardsdatascience.com/spectral-clustering-for-beginners-d08b7d25b4d8>. Acessado em: 2019-09-26. 29
- [64] WORLDQUANT: *Fuzzy approaches in financial modeling*. <https://www.weareworldquant.com/en/thought-leadership/fuzzy-approaches-in-financial-modeling/>. Acessado em: 2019-09-17. 30
- [65] Google: *Google scholar*. <https://scholar.google.com>. Acessado em: 2019-10-11. 34
- [66] Ryan, Richard M, C Scott Rigby e Andrew Przybylski: *The motivational pull of video games: A self-determination theory approach*. Motivation and emotion, 30(4):344–360, 2006. 35
- [67] Shaw, Adrienne: *Do you identify as a gamer? gender, race, sexuality, and gamer identity*. New Media Society, 14:28–44, fevereiro 2012. 35
- [68] Birk, Max e Regan L Mandryk: *Control your game-self: effects of controller type on enjoyment, motivation, and personality in game*. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, páginas 685–694. ACM, 2013. 35
- [69] Smyth, Joshua M: *Beyond self-selection in video game play: An experimental examination of the consequences of massively multiplayer online role-playing game play*. CyberPsychology & Behavior, 10(5):717–721, 2007. 35, 40
- [70] Hughes, Chelsea M, Brandon J Griffin e Everett L Worthington Jr: *A measure of social behavior in team-based, multiplayer online games: The sociality in multiplayer online games (smog) scale*. Computers in Human Behavior, 69:386–395, 2017. 35
- [71] Yee, Nick: *Motivations for play in online games*. CyberPsychology & behavior, 9(6):772–775, 2006. 35
- [72] Nacke, Lennart E, Chris Bateman e Regan L Mandryk: *Brainhex: A neurobiological gamer typology survey*. Entertainment computing, 5(1):55–62, 2014. 35
- [73] Klock, Ana Carolina Tomé, Thiago H Flores, Isabela Gasparini e M da S HOUNSELL: *Classificação de jogadores: um mapeamento sistemático da literatura*. Anais do Simpósio Brasileiro de Jogos e Entretenimento Digital. Porto Alegre: SBC, 2016. 36

- [74] Sweetser, Penelope e Peta Wyeth: *Gameflow: a model for evaluating player enjoyment in games*. Computers in Entertainment (CIE), 3(3):3–3, 2005. 36
- [75] Kang, Ha Na, Hye Ryeon Yong e Hyun Seok Hwang: *A study of analyzing on online game reviews using a data mining approach: Steam community data*. International Journal of Innovation, Management and Technology, 8(2):90, 2017. 36
- [76] Juul, Jesper: *The open and the closed: Games of emergence and games of progression*. Em *CGDC Conf.*, 2002. 36
- [77] Jakobsson, Mikael e TL Taylor: *The sopranos meets everquest: social networking in massively multiplayer online games*. Em *Proceedings of the 2003 Digital Arts and Culture (DAC) conference, Melbourne, Australia*, páginas 81–90, 2003. 36, 37
- [78] Pang, Sheng, Changjia Chen e Chao Xu: *Structure analysis of social network on multi-activities in mmog*. Em *Computing, Communication, Control, and Management*, páginas 147 – 152, setembro 2009. 36, 38
- [79] Maguluri, Naga Sai Nikhil: *Multi-class classification of textual data: Detection and mitigation of cheating in massively multiplayer online role playing games*. 2017. Department of Computer Science and Engineering, Wright State University. 36, 40
- [80] Chuang, Yao Chung: *Massively multiplayer online role-playing game-induced seizures: Anelected health problem in internet addiction*. CyberPsychology & Behavior, 9(4):451–456, 2006. 36
- [81] Nakamura, Lisa: *Don't hate the player, hate the game: The racialization of labor in world of warcraft*. Critical Studies in Media Communication, 26(2):128–144, 2009. 36, 38
- [82] Bergstrom, Kelly, Jennifer Jenson e Suzanne de Castell: *What's' choice'got to do with it?: avatar selection differences between novice and expert players of world of warcraft and rift*. Em *proceedings of the International Conference on the Foundations of Digital Games*, páginas 97–104. ACM, 2012. 36, 38
- [83] Szell, Michael e Stefan Thurner: *Measuring social dynamics in a massive multiplayer online game*. Social Networks, 32(4):313 – 329, 2010. 37, 40
- [84] Wang, Shih Ting, Jie Chi Yang, Sherry Y Chen e Wen Chi Kuo: *The clusters of gaming behavior in mmorpg: a case study in taiwan*. Em *2012 IIAI International Conference on Advanced Applied Informatics*, páginas 263–266. IEEE, 2012. 37
- [85] Barnett, Jane e Mark Coulson: *Virtually real: A psychological perspective on massively multiplayer online games*. Review of General Psychology, 14(2):167–179, 2010. 37
- [86] Ries, Michal, Philipp Svoboda e Markus Rupp: *Empirical study of subjective quality for massive multiplayer games*. Em *2008 15th International Conference on Systems, Signals and Image Processing*, páginas 181–184. IEEE, 2008. 37, 38

- [87] Suznjevic, Mirko, Maja Matijasevic e Ognjen Dobrijevic: *Action specific massive multiplayer online role playing games traffic analysis: Case study of world of war-craft*. Em *Proceedings of the 7th ACM SIGCOMM Workshop on Network and System Support for Games*, páginas 106–107. ACM, 2008. 37, 38
- [88] Khan, Hammad, Julien Gascon-Samson, Jörg Kienzle e Bettina Kemme: *Monitoring large-scale location-based information systems*. Em *2015 IEEE International Parallel and Distributed Processing Symposium*, páginas 1171–1181. IEEE, 2015. 37
- [89] Feng, Wu chang, David Brandt e Debanjan Saha: *A long-term study of a popular mmorpg*. Em *Proceedings of the 6th Workshop on Network and System Support for Games*, páginas 19–24, janeiro 2007. 37, 40
- [90] Montes, Fillipe Souza, Pedro Lima Nilo e Letícia Perani: *Evolução da progressão: uma análise de design de leveling em mmorpgs clássicos*. Em *Proceedings of SBGames 2017*. 37, 38
- [91] Halldorsson, Orvar e Arnar Birgir Jonsson: *Visualizing progression in eve online*. Em *ACM SIGGRAPH 2013 Studio Talks*, página 5. ACM, 2013. 37
- [92] En, Looi Qin e See Swee Lan: *Balancing safety and danger in gaming for better user engagement*. Em *2011 8th International Conference on Information, Communications & Signal Processing*, páginas 1–5. IEEE, 2011. 37
- [93] Gittens, Curtis e Jemar Greaves: *Transforming browserquest into an epidemiological tool for modelling disease dissemination*. Em *2015 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*, páginas 143–148. IEEE, 2015. 37
- [94] Daneva, Maya: *How practitioners approach gameplay requirements? an exploration into the context of massive multiplayer online role-playing games*. Em *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, páginas 3–12. IEEE, 2014. 37
- [95] Daneva, Maya: *Striving for balance: A look at gameplay requirements of massively multiplayer online role-playing games*. *Journal of Systems and Software*, 134:54–75, 2017. 37
- [96] Paraskeva, Fotini, Sofia Mysirlaki e Aikaterini Papagianni: *Multiplayer online games as educational tools: Facing new challenges in learning*. *Computers & Education*, 54(2):498–505, 2010. 38
- [97] Dickey, Michele D: *Game design and learning: A conjectural analysis of how massively multiple online role-playing games (mmorpgs) foster intrinsic motivation*. *Educational Technology Research and Development*, 55(3):253–273, 2007. 38
- [98] Voulgari, Iro e Demetrios G Sampson: *Applying lessons learnt from massively multiplayer online games (mmogs) to massive open online courses (moocs)*. Em *2014 IEEE 14th International Conference on Advanced Learning Technologies*, páginas 14–15. IEEE, 2014. 38

- [99] Malliarakis, Christos, Maya Satratzemi e Stelios Xinogalos: *Integrating learning analytics in an educational mmorpg for computer programming*. Em *2014 IEEE 14th International Conference on Advanced Learning Technologies*, páginas 233–237. IEEE, 2014. 38
- [100] Malliarakis, Christos, Maya Satratzemi e Stelios Xinogalos: *Cmx: the effects of an educational mmorpg on learning and teaching computer programming*. *IEEE Transactions on Learning Technologies*, 10(2):219–235, 2016. 38
- [101] Yu, Tao Wang: *Learning in the virtual world: the pedagogical potentials of massively multiplayer online role playing games*. *International Education Studies*, 2(1):32–38, 2009, ISSN 19139039. 38
- [102] Brignall, Iii, T. W. e T. L. Van Valey: *An online community as a new tribalism: The world of warcraft*. Em *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, páginas 179b–179b, Jan 2007. 38, 39
- [103] Klastrup, Lisbeth e Susana Tosca: *"because it just looks cool!" fashion as character performance: The case of wow*. *Journal For Virtual Worlds Research*, 2009. 38
- [104] Bergstrom, Kelly, Jennifer Jenson e Suzanne de Castell: *What's 'choice' got to do with it?: avatar selection differences between novice and expert players of world of warcraft and rift*. *Foundations of Digital Games 2012, FDG 2012 - Conference Program*, maio 2012. 38, 40
- [105] Platzer, C.: *Sequence-based bot detection in massive multiplayer online games*. Em *2011 8th International Conference on Information, Communications Signal Processing*, páginas 1–5, 2011. 38, 39, 40
- [106] Yee, Nick, Nicolas Ducheneaut, Han Tai Shiao e Les Nelson: *Through the azerethian looking glass: Mapping in-game preferences to real world demographics*. *Conference on Human Factors in Computing Systems - Proceedings*, maio 2012. 38, 40
- [107] Pang, Sheng, Changjia Chen e Yue Yang: *Classify players based on player relationship network in mmog*. Em *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, volume 4, páginas 143–146. IEEE, 2009. 38, 39
- [108] Bean, Anthony Martin: *Video gamers' personas: A five factor study exploring personality elements of the video gamer*. 2015. *Clinical Psychology*, Pacifica Graduate Institute. 38
- [109] Moreira, Victor Emanuel Montes e Vânia Ribas Ulbricht: *O perfil do jogador e as emoções no jogo*. Em *Proceedings of SBGames 2017*, 2017. 38
- [110] You, Sukkyung, Euikyung Kim e Donguk Lee: *Virtually real: exploring avatar identification in game addiction among massively multiplayer online role-playing games (mmorpg) players*. *Games and Culture*, 12(1):56–71, 2017. 38

- [111] Kim, Eun Joo, Kee Namkoong, Taeyun Ku e Se Joo Kim: *The relationship between online game addiction and aggression, self-control and narcissistic personality traits*. *European Psychiatry*, 23(3):212 – 218, 2008. 38
- [112] Hilgard, Joseph, Christopher R Engelhardt e Bruce D Bartholow: *Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (games)*. *Frontiers in psychology*, 4:608, 2013. 38
- [113] Lee, Yu Hao e Holin Lin: *‘gaming is my work’: identity work in internet-hobbyist game workers*. *Work, Employment and Society*, 25(3):451–467, 2011. 38
- [114] Kwon, Hyukmin, Aziz Mohaisen, Jiyoung Woo, Huy Kang Kim, Yongdae Kim e Eunjo Lee: *Crime scene reconstruction: Online gold farming network analysis*. *IEEE Transactions on Information Forensics and Security*, PP:1–1, novembro 2016. 39, 40
- [115] Westwood, Dave e Mark D Griffiths: *The role of structural characteristics in video-game play motivation: A q-methodology study*. *Cyberpsychology, Behavior, and Social Networking*, 13(5):581–585, 2010. 39
- [116] Lankoski, Petri: *Player character engagement in computer games*. *Games and Culture*, 6(4):291–311, 2011. 39
- [117] Lebres, Inês, Paulo Rita, Sérgio Moro e Pedro Ramos: *Factors determining player drop-out in massive multiplayer online games*. *Entertainment Computing*, 26:153 – 162, 2018. 39
- [118] Choi, Boreum, Inseong Lee, Kiho Lee, Seungki Jung, Sunju Park e Jinwoo Kim: *The effects of users’ motivation on their perception to trading systems of digital content accessories: Focusing on trading items in online games*. Em *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, páginas 161–161. IEEE, 2007. 39
- [119] Wigand, R. T., N. Agarwal, O. I. Osesina, W. Hering, M. A. Korsgaard, A. Picot e M. Drescher: *Social network indices as performance predictors in a virtual organization*. Em *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, páginas 144–149, 2012. 39, 40
- [120] Li, Quan: *A visual analytics approach for understanding egocentric intimacy network evolution and impact propagation in mmorpgs*. Em *Proceedings of IEEE Pacific Visualization Symposium*, abril 2017. 39, 40
- [121] Garnett, R., T. Gärtner, T. Ellersiek, E. Guðmondsson e P. Óskarsson: *Predicting unexpected influxes of players in eve online*. Em *2014 IEEE Conference on Computational Intelligence and Games*, páginas 1–8, 2014. 39
- [122] Ahmed, Iftekhhar, Channing Brown, Andrew Pilny, Dora Cai, Yannick Atouba Ada e Marshall Scott Poole: *Identification of groups in online environments: The twist and turns of grouping groups*. Em *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, páginas 629–632. IEEE, 2011. 39

