

UNIVERSIDADE FEDERAL DO PARANÁ

MATEUS RAMOS ASSOLARI

ANÁLISE COMPORTAMENTAL DE COMPRA DOS USUÁRIOS BRASILEIROS
DA PLATAFORMA STEAM

CURITIBA

2023

UNIVERSIDADE FEDERAL DO PARANÁ

MATEUS RAMOS ASSOLARI

Trabalho de Conclusão ao Curso de Graduação em Gestão da Informação, Departamento de Ciência e Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, apresentado como requisito parcial à obtenção do título de Bacharel em Gestão da Informação.

Orientador: Prof. Dr. José Marcelo Almeida Prado Cestari

CURITIBA

2023

AGRADECIMENTOS

Bellei, por ter me auxiliado no entendimento de como encontrar meu público alvo.

Isa, pelos comentários sobre a disposição do meu trabalho.

A minha mãe, que me ajudou, me motivou e sempre disse “entrega logo de uma vez assim mesmo”

A muita cafeína e estresse que me motivaram.

Ao meu chefe que me deixou faltar no trabalho e entendeu bem a importância desse processo para mim.

E a meus professores, que me auxiliaram e me guiaram sobre como aplicar aquilo que aprendi no meu curso e me fizeram entender o que eu realmente gostava de fazer.

RESUMO

A indústria de jogos digitais tem experimentado um crescimento significativo nos últimos anos, e o mercado brasileiro não é exceção. Com o avanço da tecnologia e a popularização das plataformas de distribuição digital, como a Steam, tornou-se fundamental compreender o comportamento dos usuários brasileiros nesse ambiente.

Neste contexto, este artigo descreve uma análise de dados dos usuários brasileiros da plataforma Steam, utilizando a API fornecida pela plataforma para coletar informações sobre seus hábitos de consumo de jogos. A base de dados resultante foi aplicada ao software Weka, ferramenta de mineração de dados, com o objetivo de identificar padrões no comportamento dos usuários brasileiros da Steam, resultando que os mesmos tendem a gastar com jogos de valores baixos e apresentam um padrão de tempo de jogo relativamente baixo.

A relevância deste estudo tanto para a plataforma Steam quanto para desenvolvedores de jogos, é fornecer insights sobre o comportamento do mercado brasileiro. Os desenvolvedores podem ajustar suas estratégias de precificação e conteúdo para atender melhor às preferências dos usuários brasileiros. Outro uso destes dados é que a Steam pode direcionar campanhas de marketing específicas para esse público, destacando jogos de valores mais baixos e oferecendo benefícios adicionais para incentivar maior engajamento.

Essa análise demonstra a utilidade da API da Steam para coletar dados valiosos sobre os usuários e a eficácia do software Weka na mineração desses dados. A combinação dessas ferramentas permite uma análise mais profunda e precisa do comportamento dos usuários, proporcionando informações valiosas para tomadas de decisão estratégicas e personalização de serviços na indústria de jogos.

Palavras-chave: Jogos Digitais ; Padrão de Consumo; Análise de Dados

ABSTRACT

The digital gaming industry has experienced significant growth in recent years, and the Brazilian market is no exception. With the advancement of technology and the popularization of digital distribution platforms, such as Steam, it has become essential to understand the behavior of Brazilian users in this environment.

In this context, this article describes an analysis of data from Brazilian users of the Steam platform, using the API provided by the platform to gather information about their gaming consumption habits. The resulting database was applied to the Weka software, which is a data mining tool, in order to identify patterns and interesting insights about the behavior of Brazilian users.

This finding may be relevant to both the Steam platform and game developers, providing insights into the behavior of the Brazilian market. Developers can adjust their pricing and content strategies to better cater to the preferences of Brazilian users. Additionally, Steam can target specific marketing campaigns to this audience, highlighting lower-value games and offering additional benefits to encourage greater engagement.

This analysis demonstrates the usefulness of the Steam API for collecting valuable user data and the effectiveness of the Weka software in mining this data. The combination of these tools allows for deeper and more accurate analysis of user behavior, providing valuable information for strategic decision-making and service personalization in the gaming industry.

Keywords: Digital Games; Consumption Patterns; Data Analysis

LISTA DE ABREVIATURAS OU SIGLAS

API - Application Programming Interface (Interface de Programação de Aplicativos).

ID - Identity (identidade).

CSV - Comma-separated values (valores separados por vírgulas).

SUMÁRIO

1 INTRODUÇÃO.....	8
1.1 OBJETIVOS DA PESQUISA.....	8
1.1.1 Objetivos Gerais.....	9
1.1.2 Objetivos Específicos.....	9
1.2 JUSTIFICATIVA.....	9
2 REFERENCIAL TEÓRICO.....	12
2.1 COMPORTAMENTO DO CONSUMIDOR.....	12
2.2 ESTATÍSTICA E POPULAÇÃO.....	12
2.3 APIs.....	13
2.4 PLATAFORMA STEAM.....	13
2.5 STEAMDB.....	14
2.6 BANCO DE DADOS.....	15
2.7 WEKA.....	15
2.8 MÉTODO DE APRENDIZAGEM NÃO SUPERVISIONADO.....	15
2.9 STEAM LADDER.....	16
2.10 LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS.....	16
2.11 CONFIGURAÇÕES DE PERFIL DA STEAM.....	16
2.12 CRAWLER.....	18
2.13 CSV (COMMA-SEPARATED VALUES).....	18
2.14 K-Means.....	18
3 METODOLOGIA DE PESQUISA.....	19
4 RESULTADOS.....	23
4.1 JOGOS COMPRADOS POR VALOR DA CONTA.....	28
4.2 JOGOS ADQUIRIDOS POR JOGADOS MAIS DE 12 HORAS.....	29
4.3 JOGOS ADQUIRIDOS POR JOGOS NUNCA JOGADOS.....	30
4.4 VALORES DE JOGOS POR QUANTIDADE DE JOGOS COMPRADOS.....	30
4.5 JOGOS SEM VALOR POR JOGOS COMPRADOS.....	34
4.6 JOGOS COMPRADOS POR CLUSTERS.....	35
5 CONCLUSÕES.....	36
6 TRABALHOS FUTUROS.....	37
REFERÊNCIAS.....	39

1 INTRODUÇÃO

É de conhecimento geral que na última década os jogos digitais ficaram mais comuns entre os brasileiros, desde a criança com seu videogame, até o idoso em seu celular. No último ano, o mercado de jogos brasileiro movimentou cerca de 1,6 milhões de dólares (Newzoo, 2022), ou seja, o brasileiro está entre os maiores consumidores do mercado mundial de jogos.

Existem poucos estudos (Vladilava,2021; Lindgren, 2010; Lopes Ferreira,2017) que buscam como os jogadores compram seus jogos e como escolhe quais irá gastar seu dinheiro, informação que buscaremos responder, utilizando bases de dados públicas de perfis de usuários da Steam, uma das maiores lojas online de jogos de Computador, e quais tipos de jogos ele prefere, sua taxa de abandono do produto, se as compras são feitas por impulso e etc.