- [123] Rosa, Maurício: *A construção de identidades online por meio do role playing game: relações com o ensino e aprendizagem de matemática em curso à distância*. 2008. 40
- [124] Emad, Sabine, Alexandra Broillet e Wade Halvorson: *Evolution of discourses along human-computer interaction throughout computer game based teaching technology: The competency building process of an individual*. Em *IEEE International Professional Communication Conference*, páginas 1–8, outubro 2012. 40
- [125] Iosup, Alexandru, Adrian Lăscăteu e Nicolae Țăpuș: *Cameo: enabling social networks for massively multiplayer online games through continuous analytics and cloud computing*. Em *Proceedings of the 9th Annual Workshop on Network and Systems Support for Games*, página 7. IEEE Press, 2010. 40
- [126] Samsurim, S, Nor Ashikin Mohamad Kamal, Marina Ismail e Norizan Diah: *Prediction outcome for massive multiplayer online games using data mining*. *Indonesian Journal of Electrical Engineering and Computer Science*, 11:248–255, julho 2018. 40
- [127] Chiang, Bowie: *A comparative study of translation strategies applied in dealing with role languages. : A translation analysis of the video game final fantasy xiv*, 2016. 43
- [128] Hutchinson, Emma Jane: *Performative identity and the embodied avatar : an online ethnography of final fantasy xiv*. Department of Sociology, University of Warwick, September 2013. 43
- [129] Korkeila, Henry e Juho Hamari: *The relationship between player’s gaming orientation and avatar’s capital: a study in final fantasy xiv*. Em *Proceedings of the 51st Annual Hawaii International Conference on System Sciences*, janeiro 2018. 44
- [130] Schreier, Jason: *What in the world is final fantasy? a beginner’s guide to the biggest rpg series on the planet*. <https://kotaku.com/what-in-the-world-is-final-fantasy-a-beginners-guide-t-5931304>. Acessado em: 2020-02-21. 45
- [131] *Final fantasy xiv, the lodestone*. <https://na.finalfantasyxiv.com/lodestone/>. Acessado em: 2019-01-02. 64, 65, 74
- [132] Hunter, J. D.: *Matplotlib: A 2d graphics environment*. *Computing in Science & Engineering*, 9(3):90–95, 2007. 75, 80
- [133] Waskom, Michael, Olga Botvinnik, Drew O’Kane, Paul Hobson, Saulius Lukauskas, David C Gempertline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, Jordi Warmenhoven, Julian de Ruitter, Cameron Pye, Stephan Hoyer, Jake Vanderplas, Santi Villalba, Gero Kunter, Eric Quintero, Pete Bachant, Marcel Martin, Kyle Meyer, Alistair Miles, Yoav Ram, Tal Yarkoni, Mike Lee Williams, Constantine Evans, Clark Fitzgerald, Brian, Chris Fonnesbeck, Antony Lee e Adel Qalieh: *mwaskom/seaborn: v0.8.1 (september 2017)*, setembro 2017. <https://doi.org/10.5281/zenodo.883859>. 75

- [134] McKinney, Wes: *Data structures for statistical computing in python*. Em Walt, Stéfan van der e Jarrod Millman (editores): *Proceedings of the 9th Python in Science Conference*, páginas 51 – 56, 2010. 75
- [135] Shetye, Abhini: *Feature selection with sklearn and pandas*. <https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b>. Acessado em: 2019-10-23. 80
- [136] Services, Select Statistical: *Population proportion – sample size*. <https://select-statistics.co.uk/calculators/sample-size-calculator-population-proportion/b>. Acessado em: 2019-11-24. 81
- [137] Daniel, Wayne W e Chad L Cross: *Biostatistics: a foundation for analysis in the health sciences*. Wiley, 2018. 81
- [138] pythonprogramminglanguage.com: *kmeans elbow method*. <https://pythonprogramminglanguage.com/kmeans-elbow-method/>. Acessado em: 2019-10-24. 82, 83
- [139] Banerjee, Amit e Rajesh N Dave: *Validating clusters using the hopkins statistic*. Em *2004 IEEE International conference on fuzzy systems (IEEE Cat. No. 04CH37542)*, volume 1, páginas 149–153. IEEE, 2004. 83
- [140] Cengizler, Caglar e M Kerem Un: *Evaluation of calinski-harabasz criterion as fitness measure for genetic algorithm based segmentation of cervical cell nuclei*. *Journal of Advances in Mathematics and Computer Science*, páginas 1–13, 2017. 85
- [141] Davies, David L e Donald W Bouldin: *A cluster separation measure*. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227, 1979. 85, 86

Anexo I

Tabelas de Clusters Resultantes

Descrição do Total do mes 06/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hidn	PreHw	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1387	2441927	1480559	2543065	1048516	1654331	1143634	5576717	2244248	2.479	0.0367	0.0609	0.0351	0
std	0.3456	10840264	7161737	11508801	4647759	7664238	5022448	30616030	11679457	6.707	0.188	0.2392	0.1841	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6000	1	6000	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1297	7259970	3550226	6139657	4192009	4203903	2592833	12720647	5483790	2.99	0.0607	0.0379	0.0375	0
std	0.336	38169851	20136637	32560825	22397964	23194900	13795610	83885318	33498861	7.442	0.2388	0.1909	0.1901	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1306	7629970	3614825	6348837	4294207	4516334	2761066	14271028	5981822	3.034	0.0613	0.0418	0.0428	0
std	0.337	40452326	20558580	33555898	22896913	24148377	14163579	90680055	35380006	7.565	0.2399	0.2002	0.2025	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1351	7935567	3651839	6534988	4669566	4738572	2878978	14181353	6039410	3.035	0.0606	0.0403	0.04	0
std	0.3418	41319019	21110792	34036050	24210107	24779474	14885718	90256193	35268049	7.579	0.2386	0.1967	0.196	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	16200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 10/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1304	7557094	3522055	6226623	4341056	4374566	2660517	12656380	5715429	2.975	0.0589	0.0408	0.0413	0
std	0.3367	38768534	20344942	32499076	22784600	23386667	14068433	83163366	34153787	7.413	0.2354	0.1979	0.199	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1279	7937447	3559050	6566242	4437896	4578934	2619223	13518735	5586276	2.963	0.0603	0.0388	0.0386	0
std	0.334	41057228	20075740	33708282	23286584	24624104	13972548	87562401	33789600	7.407	0.238	0.193	0.1928	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6001	1	10200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Total do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1317	8068853	3610215	6645114	4584555	4633376	2836276	14193091	6130719	3.081	0.0626	0.0416	0.0426	0
std	0.3382	40709801	20484119	34096791	23708843	24246243	14416911	88823357	35248127	7.58	0.2422	0.1997	0.2021	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.1: Descrição do domínio clusterizado em cada mês

Descrição do Cluster 1 criado pela Clusterização K-means

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	14562.0	14562	14562	14562	14562	14562	14562	14562	14562	14562.0	14562.0	14562.0	14562.0	14562
mean	0.0	485225	266063	490327	269408	311133	236494	672609	313912	1.287	0.0044	0.0246	0.0041	0
std	0.0	3085160	1977879	3227232	1750177	2233436	1741086	6453240	2912640	4.283	0.0662	0.1549	0.0635	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	68598000	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4

Descrição do Cluster 2 criado pela Clusterização K-means

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	1883.0	1883	1883	1883	1883	1883	1883	1883	1883	1883.0	1883.0	1883.0	1883.0	1883
mean	1.0	5035956	2997651	4624428	2174807	3180459	2316692	6954795	3247230	7.451	0.0589	0.1535	0.0738	0
std	0.0	11228121	7999735	11310337	6016937	8115622	6242165	25337617	10337337	9.232	0.2356	0.3605	0.2615	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	60880	2000	45280	1	6000	1	2100	2	3.0	0.0	0.0	0.0	0
75%	1.0	2858019	1480180	2267080	786980	1488280	454180	779620	474735	12.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 3 criado pela Clusterização K-means

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	555.0	555	555	555	555	555	555	555	555	555.0	555.0	555.0	555.0	555
mean	0.8559	44980554	28199150	49340844	17669342	31719125	20965090	129574362	49489177	22.987	0.809	0.7009	0.7189	2
std	0.3516	31479744	22176936	32193267	12740248	22060957	11918791	97181358	35516590	8.206	0.3934	0.4583	0.4499	2
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	19262280	6709190	19670200	3160200	9051680	7734000	31989330	13657940	15.0	1.0	0.0	0.0	0
50%	1.0	38594380	30257780	47725200	29470800	30788480	29470800	121838100	50898280	21.0	1.0	1.0	1.0	1
75%	1.0	84754800	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Figura I.2: Descrição dos Cluster Resultantes do K-means no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização K-means do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14708.0	14708	14708	14708	14708	14708	14708	14708	14708	14708.0	14708.0	14708.0	14708.0	14708
mean	0.0	1159221	498440	818201	657438	532057	400915	979133	542154	1.446	0.0284	0.0068	0.0063	0
std	0.0	9336128	4468018	6581716	5744779	5097560	3758666	11581196	6393064	4.664	0.166	0.0822	0.0793	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	2	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	302090580	171033600	233968400	200027600	200027600	100013800	800110400	256330400	32.0	1.0	1.0	1.0	5

Descrição do Cluster 2 criado pela Clusterização K-means do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	561.0	561	561	561	561	561	561	561	561	561.0	561.0	561.0	561.0	561
mean	0.8449	145123921	74834175	128622895	85097566	88970796	53371641	324356386	133404769	27.312	0.656	0.8128	0.7344	2
std	0.3623	129376541	70278868	108403797	74849283	77107704	43124393	317033891	118171602	7.506	0.4755	0.3904	0.442	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	31940060	9051680	28550480	10483300	9915680	7734000	33293340	19397380	21.0	0.0	1.0	0.0	0
50%	1.0	109927560	58941600	104126560	73769100	81074980	29470800	220646400	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	222707200	122169400	214378400	138827600	200027600	100013800	712688400	263325399	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1731.0	1731	1731	1731	1731	1731	1731	1731	1731	1731.0	1731.0	1731.0	1731.0	1731
mean	1.0	14382660	6378283	11659510	8003942	7930817	4760255	11488146	6014073	8.224	0.1427	0.0508	0.0768	0
std	0.0	40325718	21927733	33306723	24330369	26121702	16318522	52927894	25385698	10.075	0.3499	0.2197	0.2664	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	3700	45281	900	6000	1	2101	2	3.0	0.0	0.0	0.0	0
75%	1.0	8310600	1646309	5577360	7734000	2049180	454180	1249840	695410	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Figura I.3: Descrição dos Cluster Resultantes do K-means no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização K-means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	14705.0	14705	14705	14705	14705	14705	14705	14705	14705	14705.0	14705.0	14705.0	14705.0	14705
mean	0.0	962620	429749	845101	558224	484083	391698	899637	439178	1.395	0.0263	0.0056	0.0067	0
std	0.0	7636633	4301218	7005257	4936369	4447680	3751428	14296535	5487802	4.594	0.1601	0.0749	0.0814	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	316609800	129484600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	5

Descrição do Cluster 2 criado pela Clusterização K-means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	658.0	658	658	658	658	658	658	658	658	658.0	658.0	658.0	658.0	658
mean	0.8875	143635716	69763153	117717781	78518480	86076364	49820259	314922325	127542667	26.695	0.6672	0.8404	0.7766	2
std	0.3162	135198511	70249912	109032468	75431966	77296678	42816400	325034764	120281937	7.694	0.4716	0.3665	0.4168	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	27017345	7734000	16785680	8540235	9051680	7734000	22902820	9759660	21.0	0.0	1.0	1.0	0
50%	1.0	107245239	38970500	88412400	40131480	63411600	29470800	187891770	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	222593700	108611800	205055925	129484600	163311600	100013800	777547650	271047400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	1637.0	1637	1637	1637	1637	1637	1637	1637	1637	1637.0	1637.0	1637.0	1637.0	1637
mean	1.0	12853918	5637396	11023036	8019345	7954177	5129176	13536602	6908855	8.241	0.1319	0.0458	0.0727	0
std	0.0	35662367	19706540	34669242	24253873	25467377	16581760	56687883	26556995	10.235	0.3385	0.2091	0.2597	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	60880	1	37180	2000	6000	1	3703	1	3.0	0.0	0.0	0.0	0
75%	1.0	7759551	1448260	3932200	3303300	1845680	786980	1491261	797180	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.4: Descrição dos Cluster Resultantes do K-means no mês 02/2018

Descrição do Cluster 1 criado pela Clusterização K-means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	614.0	614	614	614	614	614	614	614	614	614.0	614.0	614.0	614.0	614
mean	0.8713	154067213	75926009	129460865	92334125	96129158	58028079	323149311	131567758	27.207	0.6173	0.8306	0.7443	2
std	0.3351	138457900	74317586	109149965	76287746	77078464	42934982	325490194	118630914	7.472	0.4865	0.3754	0.4366	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	29629174	7734000	30945660	15468000	18828825	8862600	33728780	15568630	21.0	0.0	1.0	0.0	0
50%	1.0	110172510	44526490	101840940	100013800	100013800	63297800	182112000	90174750	32.0	1.0	1.0	1.0	1
75%	1.0	252850495	127655800	219188000	200027600	200027600	100013800	800110400	286817150	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização K-means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	14625.0	14625	14625	14625	14625	14625	14625	14625	14625	14625.0	14625.0	14625.0	14625.0	14625
mean	0.0	974683	409123	827015	585795	552912	366425	954629	463686	1.404	0.0265	0.0059	0.0066	0
std	0.0	7796839	3751068	7498082	5366549	5137444	3701757	11853591	5515494	4.659	0.1605	0.0769	0.0812	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	297266200	107747800	300041400	200027600	200027600	100013800	728084400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	1761.0	1761	1761	1761	1761	1761	1761	1761	1761	1761.0	1761.0	1761.0	1761.0	1761
mean	1.0	14794224	5382891	11079344	8019438	7635481	4516989	16302031	8578060	8.15	0.1499	0.05	0.0716	0
std	0.0	37523065	18740359	31769536	23748110	23962826	16373249	76495925	33124343	10.142	0.3571	0.2179	0.2578	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	1	130580	900	3700	1	3600	1	3.0	0.0	0.0	0.0	0
75%	1.0	8630780	1362960	3953040	7734000	2009460	390280	983639	786980	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Figura I.5: Descrição dos Cluster Resultantes do K-means no mês 06/2018

Descrição do Cluster 1 criado pela Clusterização K-means do mes 10/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14686.0	14686	14686	14686	14686	14686	14686	14686	14686	14686.0	14686.0	14686.0	14686.0	14686
mean	0.0	1130550	450432	792695	540026	498901	359183	802309	429165	1.363	0.0247	0.0059	0.0076	0
std	0.0	9761938	4034504	7279030	4483246	4654882	3327680	10201822	5823587	4.422	0.1553	0.0763	0.087	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	109670200	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização K-means do mes 10/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	637.0	637	637	637	637	637	637	637	637	637.0	637.0	637.0	637.0	637
mean	0.8462	137284674	68364840	117493306	83880789	87817159	51332637	284375346	122614148	26.031	0.6248	0.8273	0.7347	2
std	0.3611	127511218	7194983	105437359	74396952	74702103	43839530	308458967	116697035	8.296	0.4846	0.3783	0.4418	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	25251180	7734000	20214480	9214180	13127699	7734000	18388320	16159780	21.0	0.0	1.0	0.0	0
50%	1.0	105568580	35258480	101035380	79596800	79409600	29470800	150468000	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	208641800	108611800	200814580	129484600	149952600	100013800	569413400	249966400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 10/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1677.0	1677	1677	1677	1677	1677	1677	1677	1677	1677.0	1677.0	1677.0	1677.0	1677
mean	1.0	14559928	5790988	11549097	7415062	6619695	4326137	13255011	7605462	8.335	0.1431	0.0483	0.0727	0
std	0.0	37843641	20130510	32204961	21173975	21375318	15216067	61944417	30397361	10.316	0.3503	0.2145	0.2598	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	113960	2000	78760	1	16199	1	2402	2	3.0	0.0	0.0	0.0	0
75%	1.0	7930480	1448260	4271260	7734000	1845680	454180	1405140	608381	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.6: Descrição dos Cluster Resultantes do K-means no mês 10/2018

Descrição do Cluster 1 criado pela Clusterização K-means do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14724.0	14724	14724	14724	14724	14724	14724	14724	14724	14724.0	14724.0	14724.0	14724.0	14724
mean	0.0	1126768	482324	860578	593554	652207	375617	1233021	639431	1.373	0.0276	0.006	0.0072	0
std	0.0	9978229	4534859	7588165	5012771	6311552	3569194	17190686	7738536	4.461	0.164	0.0775	0.0845	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	200027600	300041400	163311600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização K-means do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	582.0	582	582	582	582	582	582	582	582	582.0	582.0	582.0	582.0	582
mean	0.8265	158656668	75945917	137211728	89052552	94406402	54776001	316716566	126636750	26.316	0.6031	0.8316	0.7388	2
std	0.379	134181122	69597849	104671632	75630026	80030475	43720054	320519939	118986195	8.109	0.4897	0.3745	0.4396	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	39836125	9051680	40536910	12106225	9915680	7734000	29959265	14457829	21.0	0.0	1.0	0.0	0
50%	1.0	112228830	58941600	112107030	100013800	96391200	33940800	187477390	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	258969200	114953349	228776150	156316600	200027600	100013800	765748900	259248650	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1694.0	1694	1694	1694	1694	1694	1694	1694	1694	1694.0	1694.0	1694.0	1694.0	1694
mean	1.0	15352947	5431875	11273749	8781680	7847849	4201050	16136038	6994755	8.767	0.1576	0.0508	0.0714	0
std	0.0	40167878	17846508	31230103	25585451	25001753	14563063	78973048	28851026	10.585	0.3645	0.2196	0.2576	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	600	130580	1	6000	1	1652	1	3.0	0.0	0.0	0.0	0
75%	1.0	8598000	1448260	3409604	7734000	1850840	333680	1128933	552515	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	5

Figura I.7: Descrição dos Cluster Resultantes do K-means no mês 02/2019

Descrição do Cluster 1 criado pela Clusterização K-means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14660.0	14660	14660	14660	14660	14660	14660	14660	14660	14660.0	14660.0	14660.0	14660.0	14660
mean	0.0	1150924	442341	868744	591800	567403	445511	925743	510313	1.388	0.0262	0.0058	0.0078	0
std	0.0	9273386	4203569	7648636	5138392	5319354	4391709	11213269	5862950	4.532	0.1597	0.0759	0.0882	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	2	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	303058500	200027600	300041400	133954599	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização K-means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	636.0	636	636	636	636	636	636	636	636	636.0	636.0	636.0	636.0	636
mean	0.8412	149807436	71454754	124948649	87711690	90791786	51196371	314302813	131649531	26.796	0.6682	0.8302	0.772	2
std	0.3658	131211780	71509076	108117281	76494612	76766529	42959909	315982187	116235622	7.647	0.4712	0.3758	0.4199	2
min	0.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
25%	1.0	32088720	7730420	27224535	9755535	10880475	7734000	31945449	22649310	21.0	0.0	1.0	1.0	0
50%	1.0	114246755	43975700	102069930	73600600	81081680	29470800	203852190	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	242581495	112492600	207761600	149952600	200027600	100013800	698015200	256330400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização K-means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1704.0	1704	1704	1704	1704	1704	1704	1704	1704	1704.0	1704.0	1704.0	1704.0	1704
mean	1.0	14683343	5542084	12185340	7909047	7456394	5354820	16323096	7636109	8.804	0.1496	0.0552	0.0698	0
std	0.0	37742000	17901131	34609129	21744484	22463291	17075121	72698295	30166492	10.365	0.3568	0.2284	0.2549	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	10200	104180	10200	16199	1	10200	2	3.0	0.0	0.0	0.0	0
75%	1.0	8730080	1910510	4823120	7734000	2049180	814430	1924625	1163730	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.8: Descrição dos Cluster Resultantes do K-means no mês 06/2019

Descrição do Cluster 1 criado pela Clusterização WARD

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	2475.0	2475	2475	2475	2475	2475	2475	2475	2475	2475.0	2475.0	2475.0	2475.0	2475
mean	0.7996	6425904	3674003	6366839	2626945	3950928	3077781	9685892	4357234	9.864	0.1232	0.3046	0.1022	0
std	0.4004	13640090	9222135	14944538	6780542	9345414	7515619	31939095	12785679	10.372	0.3288	0.4604	0.303	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	1.0	0.0	0.0	0.0	0
50%	1.0	100440	3700	78760	1	16200	1	3700	3	6.0	0.0	0.0	0.0	0
75%	1.0	4928860	1771860	3690149	786980	2049179	786980	1320582	691780	15.0	0.0	1.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização WARD

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	14103.0	14103	14103	14103	14103	14103	14103	14103	14103	14103.0	14103.0	14103.0	14103.0	14103
mean	0.0	344969	199035	357839	216304	227899	183347	466004	214206	0.812	0.0	0.0	0.0	0
std	0.0	2333246	1518740	2320884	1399340	1702030	1413058	4466805	1862846	2.884	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	68598000	37204800	59963180	29470800	58941600	29470800	173372080	58941600	32.0	0.0	0.0	0.0	4

Descrição do Cluster 3 criado pela Clusterização WARD

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	422.0	422	422	422	422	422	422	422	422	422.0	422.0	422.0	422.0	422
mean	0.8981	49155337	31443991	53145903	19603160	35855502	21892255	152273818	57694550	22.929	0.7559	0.6682	0.8152	2
std	0.3029	32267766	22598833	32289262	12390250	22133037	11631898	92947838	34405484	8.627	0.43	0.4714	0.3886	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	23202000	9051680	29642965	7734000	12889100	8862600	58941600	29470800	15.0	1.0	0.0	1.0	0
50%	1.0	44938800	30788480	59395780	29470800	35710300	29470800	202855200	72039600	21.0	1.0	1.0	1.0	1
75%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Figura I.9: Descrição dos Cluster Resultantes do WARD no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização WARD do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1871.0	1871	1871	1871	1871	1871	1871	1871	1871	1871.0	1871.0	1871.0	1871.0	1871
mean	0.7969	8268067	3611119	6777367	4967228	4630933	1995574	6877590	3333862	9.61	0.3201	0.0	0.0	0
std	0.4024	26176965	14473416	23670243	17583216	17730118	8355428	35580770	16304336	10.994	0.4667	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	6000	0	901	0	1	0	0	0	6.0	0.0	0.0	0.0	0
75%	1.0	3570170	454180	1464450	786980	896780	130580	334130	130580	15.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	149952600	200027600	100013800	768865400	245504400	32.0	1.0	0.0	0.0	4

Descrição do Cluster 2 criado pela Clusterização WARD do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14155.0	14155	14155	14155	14155	14155	14155	14155	14155	14155.0	14155.0	14155.0	14155.0	14155
mean	0.0	815455	347159	563271	492311	381183	286677	694798	358510	0.884	0.0	0.0	0.0	0
std	0.0	8027347	3470951	4945832	4812683	4377792	3083961	9933055	5088436	3.138	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	363339200	105016700	201683660	200027600	200027600	100013800	800110400	256330400	32.0	0.0	0.0	0.0	5

Descrição do Cluster 3 criado pela Clusterização WARD do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	974.0	974	974	974	974	974	974	974	974	974.0	974.0	974.0	974.0	974
mean	0.7331	98920333	49982954	85955448	56470032	58938626	37255156	198714739	84098636	20.874	0.4446	0.6612	0.655	1
std	0.4426	119092457	64797920	101290054	69980755	72831598	42091719	286826899	110181403	11.096	0.4972	0.4735	0.4756	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	8188181	1554660	2984595	1317680	1478760	196480	1056481	296355	12.0	0.0	0.0	0.0	0
50%	1.0	40404880	12570340	34581540	16568400	16980750	14350799	35800860	22299120	21.0	0.0	1.0	1.0	0
75%	1.0	153739685	100013800	135167055	102513200	102513200	100013800	257311150	143209750	32.0	1.0	1.0	1.0	2
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.10: Descrição dos Cluster Resultantes do WARD no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização WARD do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	2514.0	2514	2514	2514	2514	2514	2514	2514	2514	2514.0	2514.0	2514.0	2514.0	2514
mean	0.7621	20481270	9455181	17816865	11672080	12621923	8466247	25073525	11746257	12.095	0.3365	0.1798	0.1925	0
std	0.4259	47219634	26507123	44766943	29982839	33284652	21860088	78734664	33871281	11.665	0.4726	0.3841	0.3944	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	1.0	0.0	0.0	0.0	0
50%	1.0	161530	10200	196480	60880	130580	2000	23550	2000	9.0	0.0	0.0	0.0	0
75%	1.0	13984465	3303300	9214180	7734000	5746280	3608200	7165986	3046680	21.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização WARD do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14166.0	14166	14166	14166	14166	14166	14166	14166	14166	14166.0	14166.0	14166.0	14166.0	14166
mean	0.0	594079	286013	530935	368006	322175	218361	478488	244430	0.852	0.0	0.0	0.0	0
std	0.0	5609583	3196485	4522147	3346814	3175153	1947149	5058322	2644306	3.104	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	316609800	129484600	155346300	129484600	101331480	100013800	235766400	117369180	32.0	0.0	0.0	0.0	5

Descrição do Cluster 3 criado pela Clusterização WARD do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	320.0	320	320	320	320	320	320	320	320	320.0	320.0	320.0	320.0	320
mean	0.9531	218137026	105093872	173804368	120139755	126506947	70502071	539982420	214682150	28.441	0.6125	0.8094	0.7625	2
std	0.2117	142430170	73774154	112419496	76042171	75314952	40128932	318756906	109266254	6.83	0.4879	0.3934	0.4262	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	101973109	31421705	79850085	41354700	58941600	29470800	217429800	144194535	32.0	0.0	1.0	1.0	0
50%	1.0	201362840	101961229	175627900	129484600	129484600	100013800	784487900	284418900	32.0	1.0	1.0	1.0	2
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.11: Descrição dos Cluster Resultantes do WARD no mês 02/2018

Descrição do Cluster 1 criado pela Clusterização WARD do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	2547.0	2547	2547	2547	2547	2547	2547	2547	2547	2547.0	2547.0	2547.0	2547.0	2547
mean	0.775	20902797	8568904	16267397	11704399	11779665	6884747	24277395	12890882	11.665	0.3168	0.1429	0.1665	0
std	0.4176	50139070	25350221	40197935	30621039	31008815	20109378	85707176	40151990	11.61	0.4653	0.3501	0.3726	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	1.0	0.0	0.0	0.0	0
50%	1.0	138580	6000	196480	10200	60880	1	10200	300	9.0	0.0	0.0	0.0	0
75%	1.0	12903445	2267080	8667290	7734000	5002900	1480180	3755630	2330350	21.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização WARD do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14094.0	14094	14094	14094	14094	14094	14094	14094	14094	14094.0	14094.0	14094.0	14094.0	14094
mean	0.0	628935	289925	545404	379364	334396	196920	566297	285272	0.834	0.0	0.0	0.0	0
std	0.0	5195362	2998413	5604592	3542086	3685210	1685848	6246074	3893470	3.023	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	149962680	107747800	300041400	133954599	200027600	56302799	250992200	300041400	32.0	0.0	0.0	0.0	4

Descrição do Cluster 3 criado pela Clusterização WARD do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	359.0	359	359	359	359	359	359	359	359	359.0	359.0	359.0	359.0	359
mean	0.8969	202788322	100752281	172631780	123188765	127687843	79754264	477067117	183332249	28.192	0.6212	0.8942	0.7131	2
std	0.3045	144061267	78521838	111926023	74204961	74214541	34786585	331882156	117648377	7.36	0.4858	0.3081	0.4529	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	80409930	28166789	88412400	45477990	58941600	63297800	115566850	70763300	32.0	0.0	1.0	0.0	0
50%	1.0	200061080	101331480	183893400	129484600	133954599	100013800	552219400	207441400	32.0	1.0	1.0	1.0	2
75%	1.0	349980200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.12: Descrição dos Cluster Resultantes do WARD no mês 06/2018

Descrição do Cluster 1 criado pela Clusterização WARD do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1775.0	1775	1775	1775	1775	1775	1775	1775	1775	1775.0	1775.0	1775.0	1775.0	1775
mean	0.8152	8632396	3309763	7263212	4570998	4121000	1758674	6580077	3920658	9.632	0.3025	0.0	0.0	0
std	0.3882	28430536	14345334	24775690	16032813	16137724	5929197	26876671	18817210	11.177	0.4595	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	2900	0	2000	0	1	0	0	0	3.0	0.0	0.0	0.0	0
75%	1.0	2499400	390280	1514610	786980	896780	130580	392062	136420	15.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	138827600	200027600	63297800	258307000	300041400	32.0	1.0	0.0	0.0	7

Descrição do Cluster 2 criado pela Clusterização WARD do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14162.0	14162	14162	14162	14162	14162	14162	14162	14162	14162.0	14162.0	14162.0	14162.0	14162
mean	0.0	696190	287352	446925	382425	366663	264549	521065	255002	0.846	0.0	0.0	0.0	0
std	0.0	6152823	2812117	3602471	3109044	3937578	2597531	5879263	2930793	2.929	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	249966400	109670200	134648200	129484600	200027600	100013800	414534400	171510400	32.0	0.0	0.0	0.0	4

Descrição do Cluster 3 criado pela Clusterização WARD do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1063.0	1063	1063	1063	1063	1063	1063	1063	1063	1063.0	1063.0	1063.0	1063.0	1063
mean	0.7234	97167125	46971432	81496739	56696637	58193946	36087110	184477423	81459819	20.236	0.4365	0.6529	0.6604	1
std	0.4475	116343137	64396890	97925475	69271949	70705379	42697103	278317366	108718758	11.113	0.4962	0.4763	0.4738	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	8195599	1448260	2749300	1480180	1483180	130580	740941	220159	12.0	0.0	0.0	0.0	0
50%	1.0	44938800	9915680	32618480	15468000	22084800	8598000	30199481	23202000	21.0	0.0	1.0	1.0	0
75%	1.0	141017900	77180800	125463990	105611600	101961229	100013800	235766400	130204790	32.0	1.0	1.0	1.0	2
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.13: Descrição dos Cluster Resultantes do WARD no mês 10/2018