Durante as pesquisas em periódicos como o SciELO não foram encontradas pesquisas brasileiras utilizando dados de usuários da Steam para realizar estudos comportamentais, mas ao expandirmos para o Google Scholar, encontramos artigos que abordam fatos como sistemas de conquistas e medalhas nos perfis dos usuários (Rizani, M.N.; Khalid, M.N.A.; Iida, H, 2023) ou dados dos usuários como quantidade de posts, vídeos postados, nível na plataforma e etc. para definir em grupos relacionados (Li, X., Lu, C., Peltonen, J., & Zhang, Z. 2019), então um estudo analisando valor monetário, horas gastas, porcentagem de jogos jogados pode auxiliar a compreender o comportamento consumidor de forma mais direta e assim auxiliar a área.

No trabalho proposto serão informados dados que justifiquem a escolha da plataforma escolhida para a pesquisa, quais as questões que buscam ser respondidas com a análise dos dados, qual a metodologia utilizada para se conseguir realizar essa análise, e a fundamentação teórica que auxilia ao melhor entendimento do estudo como um todo.

1.1 OBJETIVOS DA PESQUISA

Os Objetivos de Pesquisa deste trabalho são divididos em Gerais e Específicos:

1.1.1 OBJETIVOS GERAIS

Este artigo de pesquisa tem como objetivo investigar como o usuário brasileiro jogador compra e aproveita seus jogos digitais, respondendo seus padrões de consumo de acordo com horas jogadas, porcentagem de jogos que foram jogados, e quais as faixas de preços que o usuário costuma gastar, sugerindo assim uma nova forma de realização de pesquisa do mercado consumidor de jogos brasileiro.

1.1.2 OBJETIVOS ESPECÍFICOS

- Criar sistema de agrupamentos dos usuários de acordo com suas particularidades.
- Identificar padrões de consumo do público brasileiro na plataforma.
- Identificar a faixa de preço mais comum dos jogos que o público brasileiro tem interesse, identificando assim quais os tipos de jogos que são mais prováveis de grande números de vendas de acordo com seu valor.
- Realizar a formatação dos dados de horas de jogo, quantidade de jogos na biblioteca, e valores de conta, adquiridos na plataforma Steam e como eles se relacionam no sistema da plataforma.

1.2 JUSTIFICATIVA

Segundo a Pesquisa da Indústria Brasileira de Games (2022), são encontrados cerca de 12 mil empregos relacionados no Brasil com a área de desenvolvimento de jogos, demarcando assim um crescimento de 600% no mercado, faturando mais de 80 milhões de reais no último ano. No ano de 2014 foi realizado o Primeiro Censo da Indústria Brasileira de Jogos Digitais, pela GEDIGames, que aborda informações acerca da indústria dos games, como os desenvolvedores, associações, perfil corporativo, comercial e outras diversas informações, demonstrando que o mercado está em crescimento na área de desenvolvimento internamente no Brasil. No segundo censo, realizado em 2018 pela

Homo Ludens, é apresentado os dados, onde, 72% das empresas brasileiras de games atuam no mercado nacional, e 54% tinham planejamento de atuar futuramente no mesmo.

Desde 2013 é realizada a Pesquisa Game Brasil, em formato de questionário, com foco no público brasileiro e em outros países-chave da América latina, tendo como foco identificar o perfil demográfico, sociocultural e comportamental, mas não focando em seu perfil de consumo. Nessa pesquisa, foi identificado que 74,5% dos brasileiros jogam jogos eletrônicos, onde 48% preferem jogar em Smartphones, 23% em computadores e 20% em videogames. Na pesquisa, é questionado quais plataformas os usuários utilizam, e grande parte delas, como Steam, Xbox Live, Epic Online, Origin, Uplay e Battle.net se encontram disponíveis em computadores, demonstrando assim que existem diversas plataformas ligadas ao PC, totalizando assim a plataforma como a que possui maior número de jogos disponíveis, e, destas citadas, a Steam se encontra como a mais antiga, online desde 2003, com a maior biblioteca disponível a compra (cerca de 50,261 jogos no início de 2023), e o maior número de usuários ativos, cerca de de 120 milhões. Devido a esses motivos, foi entendido que um estudo na área é necessário para melhor o entendimento sobre o padrão de consumo dos usuários brasileiros, e, a plataforma Steam foi visualizada como a principal escolha para o estudo pela sua data de lançamento e grande volume de jogos disponíveis na biblioteca em comparação com as suas concorrentes, chegando a mais de 50 mil jogos (Statista, 2023a e Statista, 2023b) .

Vladilava (2021) conclui que na área de marketing os principais desejos que importam são o da criatividade e imersão no produto final, e devem ser o foco das campanhas publicitárias no mercado dos games, que o entendimento sobre as relações de motivação, engajamento e intenção de compra podem auxiliar o processo de criação de games e alocação de recursos. Também é demonstrado que existem somente conhecimentos teóricos sobre o tema, não tendo pesquisas que compreendam os fatores motivacionais para o ato de jogar e comprar jogos, e que a análise das relações no processo de jogos pode ser realizada em futuras pesquisas, analisando lançamentos virais, ou seja, que tem grande espera do público, podem auxiliar um melhor entendimento do mercado ao todo, informação combinada com os dados da pesquisa da Arstechnica(Orland, 2014), onde cerca de 37% dos 781

milhões de jogos vendidos ao todo na plataforma Steam nunca foram jogados pelos seus donos, evidencia que boa parte da motivação da compra de um jogo é movida pelo impulso de compra.

Analisando as pesquisas encontradas no assunto com origem brasileira, observamos que as mesmas utilizam questionários, o que se faz também outra necessidade de desenvolvimento do estudo, visto que com a análise de dados diretos da plataforma, podemos acumular uma maior quantidade de dados em comparativo com a população utilizada nos questionários.

2 REFERENCIAL TEÓRICO

Para entendermos melhor os conceitos que utilizaremos neste estudo, iremos desenvolvê-los de forma separada, para que todos os assuntos sejam abordados e juntos auxiliem na compreensão dos temas abordados de forma conjunta.

2.1 COMPORTAMENTO DO CONSUMIDOR

Resume-se ao conjunto de ações que tenham relação com o processo de consumo ou aquisição de serviços, bens de consumo, atividades, experiências, pessoas ou ideias, pelo processo de escolhas pelo tempo (Hoyer et al,2012) ou seja, como indivíduos, grupos e organizações selecionam, usam e disponibilizam produtos e bens de consumo, para satisfazer suas necessidades e desejos (Solomon M.,2006). Existem teorias de como o comportamento funciona, sendo que a A Teoria da Ação Racional (TAR - Martin Fishbein, 1963) funciona como uma base para a que iremos utilizar no estudo, ela busca explicar de forma matemática que os seres humanos são racionais e utilizam as informações encontradas para avaliar seu próprio comportamento, e assim definirem o que será realizado. Nessa teoria, podemos encontrar o fator das crenças, informações que o indivíduo tem sobre o objeto, atributos, etc., e essas podem ser comportamentais, ou seja, intrigada ao indivíduo, ou normativas, as quais são julgadas pelo meio que ele se encontra, a pressão social que sofrerá caso realize ou não tal comportamento.