Descrição do Cluster 1 criado pela Clusterização WARD do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	2488.0	2488	2488	2488	2488	2488	2488	2488	2488	2488.0	2488.0	2488.0	2488.0	2488
mean	0.7524	23355977	9671247	18851095	11499732	11104989	5680219	25522931	12027007	11.9	0.3493	0.1612	0.1728	0
std	0.4317	55308043	26792085	46355732	29571477	29426984	15851323	89570829	38236826	11.521	0.4768	0.3678	0.3782	1
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	1.0	0.0	0.0	0.0	0
50%	1.0	174910	10200	162080	16199	130580	2000	6900	900	9.0	0.0	0.0	0.0	0
75%	1.0	15468000	3308940	9081805	7734000	4945135	2049180	4501510	1938705	21.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização WARD do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14184.0	14184	14184	14184	14184	14184	14184	14184	14184	14184.0	14184.0	14184.0	14184.0	14184
mean	0.0	724986	316254	556167	468207	443594	266062	828633	399527	0.854	0.0	0.0	0.0001	0
std	0.0	7779045	3663376	5433237	4728569	5565553	3198196	15109865	6392944	3.091	0.0	0.0	0.0084	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	200027600	263325399	200027600	200027600	100013800	800110400	300041400	32.0	0.0	0.0	1.0	4

Descrição do Cluster 3 criado pela Clusterização WARD do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	328.0	328	328	328	328	328	328	328	328	328.0	328.0	328.0	328.0	328
mean	0.9238	202877238	97426974	173280227	122536144	133904692	81160321	471232025	181026237	26.393	0.4756	0.7866	0.689	2
std	0.2658	139848981	73018699	104165437	74270234	75823231	35991722	330803823	118763231	8.762	0.5002	0.4103	0.4636	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	87944100	30788479	101331480	43353600	68183850	100013800	110415749	62399534	21.0	0.0	1.0	0.0	0
50%	1.0	188426200	101331480	161190400	129484600	171033600	100013800	539650100	217485700	32.0	0.0	1.0	1.0	2
75%	1.0	340184200	172869100	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.14: Descrição dos Cluster Resultantes do WARD no mês 02/2019

Descrição do Cluster 1 criado pela Clusterização WARD do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	2428.0	2428	2428	2428	2428	2428	2428	2428	2428	2428.0	2428.0	2428.0	2428.0	2428
mean	0.7727	21648272	8244912	18640497	12488308	12238016	8900718	28654103	13351201	11.926	0.3266	0.1404	0.1318	0
std	0.4192	49767417	23222984	46330804	31813478	32374309	23377996	100628695	41840272	11.475	0.4691	0.3475	0.3383	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
50%	1.0	196480	39380	196480	45280	130580	2000	23554	900	9.0	0.0	0.0	0.0	0
75%	1.0	15468000	3344300	9060050	7734000	7544595	2749300	7379315	3305850	21.0	1.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização WARD do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14186.0	14186	14186	14186	14186	14186	14186	14186	14186	14186.0	14186.0	14186.0	14186.0	14186
mean	0.0	824524	316430	566385	415810	372720	254551	624328	333389	0.89	0.0	0.0	0.0054	0
std	0.0	7519296	3436520	5462904	3762615	3803767	2219691	9704062	4893777	3.158	0.0	0.0	0.0735	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	303058500	200027600	249966400	129484600	171033600	79594800	800110400	300041400	32.0	0.0	0.0	1.0	6

Descrição do Cluster 3 criado pela Clusterização WARD do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	386.0	386	386	386	386	386	386	386	386	386.0	386.0	386.0	386.0	386
mean	0.9404	188990666	95508109	154593471	108075493	113383639	59571738	421900673	173772134	27.987	0.7021	0.9482	0.8497	3
std	0.237	139999688	75428355	111743038	78388042	76978696	42959565	326017401	114871615	6.869	0.4579	0.2219	0.3578	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	62577900	20973900	46548000	29470800	37204800	12478800	97356300	68914800	21.0	0.0	1.0	1.0	0
50%	1.0	166689400	100432600	131066280	107747800	105229349	100013800	330739900	200310280	32.0	0.0	1.0	1.0	2
75%	1.0	311550000	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.15: Descrição dos Cluster Resultantes do WARD no mês 06/2019

Descrição do Cluster 1 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	14083.0	14083	14083	14083	14083	14083	14083	14083	14083	14083.0	14083.0	14083.0	14083.0	14083
mean	0.0	319023	179171	338074	200573	211539	163951	422413	188567	0.788	0.0	0.0	0.0	0
std	0.0	2069312	1331107	2167533	1262828	1524647	1202745	3945765	1578483	2.801	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	68598000	30788480	59963180	29470800	33079000	29470800	114691200	33136800	32.0	0.0	0.0	0.0	4

Descrição do Cluster 2 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	318.0	318	318	318	318	318	318	318	318	318.0	318.0	318.0	318.0	318
mean	0.0	1441829	721696	903639	354548	987505	500818	1140558	504559	18.091	0.0	1.0	0.0	0
std	0.0	4679479	3365676	3444793	1637457	3913885	2045029	6615944	2819867	10.355	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	12.0	0.0	1.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	21.0	0.0	1.0	0.0	0
75%	0.0	148730	2000	130579	1	33480	6000	5	0	32.0	0.0	1.0	0.0	0
max	0.0	31737880	30788480	29677160	14350800	30788480	19160400	61872000	28816800	32.0	0.0	1.0	0.0	2

Descrição do Cluster 3 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	1343.0	1343	1343	1343	1343	1343	1343	1343	1343	1343.0	1343.0	1343.0	1343.0	1343
mean	1.0	2784648	1430830	2839447	1532192	1765959	1141817	3204896	1504703	3.863	0.0	0.0	0.0	0
std	0.0	6840274	4247968	7791989	4670589	5088747	3496268	10790480	4594388	5.734	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	3700	0	901	0	0	0	1	0	2.0	0.0	0.0	0.0	0
75%	1.0	1274070	282680	1317680	333680	282680	130580	316755	84780	6.0	0.0	0.0	0.0	0
max	1.0	59102680	37204800	88412400	29470800	39933600	29470800	67250400	30795300	32.0	0.0	0.0	0.0	4

Descrição do Cluster 4 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	201.0	201	201	201	201	201	201	201	201	201.0	201.0	201.0	201.0	201
mean	1.0	3619663	1505562	2879048	681098	1956901	1343311	3044870	1183454	21.955	0.0	1.0	0.0	0
std	0.0	9725091	4627654	7305496	2045975	6398456	3656301	8986912	3783783	10.059	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	15.0	0.0	1.0	0.0	0
50%	1.0	3700	0	6000	1	300	0	1	1	21.0	0.0	1.0	0.0	0
75%	1.0	520860	390280	1480180	282680	390280	161080	429160	234759	32.0	0.0	1.0	0.0	0
max	1.0	67539600	30788480	40919980	16568400	46039200	25582800	62736000	30854360	32.0	0.0	1.0	0.0	2

Descrição do Cluster 5 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	61.0	61	61	61	61	61	61	61	61	61.0	61.0	61.0	61.0	61
mean	1.0	7620530	4807161	5665324	1508611	4967778	1026358	2234123	2732254	7.525	0.0	0.0	1.0	0
std	0.0	13174233	9415644	9370605	3681921	9718874	3034816	8098733	6915331	6.7	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	1.0	3700	0	0	0	1	0	0	0	3.0	0.0	0.0	1.0	0
50%	1.0	291660	161080	1152160	2000	236980	1	9700	901	6.0	0.0	0.0	1.0	0
75%	1.0	8816060	4274400	8728580	786980	3584760	130580	390280	196480	12.0	0.0	0.0	1.0	0
max	1.0	59546280	30788480	32424380	22155600	31519980	16568400	52376482	29494351	21.0	0.0	0.0	1.0	1

Descrição do Cluster 6 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	27.0	27	27	27	27	27	27	27	27	27.0	27.0	27.0	27.0	27
mean	1.0	88412400	58312266	88412400	29470800	58941600	29470800	235766400	87997466	29.556	1.0	1.0	1.0	5
std	0.0	0	3270111	0	0	0	0	0	2156056	4.66	0.0	0.0	0.0	2
min	1.0	88412400	41949600	88412400	29470800	58941600	29470800	235766400	77209200	21.0	1.0	1.0	1.0	1
25%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4
50%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	6
75%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 7 criado pela Clusterização DBSCAN

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	25.0	25	25	25	25	25	25	25	25	25.0	25.0	25.0	25.0	25
mean	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
std	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
75%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
max	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0

Figura I.16: Descrição dos Cluster Resultantes do DBSCAN no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	1264.0	1264	1264	1264	1264	1264	1264	1264	1264	1264.0	1264.0	1264.0	1264.0	1264
mean	1.0	7338460	2777723	6106168	4307826	3613917	1481206	4957021	2483038	4.499	0.0	0.0	0.0	0
std	0.0	21878144	10963045	20126174	15047219	14041143	5421524	19880271	9564626	6.878	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	10200	1	2	0	1	0	1	0	1.0	0.0	0.0	0.0	0
75%	1.0	3937410	525580	1488280	1317680	786980	130580	454180	182080	6.0	0.0	0.0	0.0	0
max	1.0	202076780	110937399	202076780	133954599	119174200	79594800	249176400	117820400	32.0	0.0	0.0	0.0	4

Descrição do Cluster 2 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	14135.0	14135	14135	14135	14135	14135	14135	14135	14135	14135.0	14135.0	14135.0	14135.0	14135
mean	0.0	682234	297052	484344	393698	291663	221651	521980	238276	0.866	0.0	0.0	0.0	0
std	0.0	6258912	2890938	4088051	3198311	2693970	1919723	5239449	2366212	3.072	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	302090580	105016700	201683680	108611800	102062980	71019800	235766400	91601200	32.0	0.0	0.0	0.0	4

Descrição do Cluster 3 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	193.0	193	193	193	193	193	193	193	193	193.0	193.0	193.0	193.0	193
mean	1.0	12665322	3891368	8881692	4079739	3866119	1709752	4912764	2823751	23.192	1.0	0.0	0.0	0
std	0.0	28545394	12604850	22880969	12703941	12638100	5234784	26692135	11441467	9.663	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	15.0	1.0	0.0	0.0	0
50%	1.0	29549	0	96680	0	10200	0	0	0	21.0	1.0	0.0	0.0	0
75%	1.0	8236709	1317680	4072980	786980	1843260	161080	191460	242980	32.0	1.0	0.0	0.0	0
max	1.0	139267780	101493980	145461280	101493980	101331480	29470800	266863500	102517860	32.0	1.0	0.0	0.0	2

Descrição do Cluster 4 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	370.0	370	370	370	370	370	370	370	370	370.0	370.0	370.0	370.0	370
mean	0.0	3562360	1323210	1782430	1384671	1571036	899883	3038180	1234736	18.438	1.0	0.0	0.0	0
std	0.0	13801229	6499327	7813902	7665708	6372205	3830660	20650535	7988205	10.305	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	12.0	1.0	0.0	0.0	0
50%	0.0	0	0	1	0	0	0	0	0	21.0	1.0	0.0	0.0	0
75%	0.0	326510	42330	229825	60880	130580	14699	302	1	32.0	1.0	0.0	0.0	0
max	0.0	136914960	74952000	101493980	110937399	58941600	29470800	237525200	88412400	32.0	1.0	0.0	0.0	2

Descrição do Cluster 5 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	61.0	61	61	61	61	61	61	61	61	61.0	61.0	61.0	61.0	61
mean	1.0	28640500	6621020	19944942	8796351	9115507	7542087	11773593	3916652	11.148	0.0	0.0	1.0	0
std	0.0	37730112	12282514	48000654	19902451	21919416	13602956	22774664	7285406	8.442	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	1.0	908360	1	130580	0	0	0	48980	900	6.0	0.0	0.0	1.0	0
50%	1.0	10162160	1319680	1318580	333680	282680	16199	1218060	524240	9.0	0.0	0.0	1.0	0
75%	1.0	33940800	7734000	10923600	7740000	7734000	7734000	11063440	3794440	15.0	0.0	0.0	1.0	0
max	1.0	119885501	58941600	238841400	107747800	101493980	56302799	104969000	33361680	32.0	0.0	0.0	1.0	1

Descrição do Cluster 6 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	48.0	48	48	48	48	48	48	48	48	48.0	48.0	48.0	48.0	48
mean	0.0	19068267	7027398	11668137	3186666	7447942	3829709	5716883	7134228	6.167	0.0	0.0	1.0	0
std	0.0	36042621	11538436	26181255	7296947	17633923	9619066	15690235	23528490	7.194	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	0.0	1575	0	0	0	0	0	0	0	1.0	0.0	0.0	1.0	0
50%	0.0	6478770	724130	384370	9099	130580	1000	34415	4850	3.0	0.0	0.0	1.0	0
75%	0.0	14690269	7950000	6325870	3088650	9051680	513580	585412	239675	7.5	0.0	0.0	1.0	0
max	0.0	134289180	38813800	103030901	38813800	101669879	44129800	77740800	144325000	21.0	0.0	0.0	1.0	1

Descrição do Cluster 7 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	50.0	50	50	50	50	50	50	50	50	50.0	50.0	50.0	50.0	50
mean	0.0	23951096	7380342	12678685	10668182	5184934	4493119	20175894	11719858	7.04	0.0	1.0	0.0	0
std	0.0	40844866	16235966	22560492	23458216	11540355	9175180	41738737	21931328	5.631	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	0.0	66605	130580	664910	225	500	0	14474	0	3.0	0.0	1.0	0.0	0
50%	0.0	7749226	1383970	4710380	565130	221880	3700	1258450	291670	6.0	0.0	1.0	0.0	0
75%	0.0	31650560	7734000	12419590	7734000	5327750	4578000	24353185	11636380	12.0	0.0	1.0	0.0	0
max	0.0	171271480	100013800	128488300	107747800	58941600	29470800	228451200	104286200	21.0	0.0	1.0	0.0	2

Descrição do Cluster 8 criado pela Clusterização DBSCAN do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	31.0	31	31	31	31	31	31	31	31	31.0	31.0	31.0	31.0	31
mean	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
std	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
50%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
75%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
max	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0

Figura I.17: Descrição dos Cluster Resultantes do DBSCAN no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	14089.0	14089	14089	14089	14089	14089	14089	14089	14089	14089.0	14089.0	14089.0	14089.0	14089
mean	0.0	586905	269400	473376	357148	304603	206062	547200	252190	0.831	0.0	0.0	0.0	0
std	0.0	4775176	2701285	3771945	3286824	3102784	1997613	5919004	2819363	3.018	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	149962680	107747800	105621380	133954599	138827600	100013800	250992200	166520999	32.0	0.0	0.0	0.0	4