2.2 ESTATÍSTICA E POPULAÇÃO

A estatística engloba uma série de conceitos fundamentais para a compreensão e análise de dados. A população refere-se ao conjunto completo de elementos que são de interesse na pesquisa, enquanto a amostra é um subconjunto representativo selecionado da população. Para o seguinte estudo, utilizaremos o cálculo de tamanho da amostra, que nos retorna o intervalo de confiança, de acordo com a seguinte fórmula:

$$z \cdot \frac{\sigma}{\sqrt{n}}$$

Onde n é o tamanho da amostra, σ é o desvio padrão populacional desejado e z é o score de nível de confiança desejado, encontrado em uma tabela de distribuição normal padrão. Neste estudo, utilizaremos o nível de confiança de 95%, ou seja, $Z=1.96$ (TRIOLA,2017).

2.3 APIs

API é a sigla para Application Programming Interface (Interface de Programação de Aplicativos, em português). Ela é um conjunto de regras e protocolos que permite a comunicação e interação entre diferentes softwares. Em termos simples, uma API define como os componentes de software devem se comunicar e interagir uns com os outros. Uma API atua como uma camada de abstração que permite que os desenvolvedores acessem e utilizem funcionalidades específicas de um software ou serviço sem precisar conhecer todos os detalhes internos de sua implementação. Através de uma API, os desenvolvedores podem enviar solicitações e receber respostas estruturadas, utilizando um conjunto de métodos e parâmetros definidos(IBM,2023a)

As APIs podem ter diferentes finalidades e funcionalidades. Elas podem permitir que um software acesse recursos de outro software, como bancos de dados, serviços web, sistemas operacionais, bibliotecas de funções, entre outros. As APIs também podem fornecer funcionalidades específicas, como processamento de pagamentos, integração de redes sociais, acesso a dados em tempo real e muitas outras.

2.4 PLATAFORMA STEAM

A plataforma Steam é uma plataforma de distribuição digital de jogos para computadores desenvolvida pela Valve Corporation. Ela foi lançada em 2003 e se tornou uma das principais plataformas de jogos para PC (VALVE, 2023).

2.5 STEAMDB

A SteamDB é uma plataforma online que fornece uma variedade de informações e estatísticas sobre jogos disponíveis na plataforma Steam. Ela atua como uma ferramenta para os usuários explorarem e acompanharem dados relacionados a jogos, preços, atualizações, análises de popularidade e muito mais. A principal função do SteamDB é monitorar e coletar dados diretamente da plataforma Steam, oferecendo aos usuários acesso a informações detalhadas sobre os jogos. Essas informações podem incluir dados sobre o número de jogadores ativos, conquistas desbloqueadas, vendas, análises de popularidade, histórico de preços, atualizações e até mesmo informações técnicas sobre os jogos, como requisitos de sistema e configurações (SteamDB,2023).

A importância da SteamDB reside na sua capacidade de fornecer dados e informações detalhadas aos usuários da Steam. Ela ajuda os jogadores a tomarem decisões informadas sobre quais jogos comprar, acompanhar o desempenho de jogos populares, identificar tendências e descobrir novos jogos com base em suas preferências. Além disso, a plataforma também é útil para desenvolvedores e profissionais da indústria de jogos, pois oferece insights e análises valiosas sobre o mercado e a performance dos jogos.

É importante mencionar que a SteamDB não é afiliada à Valve Corporation, a empresa por trás da plataforma Steam, mas é uma fonte de dados confiável e amplamente utilizada por jogadores e entusiastas da indústria de jogos, visto que ela busca os dados diretamente da API da Steam.

2.6 BANCO DE DADOS

Um banco de dados é um conjunto organizado de informações estruturadas que são armazenadas e gerenciadas em um sistema computacional. Ele é projetado para armazenar, recuperar, atualizar e gerenciar grandes volumes de dados de forma eficiente e confiável. Os bancos de dados são amplamente utilizados em diversas áreas, desde empresas e organizações governamentais até aplicações pessoais. Eles fornecem uma estrutura centralizada para armazenar dados de forma persistente, permitindo o acesso rápido e seguro às informações quando necessário(IBM,2023b).

2.7 WEKA

Weka é um software de código aberto desenvolvido pela Universidade de Waikato, na Nova Zelândia, que fornece uma ampla gama de ferramentas e algoritmos para análise de dados e mineração de dados. O nome "Weka" é uma abreviação para "Waikato Environment for Knowledge Analysis". É amplamente utilizado por pesquisadores, cientistas de dados e profissionais da área de aprendizado de máquina e mineração de dados. Ele oferece uma interface gráfica intuitiva que permite aos usuários carregar conjuntos de dados, pré-processá-los, visualizá-los e aplicar diversos algoritmos de análise de dados(Weka, 2023).

2.8 MÉTODO DE APRENDIZAGEM NÃO SUPERVISIONADO

O aprendizado não supervisionado é uma abordagem em que os dados não são rotulados ou classificados previamente, e o objetivo é encontrar padrões, estruturas ou agrupamentos intrínsecos nos dados (IBM,2023c).

2.9 STEAM LADDER

Steam Ladder é um site e serviço online que oferece informações e estatísticas detalhadas sobre usuários e jogos da plataforma de distribuição digital de jogos para PC, a Steam. A plataforma Steam Ladder rastreia e compila dados de perfis de jogadores da Steam, fornecendo rankings, conquistas, estatísticas de jogos, horas jogadas, entre outros dados (Steam Ladder,2023).

O objetivo principal da Steam Ladder é fornecer aos usuários da Steam uma plataforma para acompanhar e comparar seu progresso em jogos, ver seu ranking em relação a outros jogadores e descobrir informações detalhadas sobre suas estatísticas de jogos. Além disso, a Steam Ladder oferece recursos de busca e filtragem para explorar jogadores e jogos específicos, bem como a possibilidade de criar listas de amigos e grupos.

No geral, a Steam Ladder é uma ferramenta útil para os jogadores da Steam que desejam monitorar seu desempenho, comparar-se com outros jogadores e explorar estatísticas e conquistas de jogos.

2.10 LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS

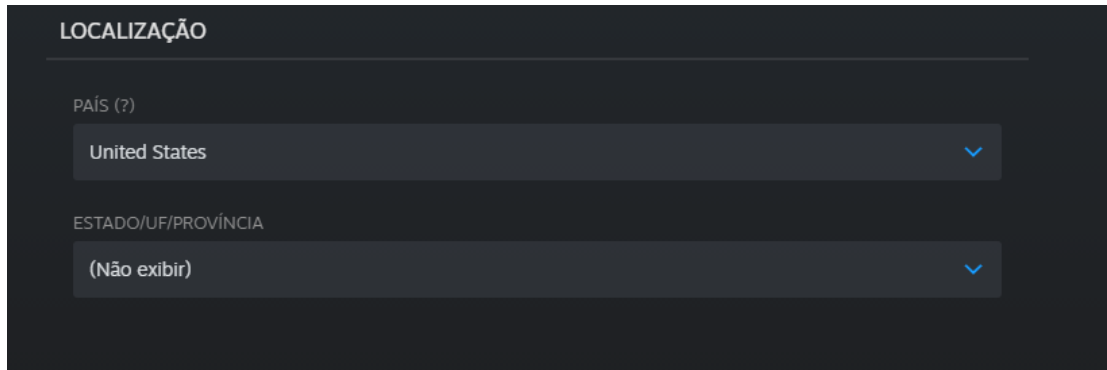
Para a realização do estudo, devemos armazenar dados das IDs dos usuários encontrados, situação que deve-se evitar de acordo com a Lei nº 13709 (Brasil, 2018), a chamada Lei geral de proteção de dados pessoais, portanto, após feita a mineração dos dados, devemos descaracterizá-los de forma que não seja possível encontrar o perfil original do usuário visado, e os IDs utilizados devem ser descartados.

2.11 CONFIGURAÇÕES DE PERFIL DA STEAM

A plataforma *Steam* fornece configurações de perfil dos usuários da Steam, sendo este, é quem define se seus dados serão públicos, se a lista de jogos será pública e a nacionalidade da conta, portanto os dados visados neste estudo podem estar sujeitos a erros, visto que os usuários podem ou não ser brasileiros, e vários

usuários não possam ser analisados devido a publicidade ou não de seus dados (Figura 2 e 3).

Figura 2 - Configuração de localização do perfil.



The image shows the 'LOCALIZAÇÃO' (Location) settings page on Steam. It features two dropdown menus. The first is labeled 'PAÍS (?)' (Country) and is currently set to 'United States'. The second is labeled 'ESTADO/UF/PROVÍNCIA' (State/Province) and is currently set to '(Não exibir)' (Do not display). Both dropdowns have a blue downward arrow icon on the right side.

FONTE: STEAM (2023)

Figura 3 - Configurações de privacidade do perfil na Steam



The image shows the 'Privacidade de dados básicos' (Basic data privacy) settings page on Steam. At the top right, there is a link 'Voltar ao seu perfil' (Return to your profile). The main heading is 'Privacidade de dados básicos: Pública'. Below this, a paragraph explains that the user's name and avatar are visible to everyone who accesses their profile and are used for identification in multiplayer games. It also mentions that the profile background and selected mini-profiles will be visible, even with a private profile. The next section is 'Privacidade do perfil: Pública', which includes a dropdown arrow. Below this, a paragraph states that the user's community profile includes a summary, friends list, badges, Steam level, highlights, comments, and group participation. It then says 'Os seguintes elementos podem ser controlados individualmente:' (The following elements can be controlled individually:). The next section is 'Privacidade de jogos: Pública', which includes a dropdown arrow. Below this, a paragraph explains that it includes a list of all games on the user's Steam account, games in the wishlist, achievements, and play time. It also mentions that the user can choose whether the 'In-game' status will be displayed next to the game name. At the bottom, there is a checkbox labeled 'Não exibir o meu tempo em jogos para outros usuários.' (Do not display my game time to other users.), which is currently unchecked.

FONTE: STEAM (2023)

2.12 CRAWLER

Um "crawler" é um programa de software ou um script que percorre a web de forma sistemática, automatizada e metódica para coletar informações de diferentes sites. Também conhecido como "spider" ou "web crawler", ele navega pelas páginas da web, segue os links internos e extrai dados relevantes, como texto, imagens, URLs e metadados(IBM, 2022a).

2.13 CSV (COMMA-SEPARATED VALUES)

CSV (Comma-Separated Values) é um formato de arquivo amplamente utilizado para armazenar dados tabulares. Como o nome sugere, os valores são separados por vírgulas em um arquivo CSV. É um formato simples e comum que pode ser facilmente lido e manipulado por programas de planilhas, bancos de dados e outras aplicações(The Internet Society, 2005).

2.14 K-MEANS

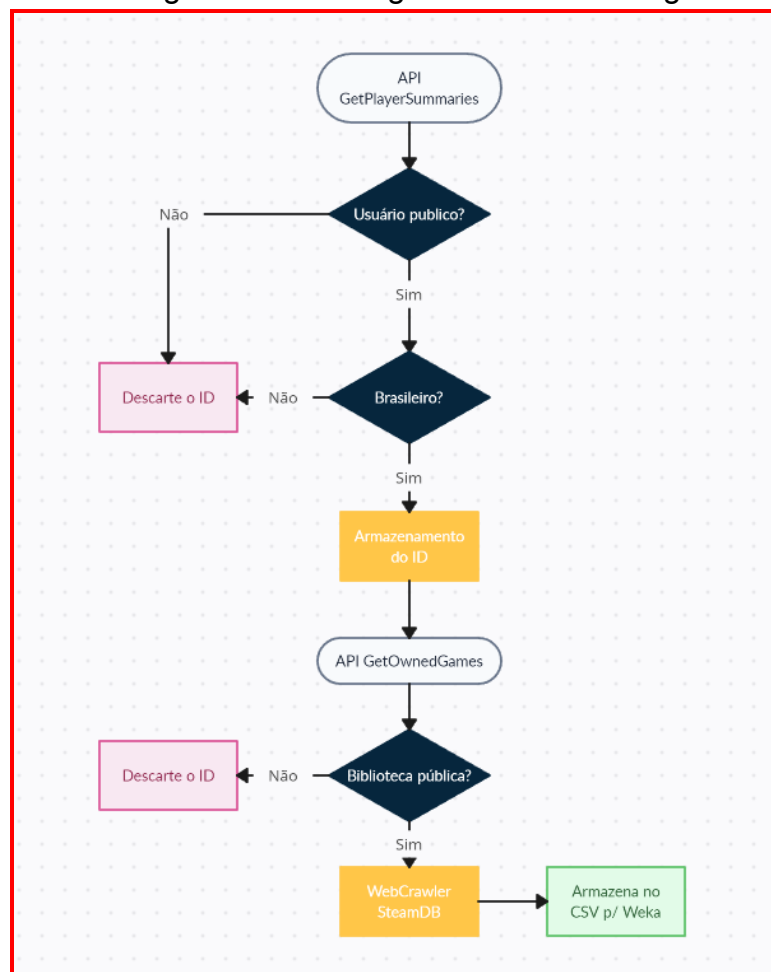
O K-means é um algoritmo de aprendizado de máquina não supervisionado utilizado para agrupar dados em clusters. O objetivo do algoritmo é particionar um conjunto de dados em grupos ou clusters, onde os pontos dentro de um mesmo cluster são mais semelhantes entre si do que com pontos de outros clusters(IBM,2022b).

O resultado final do algoritmo K-means é um conjunto de "k" clusters, onde cada ponto de dado pertence a um cluster específico. O algoritmo busca minimizar a variância dentro de cada cluster e maximizar a variância entre os clusters.

É importante observar que o resultado do algoritmo K-means pode variar dependendo da inicialização aleatória dos centroides, e diferentes execuções podem levar a diferentes agrupamentos. Portanto, é comum executar o algoritmo várias vezes e escolher o resultado que apresenta o melhor critério de avaliação, como a menor soma dos erros quadrados ou o maior coeficiente de silhueta.