Descrição do Cluster 2 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	58.0	58	58	58	58	58	58	58	58	58.0	58.0	58.0	58.0	58
mean	0.0	11199404	7048822	7747614	5739184	9742935	2621893	12254286	4728183	6.172	0.0	0.0	1.0	0
std	0.0	19495337	18082817	16533305	12964639	18677411	6449665	37902524	13716492	5.478	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	0.0	3	0	0	0	75	0	1	0	2.0	0.0	0.0	1.0	0
50%	0.0	1260960	178780	1177730	368430	1480180	130580	29778	5700	6.0	0.0	0.0	1.0	0
75%	0.0	10430185	2716290	7984261	7734000	10794080	1317680	3142385	309317	9.0	0.0	0.0	1.0	0
max	0.0	102515639	101331480	101493981	66285400	104642999	29470800	235766400	88412400	21.0	0.0	0.0	1.0	2

Descrição do Cluster 3 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	1272.0	1272	1272	1272	1272	1272	1272	1272	1272	1272.0	1272.0	1272.0	1272.0	1272
mean	1.0	7994889	2471634	5278727	4501101	3572040	1332664	6286712	2887177	4.519	0.0	0.0	0.0	0
std	0.0	24671751	9486343	18068515	16136678	14544773	4628404	27404825	11921312	6.899	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	2000	0	2900	0	0	2	0	0	1.0	0.0	0.0	0.0	0
75%	1.0	3363220	333680	1317680	927980	454180	130580	278691	130655	6.0	0.0	0.0	0.0	0
max	1.0	238096400	101331480	200043800	129484600	171033600	49938800	342484400	178375600	32.0	0.0	0.0	0.0	4

Descrição do Cluster 4 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	343.0	343	343	343	343	343	343	343	343.0	343.0	343.0	343.0	343.0	343
mean	0.0	1999892	530442	2895741	1188406	1688071	675811	970545	581464	19.706	1.0	0.0	0.0	0
std	0.0	8615544	2598035	15467393	17169222	7848021	3768266	5317670	3040038	10.678	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	12.0	1.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	21.0	1.0	0.0	0.0	0
75%	0.0	33480	2000	217730	2000	45280	2000	4	1	32.0	1.0	0.0	0.0	0
max	0.0	107878380	29470800	229498400	101331480	100013800	49938800	61872000	29470800	32.0	1.0	0.0	0.0	1

Descrição do Cluster 5 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	42.0	42	42	42	42	42	42	42	42	42.0	42.0	42.0	42.0	42
mean	1.0	24187269	11773465	21455694	9028846	17470144	4710395	29015714	9605900	7.381	0.0	1.0	0.0	0
std	0.0	36950985	21206920	34656493	14024657	26278982	8708842	44760476	16391115	5.872	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	1.0	381605	1317680	814430	786980	328904	43707	108000	32272	3.0	0.0	1.0	0.0	0
50%	1.0	9685180	1873630	4001940	7734000	7027290	361980	2529672	637489	6.0	0.0	1.0	0.0	0
75%	1.0	28165260	8938260	30766560	7760035	28828220	7734000	53746170	15542530	12.0	0.0	1.0	0.0	0
max	1.0	140504260	100013800	134115680	64036800	101331480	29470800	160800000	89323000	21.0	0.0	1.0	0.0	3

Descrição do Cluster 6 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	206.0	206	206	206	206	206	206	206	206	206.0	206.0	206.0	206.0	206
mean	1.0	13814249	2802597	10564182	4848876	5591115	1828600	4995869	4023587	23.636	1.0	0.0	0.0	0
std	0.0	35586012	8043251	28539442	14230799	15080703	5580293	15191848	17627202	9.372	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	21.0	1.0	0.0	0.0	0
50%	1.0	117380	0	130580	0	33480	0	0	0	21.0	1.0	0.0	0.0	0
75%	1.0	7847155	1317680	3246539	962930	2267080	282680	717415	393792	32.0	1.0	0.0	0.0	0
max	1.0	237232400	56302799	207761600	100019800	102763100	33940800	101395261	200027600	32.0	1.0	0.0	0.0	2

Descrição do Cluster 7 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	69.0	69	69	69	69	69	69	69	69	69.0	69.0	69.0	69.0	69
mean	1.0	25952377	8736848	15894356	11599367	15482011	5722287	13714770	9529470	10.29	0.0	0.0	1.0	0
std	0.0	44247278	17388629	23784118	21149412	27670642	9560746	26298030	18645762	8.655	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	1.0	1242260	900	691780	45280	130580	0	10200	1	3.0	0.0	0.0	1.0	0
50%	1.0	8839760	1317681	5413560	2499400	1771860	525580	836660	381110	9.0	0.0	0.0	1.0	0
75%	1.0	29611580	7734000	18177380	9051680	20478080	7734000	14939600	10952860	15.0	0.0	0.0	1.0	0
max	1.0	232247699	101331480	102326160	107747800	108611800	33940800	154499800	88412400	32.0	0.0	0.0	1.0	1

Descrição do Cluster 8 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	49.0	49	49	49	49	49	49	49	49	49.0	49.0	49.0	49.0	49
mean	0.0	20306986	5608062	10670570	14647138	8120333	5247744	15001134	9211314	9.061	0.0	1.0	0.0	0
std	0.0	33743978	9252136	13481879	26366381	11671825	8694605	25945364	15225479	7.644	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	0.0	341160	130580	1317680	130580	1317680	1	235940	84180	3.0	0.0	1.0	0.0	0
50%	0.0	8941060	1448260	3364180	7734000	2267080	454180	1997080	1110340	9.0	0.0	1.0	0.0	0
75%	0.0	29531681	7734000	15468000	15468000	9051680	7734000	11742360	11104660	12.0	0.0	1.0	0.0	0
max	0.0	200158179	33403000	61208680	129484600	51256479	38813800	107348300	62089200	32.0	0.0	1.0	0.0	4

Descrição do Cluster 9 criado pela Clusterização DBSCAN do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	29.0	29	29	29	29	29	29	29	29	29.0	29.0	29.0	29.0	29
mean	1.0	0	0	0	0	1	1	1	1	0.931	1.0	1.0	1.0	0
std	0.0	0	0	0	0	0	0	0	0	0.258	0.0	0.0	0.0	0
min	1.0	0	0	0	0	1	1	1	1	0.0	1.0	1.0	1.0	0

Descrição do Cluster 1 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	14100.0	14100	14100	14100	14100	14100	14100	14100	14100	14100.0	14100.0	14100.0	14100.0	14100
mean	0.0	662788	253562	446351	360535	298936	234485	451306	237717	0.0	0.0	0.0	0.0	0
std	0.0	5906192	2485629	3783644	3207402	2950843	2120279	5040802	2342141	3.032	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	300102280	103030900	104286200	107747800	110937399	79594800	235766400	88412400	32.0	0.0	0.0	0.0	5

Descrição do Cluster 2 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	1229.0	1229	1229	1229	1229	1229	1229	1229	1229	1229.0	1229.0	1229.0	1229.0	1229
mean	1.0	7665106	2811449	5362756	4582753	3660801	2703850	6217615	3045392	5.077	0.0	0.0	0.0	0
std	0.0	22976509	10397426	17257033	14780026	13642927	10385133	21414280	11864633	7.478	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	10200	1	900	1	1	1	300	0	2.0	0.0	0.0	0.0	0
75%	1.0	3932200	1161780	14605340	1656680	1021580	161080	695480	333680	6.0	0.0	0.0	0.0	0
max	1.0	266534800	129484600	233968400	156316600	179608800	100013800	235766400	229498400	32.0	0.0	0.0	0.0	5

Descrição do Cluster 3 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	203.0	203	203	203	203	203	203	203	203	203.0	203.0	203.0	203.0	203
mean	1.0	12399727	5933755	9655952	5083391	6138848	4627098	14117802	6539183	23.123	1.0	0.0	0.0	0
std	0.0	28882279	17280443	25115708	12336251	17138716	11126018	52253118	19996470	9.533	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	15.0	1.0	0.0	0.0	0
50%	1.0	196480	1	130580	2000	23550	1	8	0	21.0	1.0	0.0	0.0	0
75%	1.0	9551570	2223200	3060030	2506190	1577510	1945600	1489840	467310	32.0	1.0	0.0	0.0	0
max	1.0	194587000	107747800	157806600	100467980	129484600	71018800	508563399	149071400	32.0	1.0	0.0	0.0	4

Descrição do Cluster 4 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	353.0	353	353	353	353	353	353	353	353	353.0	353.0	353.0	353.0	353
mean	0.0	3934869	1461775	2889804	1066675	1494116	1008127	2602020	2050599	17.626	1.0	0.0	0.0	0
std	0.0	14769558	7109504	11549076	6639067	6624295	5018864	17114190	9696092	10.503	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	1.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	9.0	1.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	21.0	1.0	0.0	0.0	0
75%	0.0	196480	2000	282680	3700	130580	6000	6000	2	21.0	1.0	0.0	0.0	0
max	0.0	110937399	101331480	100154580	102763100	64615480	49938800	235766400	88412400	32.0	1.0	0.0	0.0	1

Descrição do Cluster 5 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	40.0	40	40	40	40	40	40	40	40	40.0	40.0	40.0	40.0	40
mean	0.0	13921141	10279510	16335854	6759887	12806168	2679902	20383222	10636056	8.5	0.0	1.0	0.0	0
std	0.0	23048309	14351782	26836506	17632553	19762163	5299639	24698118	15645208	6.429	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	0.0	227779	8575	130805	113155	798830	0	75580	14699	3.0	0.0	1.0	0.0	0
50%	0.0	7384030	2883200	2424070	1029280	7734000	14830	7734000	1849130	6.0	0.0	1.0	0.0	0
75%	0.0	18070360	17519625	16832170	7734000	16391100	2768410	34504175	19079250	15.0	0.0	1.0	0.0	0
max	0.0	110222560	58941600	100210280	108618000	104642999	29470800	68518000	59102680	21.0	0.0	1.0	0.0	3

Descrição do Cluster 6 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	63.0	63	63	63	63	63	63	63	63	63.0	63.0	63.0	63.0	63
mean	1.0	28801883	6601853	15699935	6615041	11503313	3850763	14716419	6164255	12.825	0.0	0.0	1.0	0
std	0.0	52318103	18233549	29752624	9984339	22638378	8398820	43519918	15430072	8.794	0.0	0.0	0.0	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	1.0	180270	6100	24840	0	24415	0	1651	0	3.0	0.0	0.0	1.0	0
50%	1.0	8728580	1317681	14495160	1656080	1845680	23550	464380	276960	12.0	0.0	0.0	1.0	0
75%	1.0	29664770	4450800	16235180	7757640	12441290	2749300	3812934	3936480	21.0	0.0	0.0	1.0	0
max	1.0	268312199	100013800	136003780	44290880	103946000	38813800	235766400	88412400	32.0	0.0	0.0	1.0	2

Descrição do Cluster 7 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	71.0	71	71	71	71	71	71	71	71	71.0	71.0	71.0	71.0	71
mean	0.0	8585281	4532882	6804640	3771557	9191982	3199897	7840713	4515674	7.549	0.0	0.0	1.0	0
std	0.0	16760620	12791389	18239502	8981554	22881913	7929291	27945382	13830154	8.427	0.0	0.0	0.0	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	1.0	0
25%	0.0	2	0	0	0	1	0	0	0	1.0	0.0	0.0	1.0	0
50%	0.0	1382260	130580	786980	3700	130580	45280	15600	10199	6.0	0.0	0.0	1.0	0
75%	0.0	8612060	1568130	4477970	2382240	2385570	1398930	1358675	697795	9.0	0.0	0.0	1.0	0
max	0.0	100074680	72337480	101462060	49737400	109870200	44129800	184838100	78503900	32.0	0.0	0.0	1.0	2

Descrição do Cluster 8 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	40.0	40	40	40	40	40	40	40	40	40.0	40.0	40.0	40.0	40
mean	1.0	25007939	5316208	13501534	8733749	9661497	9514092	19418246	6514347	9.1	0.0	1.0	0.0	0
std	0.0	31112585	8119984	17947601	17459645	17367200	16342218	42135374	15316842	7.027	0.0	0.0	0.0	1
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	1.0	0.0	0
25%	1.0	1945385	1037380	1317680	80480	512754	106305	186104	32849	3.0	0.0	1.0	0.0	0
50%	1.0	12264425	2047770	7148880	841880	2191980	2248640	1177090	239449	7.5	0.0	1.0	0.0	0
75%	1.0	31122990	7734000	15797420	7951500	10685880	7950000	2397985	5022675	15.0	0.0	1.0	0.0	1
max	1.0	109534460	32220100	69856200	79594800	101331480	74019800	235766400	88412400	21.0	0.0	1.0	0.0	4

Descrição do Cluster 9 criado pela Clusterização DBSCAN do mês 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GAExp	mmCPTime	PreHw	Hidn	Mrrg	BTC
count	32.0	32	32	32	32	32	32	32	32	32.0	32.0	32.0	32.0	32
mean	1.0	31141097	12225977	21097660	15223912	20117656	7753843	41630680	14615974	22.719	1.0	1.0	0.0	0
std	0.0	33831125	18577168	23526385	18201696	21420786	10305320	59015990	18336932	7.722	0.0	0.0	0.0	0
min	1.0	3	0	2	0	0	0	0	0	9.0	1.0	1.0	0.0	0
25%	1.0	6564186	2929785	3704020	786980	7580394								

Descrição do Cluster 1 criado pela Clusterização BANG do mes 6/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1387	2441927	1480559	2543065	1048516	1654331	1143634	5576717	2244247	2.679	0.0367	0.0609	0.0351	0
std	0.3456	10840264	7161737	11508801	4647759	7664238	5022448	30616030	11678457	6.707	0.188	0.2392	0.1841	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6000	1	6000	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1297	7256513	3550226	6139657	4192009	4203903	2592833	12720647	5483790	2.59	0.0607	0.0379	0.0375	0
std	0.336	38169851	20136637	32560825	22397964	23194900	13795610	83885318	33498861	7.442	0.2388	0.1909	0.1901	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10199	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1306	7629970	3614825	6348837	4294207	4516333	2761066	14271028	5981822	3.034	0.0613	0.0418	0.0428	0
std	0.337	40452326	20558580	33555898	22896913	24148377	14163579	90680055	35380006	7.565	0.2399	0.2002	0.2025	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10199	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1351	7935567	3651839	6534988	4669566	4738572	2878978	14181353	6039410	3.035	0.0606	0.0403	0.04	0
std	0.3418	41319019	21110792	34036050	24210107	24779474	14885718	90256193	35268049	7.579	0.2386	0.1967	0.196	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	16200	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 10/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1304	7557094	3522055	6226623	4341056	4374566	2660517	12656380	5715429	2.975	0.0589	0.0408	0.0413	0
std	0.3367	38768534	20344942	32499076	22784600	23386667	14068433	83163366	34153787	7.413	0.2354	0.1979	0.199	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10199	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 2/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1279	7937447	3559050	6566242	4437896	4578934	2619223	13518735	5586276	2.963	0.0603	0.0388	0.0386	0
std	0.334	41057228	20075740	33708282	23286584	24624104	13972548	87562401	33789600	7.407	0.238	0.193	0.1928	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6001	1	10199	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 1 criado pela Clusterização BANG do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	17000.0	17000	17000	17000	17000	17000	17000	17000	17000	17000.0	17000.0	17000.0	17000.0	17000
mean	0.1317	8068853	3610215	6645114	4584555	4633376	2826276	14193091	6130719	3.081	0.0626	0.0416	0.0426	0
std	0.3382	40709801	20484119	34096791	23708843	24246243	14416911	88823357	35248127	7.58	0.2422	0.1997	0.2021	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	10200	1	10199	1	1	1	0	0	1.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.23: Resultados da clusterização do algoritmo BANG

Descrição do Cluster 1 criado pela Clusterização SOM

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	14552.0	14552	14552	14552	14552	14552	14552	14552	14552	14552.0	14552.0	14552.0	14552.0	14552
mean	0.0	487259	273645	491921	269569	313341	233716	678409	310573	1.268	0.0047	0.0238	0.0041	0
std	0.0	3099075	2038023	3231160	1750176	2247910	1709911	6575208	2835840	4.216	0.0687	0.1524	0.0635	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	68598000	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4

Descrição do Cluster 2 criado pela Clusterização SOM

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	1904.0	1904	1904	1904	1904	1904	1904	1904	1904	1904.0	1904.0	1904.0	1904.0	1904
mean	0.9926	5045045	2949030	4703679	2228319	3242701	2402443	7409782	3412744	7.603	0.0588	0.1607	0.0762	0
std	0.0855	11126776	7850905	11570260	6105469	8250079	6398949	27071922	10874072	9.353	0.2354	0.3674	0.2653	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	60880	2850	45280	1	6000	1	3700	2	3.0	0.0	0.0	0.0	0
75%	1.0	2943965	1480180	2267380	786980	1514160	525580	786980	515560	12.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 3 criado pela Clusterização SOM

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	544.0	544	544	544	544	544	544	544	544	544.0	544.0	544.0	544.0	544
mean	0.8603	45618399	28625869	49849019	17756035	31966537	21078109	130190733	49880297	23.169	0.8143	0.7059	0.7224	2
std	0.347	31523390	22215886	32038829	12756762	22077306	11915304	96979725	35516182	8.156	0.3892	0.4561	0.4482	2
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	21240885	7734000	21717285	3231750	9051680	7734000	32426455	13736570	15.0	1.0	0.0	0.0	0
50%	1.0	39429180	30788480	50644800	29470800	30788480	29470800	122077050	54100800	21.0	1.0	1.0	1.0	1
75%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Figura I.24: Descrição dos Cluster Resultantes do SOM no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização SOM do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14702.0	14702	14702	14702	14702	14702	14702	14702	14702	14702.0	14702.0	14702.0	14702.0	14702
mean	0.0	1134792	488845	792680	640867	524511	393650	964916	511433	1.44	0.0283	0.0067	0.0059	0
std	0.0	9155314	4392425	6411753	5614284	5058435	3668496	11610062	6109438	4.652	0.1658	0.0814	0.0767	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	2	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	302090580	171033600	233968400	200027600	200027600	100013800	800110400	256330400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização SOM do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1705.0	1705	1705	1705	1705	1705	1705	1705	1705	1705.0	1705.0	1705.0	1705.0	1705
mean	1.0	13351826	5714252	10670197	7807965	7484931	4438350	10693735	5757812	8.045	0.1396	0.0463	0.0686	0
std	0.0	37409557	20078745	30868559	24013867	24939582	15657928	48173134	24805835	9.985	0.3467	0.2103	0.2529	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	2000	45280	300	6000	1	2000	2	3.0	0.0	0.0	0.0	0
75%	1.0	8067681	1514160	4352880	7734000	1845680	390280	1205480	632040	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	768865400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 3 criado pela Clusterização SOM do mes 10/17

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	593.0	593	593	593	593	593	593	593	593	593.0	593.0	593.0	593.0	593
mean	0.8432	141504475	73227706	125678745	81837334	85991883	51809982	310003333	127973495	26.89	0.6374	0.7875	0.7319	1
std	0.3639	128976483	69959654	107688812	74650998	77168720	43228320	316322113	117819114	7.78	0.4811	0.4094	0.4434	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	30043729	7734000	23202000	9384980	9051680	7734000	29476600	15548480	21.0	0.0	1.0	0.0	0
50%	1.0	107878380	51784480	102517860	54211200	59319899	29470800	181551740	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	216427580	119174200	207441400	133954599	156316600	100013800	685856400	249966400	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.25: Descrição dos Cluster Resultantes do SOM no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização SOM do mes 2/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	14703.0	14703	14703	14703	14703	14703	14703	14703	14703	14703.0	14703.0	14703.0	14703.0	14703
mean	0.0	958872	419259	841289	563360	491354	388143	886261	429525	1.396	0.0264	0.0055	0.0065	0
std	0.0	7630874	4205212	6996804	5003221	4538776	3741234	14250233	5406286	4.602	0.1603	0.074	0.0801	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	316609800	129484600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	5

Descrição do Cluster 2 criado pela Clusterização SOM do mes 2/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	1647.0	1647	1647	1647	1647	1647	1647	1647	1647	1647.0	1647.0	1647.0	1647.0	1647
mean	1.0	12857453	5636919	11166084	8021205	8201174	5192416	13263361	6917859	8.37	0.1324	0.0528	0.0747	0
std	0.0	35131977	19715285	34514152	24173263	25788839	16827728	52285833	26290286	10.313	0.339	0.2237	0.263	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	60880	900	45280	2000	6000	1	4005	2	3.0	0.0	0.0	0.0	0
75%	1.0	7794880	1448260	4010010	3455750	2009460	786980	1608190	837940	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização SOM do mes 2/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	650.0	650	650	650	650	650	650	650	650	650.0	650.0	650.0	650.0	650
mean	0.8831	145284573	70774850	118723400	79242318	86224550	50275909	319587726	129203017	26.563	0.6708	0.8354	0.7846	2
std	0.3216	135485568	70177631	109403131	75540779	77670174	42608156	326223213	120265903	7.876	0.4703	0.3711	0.4114	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	28360445	7734000	17037450	8598000	9051680	7734000	23271955	10406485	21.0	0.0	1.0	1.0	0
50%	1.0	107747801	40628690	88800100	41034600	63411600	29470800	193486100	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	224990899	110620599	207504850	129484600	171033600	100013800	786886150	277478650	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.26: Descrição dos Cluster Resultantes do SOM no mês 02/2018

Descrição do Cluster 1 criado pela Clusterização SOM do mes 6/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	14622.0	14622	14622	14622	14622	14622	14622	14622	14622	14622.0	14622.0	14622.0	14622.0	14622
mean	0.0	994680	411219	841395	592580	567089	374711	991861	473411	1.398	0.026	0.0059	0.0068	0
std	0.0	8060153	3748997	7582231	5427551	5268628	3795547	13393646	5986254	4.641	0.1591	0.0769	0.0824	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	297266200	107747800	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização SOM do mes 6/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	1791.0	1791	1791	1791	1791	1791	1791	1791	1791	1791.0	1791.0	1791.0	1791.0	1791
mean	0.9983	15438478	5852911	11540326	8663969	8293361	4907408	16531057	8664552	8.532	0.1575	0.0609	0.0782	0
std	0.0409	39483271	19799703	32201615	25031720	24938315	17086765	74382245	32478097	10.453	0.3643	0.2391	0.2685	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	900	130580	2000	10200	1	3700	2	3.0	0.0	0.0	0.0	0
75%	1.0	8862960	1463510	4272400	7734000	2049180	525580	1188510	848160	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 3 criado pela Clusterização SOM do mes 6/18

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTTime	PreHw	Hldn	Mrrg	BTC
count	587.0	587	587	587	587	587	587	587	587	587.0	587.0	587.0	587.0	587
mean	0.8654	157938860	77659036	133089917	94038768	97802958	59070579	335558603	136677243	27.039	0.6269	0.833	0.7496	2
std	0.3416	138667808	74944477	109883891	76744863	77916610	42949311	327284846	118643119	7.66	0.484	0.3732	0.4336	2
min	0.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
25%	1.0	31511259	7734000	31436070	15468000	16620490	9656400	40199110	17394440	21.0	0.0	1.0	0.5	0
50%	1.0	114664400	55053600	103380660	100013800	100013800	89187800	200814881	100335960	32.0	1.0	1.0	1.0	1
75%	1.0	258969200	129484600	229498400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.27: Descrição dos Cluster Resultantes do SOM no mês 06/2018

Descrição do Cluster 1 criado pela Clusterização SOM do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14690.0	14690	14690	14690	14690	14690	14690	14690	14690	14690.0	14690.0	14690.0	14690.0	14690
mean	0.0	1143399	450937	797621	546332	502773	360069	816792	437840	1.368	0.0251	0.0061	0.0078	0
std	0.0	9814454	4034083	7308843	4512988	4664435	3328199	10375482	5881411	4.434	0.1563	0.0776	0.0878	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	109670200	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização SOM do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1696.0	1696	1696	1696	1696	1696	1696	1696	1696	1696.0	1696.0	1696.0	1696.0	1696
mean	1.0	15285329	5913409	11873037	7895851	6824617	4497812	14915043	8431058	8.489	0.1462	0.0525	0.0796	0
std	0.0	39073045	20306270	32989377	22323210	21383322	15546669	67814675	33240978	10.418	0.3534	0.2231	0.2708	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	3700	104180	1	16199	1	3700	2	3.0	0.0	0.0	0.0	0
75%	1.0	8128064	1480180	4617575	7734000	2049180	525580	1484305	659809	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização SOM do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	614.0	614	614	614	614	614	614	614	614	614.0	614.0	614.0	614.0	614
mean	0.8469	139658212	70393366	120519331	85311029	90239967	52623912	289680559	124481018	26.205	0.627	0.8404	0.7378	2
std	0.3604	128528619	72299444	105456460	74721437	74970495	43797111	310248001	116820964	8.142	0.484	0.3665	0.4402	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	27240870	7734000	24244975	9535185	15468000	7734000	21220175	17179520	21.0	0.0	1.0	0.0	0
50%	1.0	107747801	37204800	101510260	90477100	88259600	29470800	153682700	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	216320900	114364599	201345279	129484600	156316600	100013800	590215150	249966400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.28: Descrição dos Cluster Resultantes do SOM no mês 10/2018

Descrição do Cluster 1 criado pela Clusterização SOM do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14706.0	14706	14706	14706	14706	14706	14706	14706	14706	14706.0	14706.0	14706.0	14706.0	14706
mean	0.0	1108812	494758	853690	590531	645775	373742	1177497	631012	1.335	0.0265	0.0061	0.0072	0
std	0.0	10150380	4869591	7594114	5042239	6296123	3599372	16732369	7669764	4.337	0.1607	0.078	0.0846	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	200027600	300041400	163311600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização SOM do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1745.0	1745	1745	1745	1745	1745	1745	1745	1745	1745.0	1745.0	1745.0	1745.0	1745
mean	0.9857	16900727	5896287	12956581	9558069	8287708	4587604	17398012	7575092	9.208	0.1708	0.0676	0.0779	0
std	0.1189	42800430	18854761	35781413	27200285	25869105	15357459	78783671	30102900	10.849	0.3764	0.2512	0.2681	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	2000	130580	900	10200	1	2001	1	3.0	0.0	0.0	0.0	0
75%	1.0	9579680	1481080	4535960	7734000	2049180	390280	1317680	676140	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização SOM do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	549.0	549	549	549	549	549	549	549	549	549.0	549.0	549.0	549.0	549
mean	0.8288	162365481	78212979	139275981	91222330	98147668	56512177	331771746	132000887	26.734	0.6138	0.8215	0.7559	2
std	0.377	135266656	69869432	104759852	75883474	79908742	43594444	323983529	119469594	7.826	0.4873	0.3833	0.4299	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	39840160	9783180	42735400	14350799	11615100	7734000	33069480	16727760	21.0	0.0	1.0	1.0	0
50%	1.0	117883200	58941600	115244400	100013800	100094280	49938800	210029700	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	264928400	116582200	229498400	178375600	200027600	100013800	800110400	279622400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.29: Descrição dos Cluster Resultantes do SOM no mês 02/2019

Descrição do Cluster 1 criado pela Clusterização SOM do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14642.0	14642	14642	14642	14642	14642	14642	14642	14642	14642.0	14642.0	14642.0	14642.0	14642
mean	0.0	1127237	424001	830045	566452	554018	445634	859852	483073	1.359	0.0251	0.0057	0.0075	0
std	0.0	9224897	4073165	7363631	4953344	5287278	4453872	10754752	5659660	4.453	0.1563	0.0751	0.0864	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	2	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	303058500	200027600	300041400	133954599	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização SOM do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1734.0	1734	1734	1734	1734	1734	1734	1734	1734	1734.0	1734.0	1734.0	1734.0	1734
mean	0.9913	14843969	5791969	12570039	8318274	8084320	5680359	18947347	8929121	9.084	0.1609	0.0634	0.0715	0
std	0.0926	37824144	18260429	35602673	22494503	23743260	17614297	80964597	34279405	10.54	0.3675	0.2438	0.2578	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	131031	16199	130580	16199	23550	1	10205	2	3.0	0.0	0.0	0.0	0
75%	1.0	9051680	2049180	5406075	7734000	2499400	1021580	2441380	1317905	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização SOM do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	624.0	624	624	624	624	624	624	624	624	624.0	624.0	624.0	624.0	624
mean	0.8333	152125086	72311156	126629775	88492881	90764812	51028829	313842763	130874953	26.812	0.6699	0.8237	0.7869	2
std	0.373	131129535	71780146	107749845	76702273	77206301	43115350	316456875	116282361	7.656	0.4706	0.3814	0.4099	2
min	0.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
25%	1.0	36041260	7734000	29568735	9755535	10542245	7734000	33192355	19414135	21.0	0.0	1.0	1.0	0
50%	1.0	117230139	46039200	102528470	80507100	77638100	29470800	201788090	88412400	32.0	1.0	1.0	1.0	1
75%	1.0	246193950	114364599	207761600	157289350	200027600	100013800	698119900	256330400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.30: Descrição dos Cluster Resultantes do SOM no mês 06/2019