3 METODOLOGIA DE PESQUISA

Fluxograma 1 - Fluxograma da Metodologia



FONTE: O AUTOR (2023)

Para a realização deste trabalho, será aplicada a metodologia (fluxograma 1) utilizada por Orlando (2014a,2014b) no seu texto *Introducing Steam Gauge: Ars reveals steam's most popular games*, é demonstrado que para se acessar perfil da Steam utiliza-se o ID, um código de 17 números decimais individual de cada conta, iniciados no número 76561197960265729, que caminha de forma sequencial a partir do mesmo, possuindo a partir disso cerca de 172 milhões de páginas, com alguns sendo inválidos, acessando assim todas as páginas de perfis de usuários da plataforma. Com um *crawler*, um sistema que percorre a web coletando dados de forma automática, acessando todos os perfis existentes, criando uma base de dados que possua as informações necessárias para o estudo. Utilizando dados publicados na plataforma Statista (2018) cerca de 4.77% da população total da Steam em abril

de 2018 é brasileira, o que é cerca de 4 milhões dos usuários, dos 90 milhões ativos naquele ano (Statista, 2021).

Para o engajamento dos dados será utilizada as APIs, onde a API da Steam se acessa com o ID do usuário, criando assim um banco de dados para o armazenamento e tratamento dos mesmos, utilizando filtros como país de origem da conta e se a conta é pública ou não, e por meio de análises descritiva e diagnóstica serão desenvolvidos os resultados, buscando responder as questões pertinentes a esse estudo, como padrões de consumo de tempo de jogo, valores monetários e médias entre os dados. A análise descritiva demonstra dados como padrão de consumo, faixa de preço de jogos mais comprados, enquanto a diagnóstica irá tentar achar padrões de compras ligados a compras por impulsos e porcentagem de jogos jogados. A pesquisa se qualifica como quantitativa, visto que será realizada a coleta dos dados e aplicação dos mesmo em softwares analíticos, com o objetivo de apresentar os resultados e interpretar as respostas subjetivas encontradas. Para o armazenamento dos dados será criado um banco de dados em planilha CSV, e a aplicaremos no software Weka, para identificarmos padrões com auxílio da mineração de dados.

Para a mineração dos dados citada, as seguintes APIs disponíveis da Steam e o sistema SteamDB para o levantamento dos dados estão disponíveis:

GetPlayerSummaries: Retorna dados como URL do usuário, ID da conta, código de país e privacidade da conta.

GetOwnedGames: Retorna dados como a lista de jogos do ID informado, total de horas jogadas nas últimas 2 semanas, em toda a vida da conta.

SteamDB: Site que busca armazenar dados da plataforma da Steam, utilizando as APIs da Steam para realizar buscas e demonstrar de forma simplificada ao usuário os dados encontrados. Fornece um sistema de calculadora de contas que informa dados retirados das outras APIs citadas acima, de forma simplificada.

Deve-se levar em consideração as limitações impostas pelos desenvolvedores das plataformas utilizadas, onde nas APIs da Steam pode-se realizar duas mil pesquisas diárias por chave de utilizador, um código de acesso a API dada a cada usuário da Steam, e a limitação temporária de cerca de 70 pesquisas diárias por IP na SteamDB, o que criou uma necessidade de uso de um

sistema de mudança de IP para evitar-se o bloqueio e comprometimento temporal da pesquisa. Para a identificação dos IDs dos usuários, foi realizada a filtragem das listas fornecidas na *Steam Ladder*, focando, assim, nos usuários que possuem dados disponíveis na plataforma, ainda sendo necessário identificar dados como país de origem e a abertura da conta.

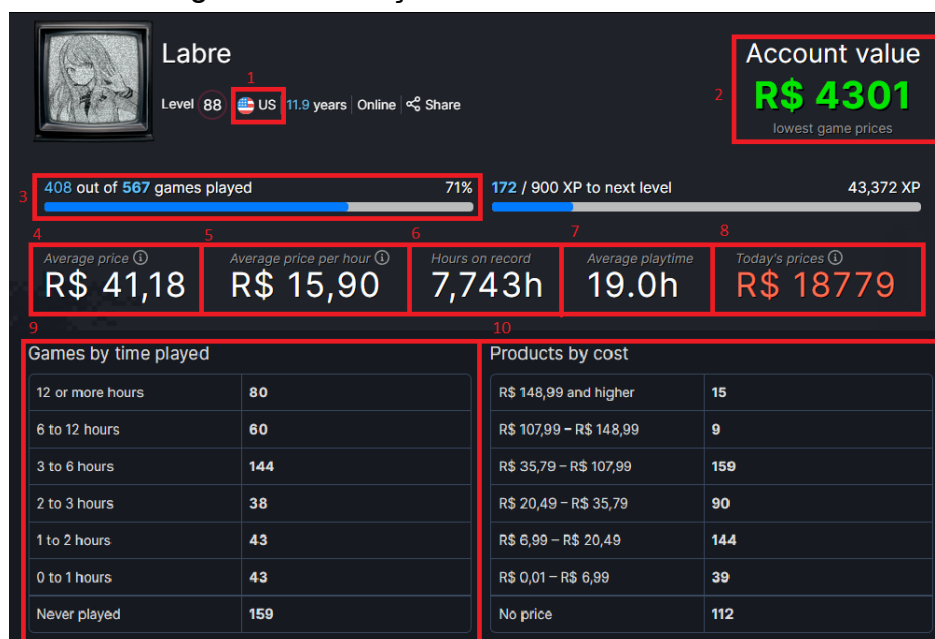
Para o processo foram utilizadas em conjunto as três plataformas citadas, da seguinte forma:

1 - Utilizando a API *GetPlayerSummaries*, iremos verificar se esse usuário encontrado tem sua Steam como Brasileira e se é pública, caso sim, iremos armazenar o seu ID.

2 - Posteriormente, iremos realizar a utilização da API *GetOwnedGames*, que retorna a lista de jogos, visto que ter uma conta pública na plataforma é diferente de ter sua biblioteca pública, então caso retorne ao menos um jogo como resposta, a biblioteca do usuário é pública e poderemos retornar as informações de sua conta

3 - Utilizando um *Crawler*, realizaremos a busca na plataforma *SteamDB*, retornando os seguintes dados e inserindo-os em uma planilha CSV para aplicação no software Weka, para posterior análise. Os dados encontrados na página da SteamDB e suas definições são as seguintes (Figura 1).

Figura 1 - Definição dos dados encontrados.



FONTE: STEAMDB (2023)

1. País de identificação do usuário (o país da sua conta é definido pelo usuário, sendo assim a única forma de identificação de nacionalidade possível em uma conta de forma pública).
2. Valor da conta, considerando os menores valores encontrados para os jogos cadastrados no perfil do usuário, em toda a vida corrida do jogo.
3. Porcentagem de jogos que o usuário jogou, ao menos uma vez.
4. Preço médio de jogos, ou seja, o valor da conta dividido pelo número total de jogos possuídos.
5. Média de valor entre preço total e número de horas jogadas.
6. Número de horas jogadas.
7. Tempo médio de jogo em cada jogo.
8. Preço nos valores atuais da conta, considerando os preços encontrados no dia da pesquisa.
9. Número de jogos jogados na conta, de acordo com a distribuição de horas, sendo esses divididos em 7 classificações, de 12 horas ou mais de jogo, a nunca jogados.
10. Jogos encontrados divididos por seus custos, sendo estes divididos em 7 classificações, de valores acima de R\$148,99 reais, até jogos de graça.