Descrição do Cluster 1 criado pela Clusterização Spectral

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	16585.0	16585	16585	16585	16585	16585	16585	16585	16585	16585.0	16585.0	16585.0	16585.0	16585
mean	0.1193	1227853	718939	1205530	589196	787668	601490	1916296	868687	2.127	0.0159	0.0434	0.0157	0
std	0.3242	5944554	4028724	6166123	3113911	4219311	3277931	13932646	5737080	5.673	0.1249	0.2038	0.1245	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3700	1	3700	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	117.0	117	117	117	117	117	117	117	117	117.0	117.0	117.0	117.0	117
mean	0.8974	83395213	55571839	82140700	28720597	52022513	28474553	210350238	79291827	25.752	0.9231	0.6752	0.735	4
std	0.3047	13462779	8745745	16168808	4095880	14125821	4336233	57899314	21223546	8.066	0.2676	0.4703	0.4432	2
min	0.0	27946799	26894400	2558840	1317680	2797860	7734000	1842239	2	3.0	0.0	0.0	0.0	0
25%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	21.0	1.0	0.0	0.0	1
50%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	5
75%	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	298.0	298	298	298	298	298	298	298	298	298.0	298.0	298.0	298.0	298
mean	0.9195	38226761	22630868	45731328	15747117	30112495	20585699	128897459	48549984	24.292	0.849	0.7953	0.8389	1
std	0.2726	27074346	19121536	29808490	12577200	20883271	12104019	94734465	34593542	7.566	0.3587	0.4042	0.3682	1
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	16869380	4629200	20272050	2499400	9051680	7734000	32500264	15854479	21.0	1.0	1.0	1.0	0
50%	1.0	32995370	18435990	42608440	9656400	30788480	29470800	122823750	46332000	21.0	1.0	1.0	1.0	1
75%	1.0	59102680	33257350	66675599	29470800	50877600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	2
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Figura I.31: Descrição dos Cluster Resultantes do Spectral no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 10/17

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16632.0	16632	16632	16632	16632	16632	16632	16632	16632	16632.0	16632.0	16632.0	16632.0	16632
mean	0.112	3346295	1533341	2669154	1909862	1843984	1208331	3799801	1903069	2.409	0.0462	0.0189	0.0199	0
std	0.3154	19665870	10458255	16562477	12190920	12782901	8319967	35131666	15823824	6.35	0.21	0.1363	0.1397	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6000	1	3700	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 10/17

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	74.0	74	74	74	74	74	74	74	74	74.0	74.0	74.0	74.0	74
mean	0.8649	336206988	175222776	257346084	183835108	179830297	95128456	696266837	270945283	29.378	0.6216	0.9189	0.7162	4
std	0.3442	88780178	43571102	75934767	38483520	45459169	17005887	217140473	69616246	6.061	0.4883	0.2748	0.4539	2
min	0.0	89233200	22084800	1480180	9579680	2049180	9656400	9421160	2571940	6.0	0.0	0.0	0.0	0
25%	1.0	293834500	158065350	218312800	200027600	200027600	100013800	724490150	300041400	32.0	0.0	1.0	0.0	2
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	5
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 10/17

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	294.0	294	294	294	294	294	294	294	294	294.0	294.0	294.0	294.0	294
mean	0.9456	145666095	74438179	139242135	88080041	93502596	57624770	345336608	141233496	29.187	0.7381	0.8878	0.8639	2
std	0.2272	114855273	64243016	100523298	69233249	72910090	41596782	304818704	112555007	5.629	0.4404	0.3162	0.3434	2
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	44121970	13276984	48053970	24076560	29153100	19160400	63246595	30139125	32.0	0.0	1.0	1.0	0
50%	1.0	123649980	58941600	128636499	100013800	100013800	53120799	235766400	103428040	32.0	1.0	1.0	1.0	1
75%	1.0	215767835	112492600	225718400	133954599	149952600	100013800	690005525	249966400	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.32: Descrição dos Cluster Resultantes do Spectral no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 2/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16539.0	16539	16539	16539	16539	16539	16539	16539	16539	16539.0	16539.0	16539.0	16539.0	16539
mean	0.1074	2881716	1352694	2471067	1727542	1780005	1170874	3508587	1722130	2.322	0.0428	0.0183	0.0199	0
std	0.3097	17086742	9938067	15604511	11225132	12299434	8003951	31957390	14071970	6.25	0.2024	0.1339	0.1396	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3700	1	3700	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 2/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	82.0	82	82	82	82	82	82	82	82	82.0	82.0	82.0	82.0	82
mean	0.9268	351035855	175038061	269706456	187377170	184742116	94003110	751957462	283278826	29.988	0.6951	0.9512	0.8049	4
std	0.262	78560513	46588489	65434003	34427285	38355479	20731406	130206404	46876074	5.438	0.4632	0.2167	0.3987	2
min	0.0	88412400	9390080	667360	43821600	11615100	2267080	235766400	8674480	3.0	0.0	0.0	0.0	0
25%	1.0	300769045	163311600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	0.0	1.0	1.0	1
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 2/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	379.0	379	379	379	379	379	379	379	379	379.0	379.0	379.0	379.0	379
mean	0.971	140537833	65241947	118589247	76688057	84932214	52413639	324323062	131872834	28.259	0.7309	0.8734	0.8786	1
std	0.1681	123345320	61760830	103591834	70679948	73709558	41689442	319328038	117441724	6.358	0.4441	0.333	0.327	2
min	0.0	0	0	0	0	0	0	0	0	6.0	0.0	0.0	0.0	0
25%	1.0	33549430	9051680	17957080	9299580	9341890	11701200	33097330	15586230	21.0	0.0	1.0	1.0	0
50%	1.0	110249560	43821600	101800460	43821600	68284600	29470800	216297400	90549400	32.0	1.0	1.0	1.0	1
75%	1.0	208393190	102397040	201514479	129484600	129484600	100013800	775403900	253148400	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.33: Descrição dos Cluster Resultantes do Spectral no mês 02/2018

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16562.0	16562	16562	16562	16562	16562	16562	16562	16562	16562.0	16562.0	16562.0	16562.0	16562
mean	0.1141	3221915	1308421	2582249	1845253	1801541	1120749	3987668	2013866	2.361	0.0447	0.0178	0.0196	0
std	0.3179	18841368	9384357	15863781	11750327	11837771	8073330	36899949	16369357	6.345	0.2067	0.1321	0.1385	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3700	1	6000	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	364.0	364	364	364	364	364	364	364	364	364.0	364.0	364.0	364.0	364
mean	0.9231	150821561	73042180	130767125	96211005	102123842	63832696	327346986	133967048	28.107	0.6126	0.8791	0.7995	2
std	0.2668	125215895	68768301	104095552	73434482	74672827	41270011	317303735	114423697	6.728	0.4878	0.3264	0.401	2
min	0.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
25%	1.0	40473385	9051680	33897930	29470800	30788479	29470800	54083119	25225125	21.0	0.0	1.0	1.0	0
50%	1.0	114758899	57436339	106147440	100800780	101331480	100013800	193880400	101666290	32.0	1.0	1.0	1.0	1
75%	1.0	227999304	103393325	208091050	171033600	200027600	100013800	773260150	245609650	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 6/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	74.0	74	74	74	74	74	74	74	74	74.0	74.0	74.0	74.0	74
mean	0.9595	360057335	186808606	280112868	186496526	183047555	96563259	755201945	277733976	30.419	0.8919	0.9595	0.8784	4
std	0.1986	80779352	31988102	49136681	35233340	44592746	14785468	134786419	60331204	4.475	0.3126	0.1986	0.3291	2
min	0.0	15468000	40394400	1610760	37204800	15196000	29470800	240236400	1141940	12.0	0.0	0.0	0.0	0
25%	1.0	348963200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	2
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.34: Descrição dos Cluster Resultantes de Spectral no mês 06/2018

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16626.0	16626	16626	16626	16626	16626	16626	16626	16626	16626.0	16626.0	16626.0	16626.0	16626
mean	0.1122	3759757	1542515	2937813	1986621	1961795	1256099	3974654	2143938	2.385	0.044	0.0211	0.0233	0
std	0.3157	21623555	10870157	18143107	12285138	12975368	8572055	35814152	17082785	6.295	0.205	0.1438	0.151	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	6000	1	4000	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	95.0	95	95	95	95	95	95	95	95	95.0	95.0	95.0	95.0	95
mean	0.9053	333127932	182783399	276515427	185893363	181361410	97387296	615749773	247544178	28.326	0.5158	0.8737	0.8632	3
std	0.2944	96835943	39629082	55497652	32327440	42794365	15208103	267397590	86582501	6.565	0.5024	0.334	0.3455	2
min	0.0	16619460	10431100	103317101	68284600	4457780	130580	307	1705760	3.0	0.0	0.0	0.0	0
25%	1.0	265014300	200027600	300041400	200027600	200027600	100013800	432896700	203645350	21.0	0.0	1.0	1.0	1
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	4
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	279.0	279	279	279	279	279	279	279	279	279.0	279.0	279.0	279.0	279
mean	0.9462	122988233	60446819	110177601	82826255	87890628	54097075	324658148	136202439	29.559	0.7921	0.9319	0.8315	2
std	0.226	101564186	56964924	86023033	67732224	68583473	42312155	309073361	116213194	5.18	0.4065	0.2524	0.3749	2
min	0.0	0	0	0	0	0	0	0	0	9.0	0.0	0.0	0.0	0
25%	1.0	39989570	9051680	38491040	16081490	24931250	9127200	61872000	26891200	32.0	1.0	1.0	1.0	0
50%	1.0	102529401	38213200	101492560	100250780	100013800	33940800	215952000	97755400	32.0	1.0	1.0	1.0	2
75%	1.0	188857100	101331480	158955400	129484600	129484600	100013800	683962400	256330400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.35: Descrição dos Cluster Resultantes do Spectral no mês 10/2018

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16609.0	16609	16609	16609	16609	16609	16609	16609	16609	16609.0	16609.0	16609.0	16609.0	16609
mean	0.1091	3523744	1438655	2821551	1932037	1903143	1095282	4303804	1999389	2.366	0.0467	0.0193	0.0198	0
std	0.3118	20689840	9731583	17242008	12104156	12718580	7627665	39693962	16433647	6.291	0.2109	0.1375	0.1393	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	3700	1	3700	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	75.0	75	75	75	75	75	75	75	75	75.0	75.0	75.0	75.0	75
mean	0.9467	342424651	160469135	272291822	173478482	178619718	93621666	708274454	275633576	30.533	0.88	0.9333	0.7867	5
std	0.2262	84610322	52919811	54645559	46634023	46985685	20729654	188612241	53922408	3.764	0.3271	0.2511	0.4124	2
min	0.0	96146400	35728800	63234080	4272400	4629200	691780	202947299	69865200	21.0	0.0	0.0	0.0	0
25%	1.0	302643700	110937399	294628400	147048100	200027600	100013800	794697400	300041400	32.0	1.0	1.0	1.0	5
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	316.0	316	316	316	316	316	316	316	316	316.0	316.0	316.0	316.0	316
mean	0.9241	160534492	77765941	140319897	96025793	103911707	61119105	332962129	130020022	27.807	0.5823	0.8513	0.8513	2
std	0.2653	125267903	66899178	97180216	74015679	78653659	42894080	317111653	114531732	7.21	0.494	0.3564	0.3564	2
min	0.0	16199	0	0	900	0	0	0	0	1.0	0.0	0.0	0.0	0
25%	1.0	52996300	15383354	62287520	29470800	23124920	13882799	52048255	22341625	21.0	0.0	1.0	1.0	0
50%	1.0	123625600	58941600	119866940	100229440	102062980	100013800	213315260	89455890	32.0	1.0	1.0	1.0	1
75%	1.0	250123175	116582200	204482625	183788600	200027600	100013800	759515900	238916150	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.36: Descrição dos Cluster Resultantes do Spectral no mês 02/2019

Descrição do Cluster 1 criado pela Clusterização Spectral do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	16543.0	16543	16543	16543	16543	16543	16543	16543	16543	16543.0	16543.0	16543.0	16543.0	16543
mean	0.1094	3374317	1267802	2641671	1833497	1747338	1251349	3822106	1900441	2.393	0.0451	0.0178	0.0203	0
std	0.3122	19333297	8827695	16475752	11535878	11581334	8440468	35066736	15676118	6.334	0.2075	0.1321	0.1409	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	5700	1	3700	1	1	0	0	0	0.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Spectral do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	93.0	93	93	93	93	93	93	93	93	93.0	93.0	93.0	93.0	93
mean	0.9032	341771069	180158747	274534233	188731980	187026929	98184982	697851077	267708699	30.516	0.7849	0.9462	0.8172	4
std	0.2973	89034568	43993319	61013557	31723931	36317786	10674144	213742398	66788558	3.922	0.4131	0.2268	0.3886	2
min	0.0	67288800	1934359	28810800	38068800	161080	29470800	8141560	31737880	15.0	0.0	0.0	0.0	0
25%	1.0	301359080	200027600	300041400	200027600	200027600	100013800	800110400	271047400	32.0	1.0	1.0	1.0	3
50%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6
75%	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Spectral do mes 6/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	364.0	364	364	364	364	364	364	364	364	364.0	364.0	364.0	364.0	364
mean	0.9478	136166083	64960567	120148614	82565468	89196904	50506683	310860704	131555791	27.374	0.6731	0.8929	0.8626	2
std	0.2227	112157509	59982785	95366371	69808953	70782282	41659958	298241171	108318180	7.233	0.4697	0.3097	0.3447	2
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	36397465	8382000	34729580	12478800	20396735	7734000	53550350	29856599	21.0	0.0	1.0	1.0	0
50%	1.0	113241179	51441439	103566670	72754600	92227900	29470800	228805700	101329970	32.0	1.0	1.0	1.0	1
75%	1.0	203062810	102280880	200031650	129484600	135172849	100013800	590716400	229676150	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.37: Descrição dos Cluster Resultantes do Spectral no mês 06/2019

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	14529.0	14529	14529	14529	14529	14529	14529	14529	14529	14529.0	14529.0	14529.0	14529.0	14529
mean	0.0	449133	242864	445311	249863	280321	212818	598567	279585	1.252	0.0034	0.0231	0.0032	0
std	0.0	2892607	1802500	2893573	1626922	1999715	1577544	5866082	2641504	4.209	0.0586	0.1503	0.0562	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	1	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	68598000	37204800	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	4

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	658.0	658	658	658	658	658	658	658	658	658.0	658.0	658.0	658.0	658
mean	0.8298	41008665	25702106	44941074	15977931	28773613	19440286	114554499	44071585	22.233	0.7599	0.6748	0.6733	1
std	0.3761	31386975	21885252	32638443	13003333	22314664	12467142	98492214	36035302	8.409	0.4275	0.4688	0.4694	2
min	0.0	0	0	0	0	0	0	0	0	3.0	0.0	0.0	0.0	0
25%	1.0	13113684	4001430	14034490	1480180	7734000	7734000	13905870	7733999	15.0	1.0	0.0	0.0	0
50%	1.0	32837660	29470800	39123680	10923600	29478900	29470800	80127630	33713440	21.0	1.0	1.0	1.0	1
75%	1.0	67539600	41560800	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	3
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	Hldn	PreHw	Mrrg	BTC
count	1813.0	1813	1813	1813	1813	1813	1813	1813	1813	1813.0	1813.0	1813.0	1813.0	1813
mean	0.9994	4414565	2608361	3966330	2030354	2822840	1962513	5918778	2808060	7.01	0.0408	0.1412	0.0596	0
std	0.0235	10113069	7406721	9956204	5753780	7408426	5581341	22109896	9280830	8.971	0.1979	0.3483	0.2368	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	33480	2000	16200	1	2000	1	900	1	3.0	0.0	0.0	0.0	0
75%	1.0	2267080	1327880	1845680	786980	1319680	333680	580860	333980	12.0	0.0	0.0	0.0	0
max	1.0	88412400	58941600	88412400	29470800	58941600	29470800	235766400	88412400	32.0	1.0	1.0	1.0	6

Figura I.38: Descrição dos Cluster Resultantes do C-means no mês 06/2017

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 10/17

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14544.0	14544	14544	14544	14544	14544	14544	14544	14544	14544.0	14544.0	14544.0	14544.0	14544
mean	0.0	1025602	419480	682473	538629	439679	320538	813808	438543	1.158	0.0192	0.0046	0.0042	0
std	0.0	8644812	3777044	5545664	4695068	4478333	3023913	10434437	5486899	3.717	0.1372	0.0677	0.0646	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	302090580	105016700	201683680	200027600	200027600	100013800	800110400	256330400	32.0	1.0	1.0	1.0	5

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 10/17

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	834.0	834	834	834	834	834	834	834	834	834.0	834.0	834.0	834.0	834
mean	0.699	107815960	55295334	94942973	63193247	65503653	39445900	227870601	94518442	26.625	0.6823	0.6319	0.5875	1
std	0.459	123967389	66831627	104914191	72324649	74767816	42782047	298990734	113969777	8.018	0.4659	0.4826	0.4926	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	7895079	1478760	2767590	1317680	1620910	236979	1074374	373142	21.0	0.0	0.0	0.0	0
50%	1.0	47043180	23473280	46913580	29470800	29942990	29470800	61872000	30202770	32.0	1.0	1.0	1.0	0
75%	1.0	176895900	101331480	158955400	110620599	109670200	100013800	362704650	200288730	32.0	1.0	1.0	1.0	2
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 10/17

Estatística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1622.0	1622	1622	1622	1622	1622	1622	1622	1622	1622.0	1622.0	1622.0	1622.0	1622
mean	1.0	11421616	5016405	9411740	6613545	6437493	4018720	8859985	4943207	7.263	0.1134	0.0308	0.0536	0
std	0.0	33799093	19007856	29166248	21410423	22887418	15036794	42123392	22711425	9.381	0.3172	0.1729	0.2254	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	80480	900	20399	1	900	0	300	1	3.0	0.0	0.0	0.0	0
75%	1.0	7857979	1319680	3163360	2499400	1480180	236979	857830	392005	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	723574400	300041400	32.0	1.0	1.0	1.0	6

Figura I.39: Descrição dos Cluster Resultantes do C-means no mês 10/2017

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14673.0	14673	14673	14673	14673	14673	14673	14673	14673	14673.0	14673.0	14673.0	14673.0	14673
mean	0.0	905371	405526	773475	522179	451326	350812	830533	401721	1.358	0.0251	0.0044	0.0053	0
std	0.0	7278571	4071163	6300498	4660754	4222217	3423229	14019568	5288438	4.512	0.1564	0.0664	0.0727	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	316609800	129484600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	5

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	754.0	754	754	754	754	754	754	754	754	754.0	754.0	754.0	754.0	754
mean	0.8594	131576732	63016402	109555449	72113973	78395933	45785042	282337719	114883603	25.871	0.6538	0.7798	0.756	1
std	0.3478	132315798	68949601	107809664	73844436	76633805	42614404	318292028	118371742	8.239	0.4761	0.4146	0.4298	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	18131845	3543210	12788815	7783180	7734000	2499400	14579349	7734000	21.0	0.0	1.0	1.0	0
50%	1.0	97542640	31519980	73292400	37204800	57156780	29470800	126582340	67371440	32.0	1.0	1.0	1.0	1
75%	1.0	208894200	102901475	200361280	129484600	129484600	100013800	643793900	237713900	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 2/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1573.0	1573	1573	1573	1573	1573	1573	1573	1573	1573.0	1573.0	1573.0	1573.0	1573
mean	1.0	10944767	5077792	8885075	6971166	7021506	4620939	11149663	5832355	7.721	0.1151	0.0369	0.0509	0
std	0.0	31825724	18649059	28854574	22546776	23407588	15636681	48921917	24166513	9.896	0.3192	0.1885	0.2198	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	33480	1	19899	1	2000	0	2000	1	3.0	0.0	0.0	0.0	0
75%	1.0	7734001	1318580	2629980	1845680	1514160	525580	1025040	559140	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.40: Descrição dos Cluster Resultantes do C-means no mês 02/2018

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14583.0	14583	14583	14583	14583	14583	14583	14583	14583	14583.0	14583.0	14583.0	14583.0	14583
mean	0.0	842496	358505	720598	523008	478501	316541	741842	376304	1.35	0.0246	0.0045	0.0051	0
std	0.0	6391580	3363649	6450989	4758914	4635778	3275524	7724452	4371759	4.54	0.155	0.0671	0.0715	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	300	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	200158179	107747800	300041400	200027600	200027600	100013800	250992200	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	724.0	724	724	724	724	724	724	724	724	724.0	724.0	724.0	724.0	724
mean	0.8329	138963376	66971846	117323827	82181423	85320355	51745927	287324988	117745372	26.276	0.6215	0.761	0.7169	1
std	0.3733	135115542	72492837	107664254	76010396	76776494	43606402	319380316	117265863	8.027	0.4853	0.4267	0.4508	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	21265950	4033285	16785680	8748830	9186935	7734000	18627245	8350480	21.0	0.0	1.0	0.0	0
50%	1.0	101710220	33743200	94322590	51745099	58941600	33940800	122012500	77655599	32.0	1.0	1.0	1.0	1
75%	1.0	222560894	103102450	202776900	135172849	156316600	100013800	658072800	229498400	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 6/18

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1693.0	1693	1693	1693	1693	1693	1693	1693	1693	1693.0	1693.0	1693.0	1693.0	1693
mean	1.0	13000024	4941272	9240321	7239372	6973298	4053429	13137293	7049432	7.604	0.1305	0.0402	0.0508	0
std	0.0	34550728	17989403	27857772	22339571	22911922	15537699	65262210	28963942	9.772	0.337	0.1964	0.2196	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	60880	1	60880	1	900	0	906	1	3.0	0.0	0.0	0.0	0
75%	1.0	7895079	1317681	2640540	7734000	1600360	282680	725260	615260	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Figura I.41: Descrição dos Cluster Resultantes do C-means no mês 06/2018

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14641.0	14641	14641	14641	14641	14641	14641	14641	14641	14641.0	14641.0	14641.0	14641.0	14641
mean	0.0	993144	391417	661532	477408	445913	325642	659791	337532	1.308	0.0227	0.0044	0.0064	0
std	0.0	8167545	3506195	5528785	3913357	4282410	3125359	6938258	3908938	4.283	0.1489	0.0665	0.0794	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	2	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	329512200	109670200	300041400	200027600	200027600	100013800	414534400	200043800	32.0	1.0	1.0	1.0	7

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	745.0	745	745	745	745	745	745	745	745	745.0	745.0	745.0	745.0	745
mean	0.8081	124811525	61152437	107066820	75258804	78156607	45961469	249293707	110027602	25.274	0.6309	0.7557	0.698	1
std	0.3941	125135978	69964743	103869931	73217443	74390100	43660071	299915913	115059793	8.68	0.4829	0.43	0.4594	2
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	20273680	3147680	12867480	8520980	7734000	2749300	8823840	6589760	21.0	0.0	1.0	0.0	0
50%	1.0	98174080	30788479	88412400	41674800	58941600	29470800	104856980	61690900	32.0	1.0	1.0	1.0	1
75%	1.0	200158179	102062980	164914600	129484600	129484600	100013800	418734400	208625600	32.0	1.0	1.0	1.0	3
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 10/18

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1614.0	1614	1614	1614	1614	1614	1614	1614	1614	1614.0	1614.0	1614.0	1614.0	1614
mean	1.0	12977314	5319480	10162525	6654532	5955598	3853639	12251947	6350636	7.808	0.1233	0.0409	0.0551	0
std	0.0	35226854	19145034	29728299	19890422	19700830	14178705	61571809	27560693	9.983	0.3289	0.1981	0.2283	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	61881	900	45280	1	6000	0	905	1	3.0	0.0	0.0	0.0	0
75%	1.0	7828360	1337892	3303300	5296000	1505665	333680	1183357	521535	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Figura I.42: Descrição dos Cluster Resultantes do C-means no mês 10/2018

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14659.0	14659	14659	14659	14659	14659	14659	14659	14659	14659.0	14659.0	14659.0	14659.0	14659
mean	0.0	979284	417614	721328	526880	533040	326761	947596	512053	1.293	0.0246	0.0037	0.0052	0
std	0.0	9021121	4029752	6511717	4584038	5460258	3249568	13809060	6656492	4.272	0.1548	0.0606	0.0718	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	400055200	200027600	300041400	163311600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	4

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	733.0	733	733	733	733	733	733	733	733	733.0	733.0	733.0	733.0	733
mean	0.7735	135102445	63501340	116193282	76026224	80770772	45801327	267815557	106991014	25.312	0.618	0.7558	0.6821	1
std	0.4188	131060060	67835179	104649623	74530551	78685033	43839095	311320073	115590417	8.465	0.4862	0.4299	0.466	2
min	0.0	0	0	0	0	0	0	0	0	1.0	0.0	0.0	0.0	0
25%	1.0	23202000	3923159	16785680	7814479	7734000	2049180	12238660	4629960	21.0	0.0	1.0	0.0	0
50%	1.0	96146400	32487900	100669960	37204800	58351980	29470800	106473940	61440999	32.0	1.0	1.0	1.0	1
75%	1.0	216598000	103317100	200028500	129484600	159189600	100013800	512864980	219188000	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 2/19

Estatistica	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1608.0	1608	1608	1608	1608	1608	1608	1608	1608	1608.0	1608.0	1608.0	1608.0	1608
mean	1.0	13402479	4872868	9877163	7458636	6730767	3833602	12200793	5619455	8.003	0.1318	0.0317	0.0504	0
std	0.0	37649046	16985342	28976936	22976836	22802915	13909878	65399806	25728130	10.128	0.3384	0.1753	0.2188	0
min	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	80481	1	45280	1	2000	1	8	0	3.0	0.0	0.0	0.0	0
75%	1.0	7864579	1317681	2271130	4272400	1480180	236979	675255	403915	12.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	4

Figura I.43: Descrição dos Cluster Resultantes do C-means no mês 02/2019

Descrição do Cluster 1 criado pela Clusterização Fuzzy C-Means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	14617.0	14617	14617	14617	14617	14617	14617	14617	14617	14617.0	14617.0	14617.0	14617.0	14617
mean	0.0	1058957	409595	746082	538716	499469	401877	778800	447015	1.332	0.0243	0.0044	0.0062	0
std	0.0	8842141	4037660	6756755	4871783	4757760	4124343	10329869	5468994	4.399	0.1539	0.0665	0.0782	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
75%	0.0	1	0	2	0	0	0	0	0	0.0	0.0	0.0	0.0	0
max	0.0	303058500	200027600	300041400	133954599	189201600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Descrição do Cluster 2 criado pela Clusterização Fuzzy C-Means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	749.0	749	749	749	749	749	749	749	749	749.0	749.0	749.0	749.0	749
mean	0.8091	134060574	63050720	113149931	78147049	80783371	45313962	278688054	116894646	25.992	0.6689	0.7744	0.7276	2
std	0.3933	128590963	69517485	105747324	75434661	76065606	42768591	308860588	115093297	7.931	0.4709	0.4183	0.4455	2
min	0.0	0	0	0	0	0	0	0	0	2.0	0.0	0.0	0.0	0
25%	1.0	23202000	4155480	15821080	7814479	7734000	3303300	16468961	10341560	21.0	0.0	1.0	0.0	0
50%	1.0	100144380	31519980	88412400	43086200	58941600	29470800	133389599	84524400	32.0	1.0	1.0	1.0	1
75%	1.0	211112280	102280880	200043800	129484600	138827600	100013800	569365200	229498400	32.0	1.0	1.0	1.0	4
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	7

Descrição do Cluster 3 criado pela Clusterização Fuzzy C-Means do mes 6/19

Estadística	FCom	PMExp	PRExp	TExp	MRExp	HExp	ArcExp	CrExp	GaExp	mmCPTime	PreHw	Hldn	Mrrg	BTC
count	1634.0	1634	1634	1634	1634	1634	1634	1634	1634	1634.0	1634.0	1634.0	1634.0	1634
mean	0.9994	13023471	4994867	10594963	7056845	6707414	5142172	12950720	6202023	8.233	0.1273	0.0379	0.0551	0
std	0.0247	35109134	17152188	32273019	19944287	21207806	16957775	61463886	26108438	10.069	0.3334	0.1911	0.2282	0
min	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
25%	1.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0
50%	1.0	130580	2850	45280	2850	4850	1	6000	1	3.0	0.0	0.0	0.0	0
75%	1.0	7960209	1480180	3730400	7734000	1656080	454180	1461790	805175	15.0	0.0	0.0	0.0	0
max	1.0	400055200	200027600	300041400	200027600	200027600	100013800	800110400	300041400	32.0	1.0	1.0	1.0	6

Figura I.44: Descrição dos Cluster Resultantes do C-means no mês 06/2019