Após a criação do CSV, utilizou-se o software Weka para encontrar similaridades entre os usuários encontrados.

4 RESULTADOS

Utilizando o estudo disponível na plataforma Statista, foi realizado o cálculo do tamanho de amostra necessária, onde com o grau de confiança de 95%, margem de erro de 5%, retornou um tamanho da amostra de 400 usuários. Para o estudo, utilizou-se o tamanho da amostra de 1534 usuários da plataforma, tendo assim uma margem de erro de 3%, número encontrado de acordo com o que foi obtido durante a utilização dos *crawlers* e após a filtragem dos mesmos, excluindo os dados daqueles que não possuíam valores nos seus perfis ou horas jogadas.

Utilizando o shell do Linux foram utilizados os seguintes comandos:

```
curl https://steamladder.com/ladder/friends/br/ --output most-friends-br.txt
curl https://steamladder.com/ladder/game_bans/br/ --output mostgameban-br.txt
curl https://steamladder.com/ladder/steam_age/br/ --output steam-age-br.txt
curl https://steamladder.com/ladder/playtime/br/ --output playtime-br.txt
```

Esses comandos geram os arquivos .txt que possuem os IDs selecionados, que foram filtrados com os seguintes comandos:

```
grep 'onclick=\"location.href' games-br.txt |grep -v donate|grep -v steamladder
|awk -F '/' '{print $3}' e “grep -v donate|grep -v steamladder”
```

Resultando assim em um arquivo .txt que foi utilizado para aplicação nas APIs

Utilizando um script Python realizamos a filtragem dos usuários que são brasileiros e tem sua configuração de perfil como pública, e posteriormente utilizamos o segundo script para realizarmos a filtragem daqueles que possuíam sua biblioteca de jogos, resultando assim em 1426 usuários identificados como válidos para a análise. Foi utilizado em sequência um script para a extração dos dados, que gerou respostas em arquivos e foi necessário transformar eles para arquivos HTML, os quais foram analisados e foi criado o script, que iria alocar as informações necessárias em um único arquivo .csv (Figura 4) para ser utilizado no software Weka.

Figura 4 - Exemplo da planilha gerada

	A	B	C	D	E	F	G	H	I	J
1	SteamID	Valor da Conta	No price	R\$ 0.01 – R\$ 6.99	R\$ 6.99 – R\$ 20.49	R\$ 20.49 – R\$ 35.79	R\$ 35.79 – R\$ 107.99	R\$ 107.99 – R\$ 148.99	R\$ 14.899 and higher	Games owned
2	39	27	4	0	0	0	3	0	0	7
3	40	28	14	1	2	4	2	0	0	23
4	55	58	38	1	8	3	6	0	0	54
5	72	109	1	0	6	2	0	0	0	9
6	73	109	0	0	6	2	0	0	0	8
7	76	109	2	0	6	2	0	0	0	10
8	79	109	1	0	6	2	0	0	0	9
9	83	115	5	0	1	0	1	0	0	7
10	88	134	15	20	11	8	7	0	0	61
11	110	200	45	58	12	14	15	1	0	143
12	112	204	7	0	2	1	7	0	0	17
13	117	217	38	0	3	11	6	1	0	59
14	126	242	11	17	12	2	2	1	1	48
15	137	276	24	37	13	6	3	0	3	88
16	141	299	3	0	1	2	1	0	0	7
17	148	335	67	15	14	14	12	1	1	124
18	154	370	33	226	51	7	3	0	0	320
19	191	515	45	11	6	7	12	5	3	89
20	200	523	47	108	95	24	27	0	3	302
21	205	547	123	105	73	29	20	0	1	351
22	213	601	30	0	10	3	12	2	3	60
23	236	697	132	149	95	49	22	0	0	447
24	243	738	3	0	8	6	1	0	0	18
25	250	754	114	43	48	38	37	2	1	281
26	263	791	161	88	48	42	29	1	0	369
27	265	811	5	0	0	1	4	2	1	13
28	271	851	3	115	27	8	12	2	4	171
29	272	859	105	141	78	53	24	1	1	403
30	280	895	109	134	62	19	23	1	2	350
31	288	923	209	68	47	43	59	1	5	430
32	295	980	88	38	55	37	48	3	6	269
33	319	1065	24	427	68	19	13	2	3	558
34	320	1099	144	225	149	72	60	2	4	658
35	323	1126	41	6	4	18	20	2	4	95
36	325	1137	69	31	26	19	28	4	4	181
37	331	1171	471	255	101	53	43	2	1	928
38	345	1259	1483	180	108	55	50	1	4	1839
39	347	1301	30	3	11	19	60	3	11	137
40	349	1322	52	24	15	10	35	5	15	158
41	375	1486	7	0	9	5	3	0	0	24
42	378	1497	48	775	284	53	26	1	0	1187
43	386	1598	5	0	7	6	3	0	1	22
44	391	1628	100	597	282	33	36	1	2	1051
45	394	1656	22	13	28	2	2	0	0	67
46	395	1659	153	67	46	40	53	2	9	370
47	396	1686	43	5	11	6	16	2	7	90
48	404	1722	55	303	194	46	57	5	10	670
49	417	1819	62	3	6	2	6	1	0	80
50	429	1980	128	92	58	44	54	7	14	395
51	433	1999	148	98	94	65	70	1	3	477

J	K	L	M	N	O	P	Q	R	S	T	U
owned	Games Bought	Never played	0 to 1 hours	1 to 2 hours	2 to 3 hours	3 to 6 hours	6 to 12 hours	12 or more hours	Total de Horas Jogadas	Jogos Jogados	Média de tempo p
3	7	0	0	0	0	0	0	0	0	0	0.00
9	23	0	0	0	0	0	0	0	0	0	0.00
18	54	0	0	0	0	0	0	0	0	0	0.00
8	9	0	0	0	0	0	0	0	0	0	0.00
8	8	0	0	0	0	0	0	0	0	0	0.00
8	10	0	0	0	0	0	0	0	0	0	0.00
8	9	0	0	0	0	0	0	0	0	0	0.00
2	7	0	0	0	0	0	0	0	0	0	0.00
48	56	0	0	0	0	0	0	0	0	0	0.00
98	137	0	0	0	0	0	0	0	0	0	0.00
10	16	0	0	0	0	0	0	0	0	0	0.00
21	58	0	0	0	0	0	0	0	0	0	0.00
35	45	0	0	0	0	0	0	0	0	0	0.00
62	85	0	0	0	0	0	0	0	0	0	0.00
4	7	0	0	0	0	0	0	0	0	0	0.00
57	122	0	0	0	0	0	0	0	0	0	0.00
287	320	0	0	0	0	0	0	0	0	0	0.00
44	89	0	0	0	0	0	0	0	0	0	0.00
255	298	0	0	0	0	0	0	0	0	0	0.00
228	348	0	0	0	0	0	0	0	0	0	0.00
30	55	0	0	0	0	0	0	0	0	0	0.00
315	444	0	0	0	0	0	0	0	0	0	0.00
15	18	0	0	0	0	0	0	0	0	0	0.00
167	277	0	0	0	0	0	0	0	0	0	0.00
208	363	0	0	0	0	0	0	0	0	0	0.00
8	13	0	0	0	0	0	0	0	0	0	0.00
168	171	0	0	0	0	0	0	0	0	0	0.00
298	400	0	0	0	0	0	0	0	0	0	0.00
241	348	0	0	0	0	0	0	0	0	0	0.00
221	418	0	0	0	0	0	0	0	0	0	0.00
183	265	0	0	0	0	0	0	0	0	0	0.00
532	553	0	0	0	0	0	0	0	0	0	0.00
512	650	0	0	0	0	0	0	0	0	0	0.00
54	95	0	0	0	0	0	0	0	0	0	0.00
112	179	0	0	0	0	0	0	0	0	0	0.00
455	910	0	0	0	0	0	0	0	0	0	0.00
376	1803	0	0	0	0	0	0	0	0	0	0.00
107	135	0	0	0	0	0	0	0	0	0	0.00
104	151	0	0	0	0	0	0	0	0	0	0.00
17	24	0	0	0	0	0	0	0	0	0	0.00
1139	1184	0	0	0	0	0	0	0	0	0	0.00
17	22	0	0	0	0	0	0	0	0	0	0.00
951	1047	0	0	0	0	0	0	0	0	0	0.00
45	65	0	0	0	0	0	0	0	0	0	0.00
217	364	0	0	0	0	0	0	0	0	0	0.00
47	67	0	0	0	0	0	0	0	0	0	0.00
615	669	0	0	0	0	0	0	0	0	0	0.00
18	78	0	0	0	0	0	0	0	0	0	0.00
269	381	0	0	0	0	0	0	0	0	0	0.00
331	472	0	0	0	0	0	0	0	0	0	0.00

FONTE: O AUTOR (2023)

Para a análise foi utilizado o pré-processamento para que todos os dados fossem normalizados, para assim facilitar a visualização dos dados, e posteriormente os dados de ID excluídos da análise, visto que eram somente para identificação. Para identificarmos os melhores atributos, utilizamos a área de Selecionar atributos disponível no Weka, com isso, podemos identificar os atributos que possuem maior significância para a análise. Utilizando o processo de correlação, normalmente referido como coeficiente de correlação de Pearson, podemos encontrar o grau de correlação entre duas variáveis em escala métrica, mas na ferramenta utilizada, ele retorna um ranking de importância dos atributos(Figura 5).

Figura 5 - Ranking de Atributos

```
=== Attribute Selection on all input data ===

Search Method:
    Attribute ranking.

Attribute Evaluator (supervised, Class (numeric): 18 Total de Horas Jogadas):
    Correlation Ranking Filter

Ranked attributes:
0.352859 17 12 or more hours
0.167662 2 No price
0.128479 9 Games owned
0.109664 14 2 to 3 hours
0.093852 11 Never played
0.069725 4 R$ 6.99 â€" R$ 20.49
0.053858 3 R$ 0.01 â€" R$ 6.99
0.052602 10 Games Bought
0.036102 5 R$ 20.49 â€" R$ 35.79
0.01662 1 Valor da Conta
0.014557 13 1 to 2 hours
-0.000945 6 R$ 35.79 â€" R$ 107.99
-0.010462 15 3 to 6 hours
-0.010994 12 0 to 1 hours
-0.019522 7 R$ 107.99 â€" R$ 148.99
-0.020152 16 6 to 12 hours
-0.025619 8 R$ 14.899 and higher

Selected attributes: 17,2,9,14,11,4,3,10,5,1,13,6,15,12,7,16,8 : 17
```

FONTE: O AUTOR (2023)

Para melhor identificação do dataset em geral, foi aplicado a clusterização dos dados, utilizando o processo de SimpleKMeans, um método de aprendizado não-supervisionado, onde os dados são separados em grupos de dados ou clusters baseados em uma ou mais medidas/características similares(Figura 6).

Figura 6 - Criação dos Clusters pelo SimpleKMeans

Final cluster centroids:								
Attribute	Cluster#							
	Full Data (1425.0)	0 (63.0)	1 (918.0)	2 (39.0)	3 (109.0)	4 (33.0)	5 (18.0)	6 (245.0)
Valor da Conta	0.0409	0.0092	0.0088	0.3263	0.1437	0.0835	0.3064	0.0528
No price	0.0485	0.0576	0.0093	0.1147	0.0689	0.8327	0.4723	0.0369
R\$ 0.01 até R\$ 6.99	0.0281	0.0127	0.0061	0.1505	0.083	0.0577	0.3968	0.0395
R\$ 6.99 até R\$ 20.49	0.0551	0.0169	0.0094	0.3744	0.1805	0.1131	0.7658	0.0694
R\$ 20.49 até R\$ 35.79	0.0692	0.019	0.0155	0.5181	0.2295	0.1417	0.7067	0.0839
R\$ 35.79 até R\$ 107.99	0.0814	0.0175	0.0221	0.5856	0.2851	0.1476	0.5316	0.1072
R\$ 107.99 até R\$ 148.99	0.0551	0.0115	0.0161	0.3702	0.1933	0.1122	0.3221	0.0736
R\$ 14.899 and higher	0.0424	0.0074	0.0153	0.2359	0.1375	0.0923	0.2154	0.0605
Games owned	0.062	0.0395	0.0126	0.311	0.1597	0.4623	0.6977	0.0692
Games Bought	0.0492	0.0163	0.0105	0.3187	0.1581	0.0991	0.5932	0.0646
Never played	0.0439	0.0291	0.0071	0.2322	0.1032	0.4107	0.5626	0.0418
0 to 1 hours	0.0419	0.0117	0.016	0.2032	0.1242	0.115	0.1862	0.064
1 to 2 hours	0.0304	0.0091	0.007	0.2047	0.0871	0.0701	0.2943	0.0459
2 to 3 hours	0.0223	0.0061	0.0048	0.1517	0.067	0.0406	0.266	0.0316
3 to 6 hours	0.043	0.0148	0.0093	0.3049	0.1649	0.0679	0.3179	0.057
6 to 12 hours	0.043	0.0109	0.0121	0.2859	0.1525	0.0739	0.2889	0.0574
12 or more hours	0.0059	0.0181	0.0021	0.0144	0.0083	0.0308	0.0721	0.0061
Total de Horas Jogadas	0.0322	0.4083	0.0105	0.0147	0.018	0.0653	0.1445	0.013
Time taken to build model (full training data) : 0.03 seconds								
=== Model and evaluation on training set ===								
Clustered Instances								
0	63 (4%)							
1	918 (64%)							
2	39 (3%)							
3	109 (8%)							
4	33 (2%)							
5	18 (1%)							
6	245 (17%)							

FONTE: O AUTOR (2023)

Cruzando os dados com os ranking identificado anteriormente, foram definidas as seguintes associações a serem analisadas:

4.1 JOGOS COMPRADOS POR VALOR DA CONTA

Na distribuição dos casos, pode-se perceber que grande parte das contas ficam acumuladas na esquerda inferior do gráfico, ou seja, baixos valores de conta com poucos jogos comprados, demonstrando assim que o grupo de usuários analisado não possui grande variância de perfil de consumo no caso, sendo observados casos isolados de contas com valores altos e pouco volume de jogos, ou grande volume de jogos mas com valores de conta baixo(Figura 7).

Figura 7 - Cluster - Jogos comprados por Valor da Conta

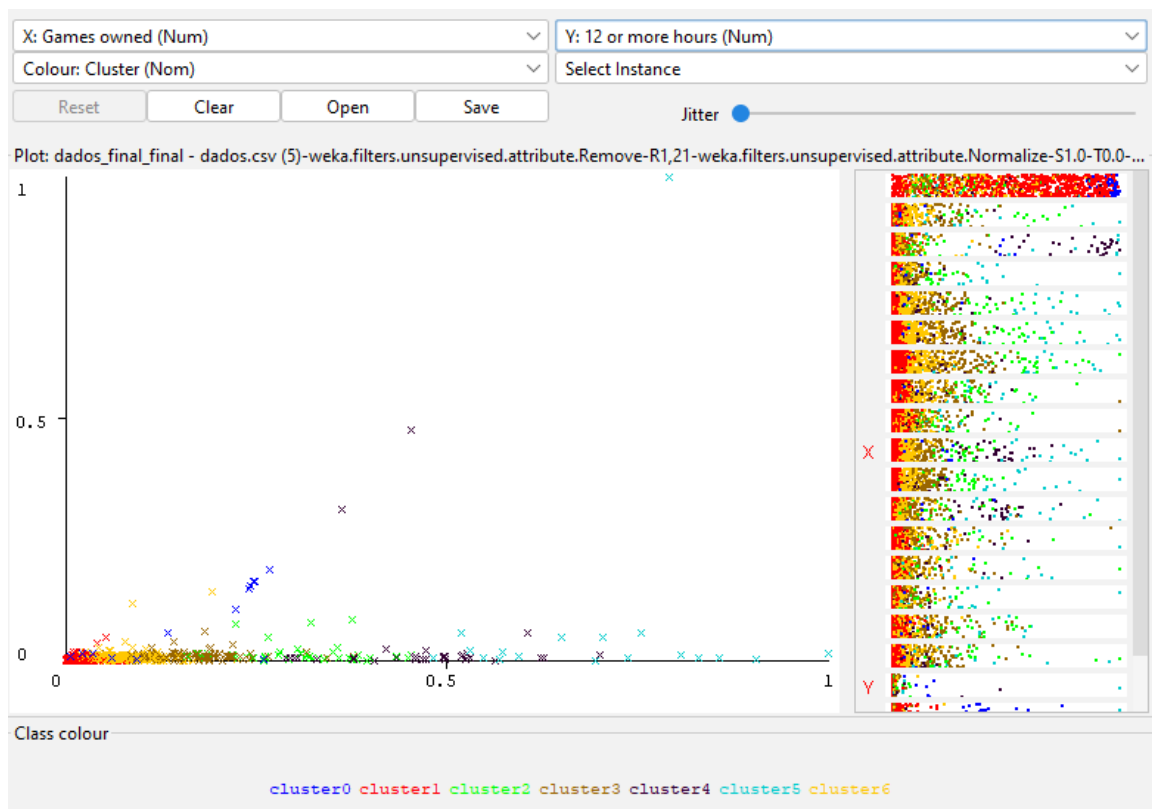


FONTE: O AUTOR (2023)

4.2 JOGOS ADQUIRIDOS POR JOGADOS MAIS DE 12 HORAS

A grande concentração de casos próximos de zero no eixo Y demonstra que os usuários do grupo analisado possuem poucos jogos em que ele jogou mais de 12 horas, demonstrando assim que os usuários analisados não possuem o costume de focar somente em um jogo por um longo tempo. O grupo analisado tem um padrão de não jogar os jogos que possui por tempo considerado extenso, demonstrando assim que os mesmos possam ter problemas com o aproveitamento dos produtos comprados(Figura 8).

Figura 8 - Jogos Adquiridos por jogadores mais de 12 horas

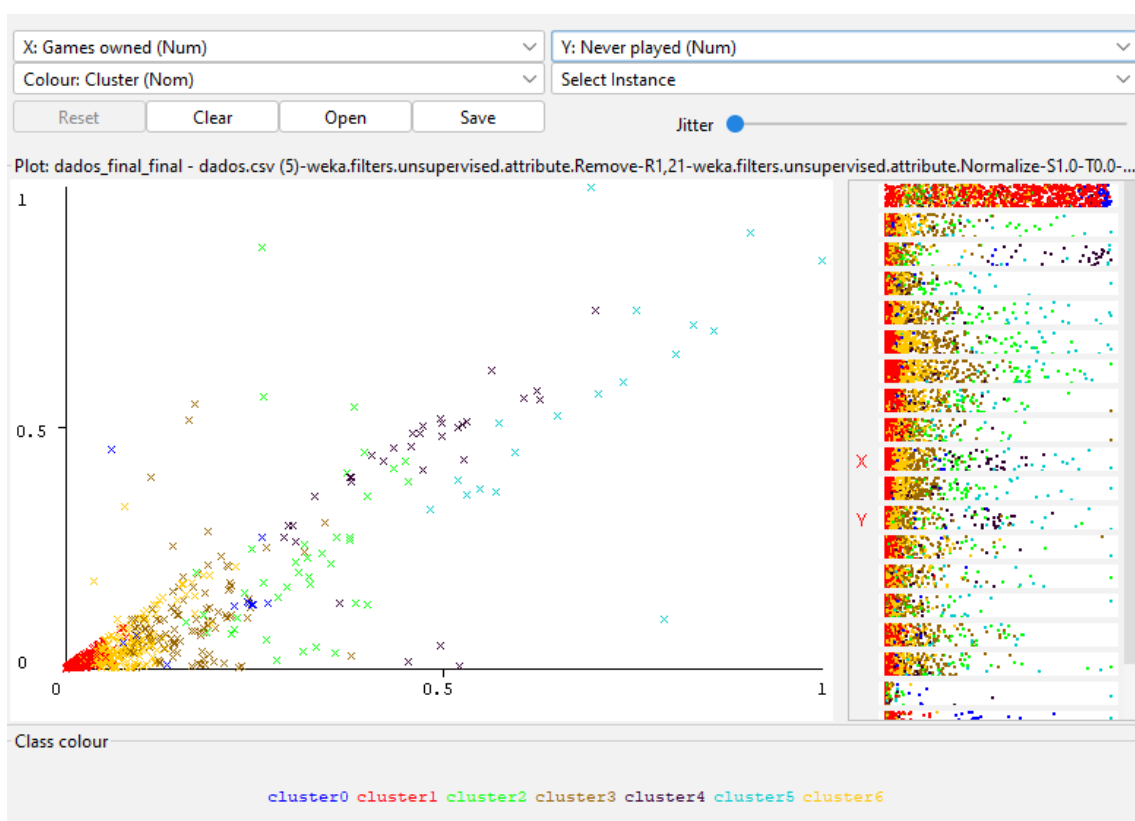


FONTE: O AUTOR (2023)

4.3 JOGOS ADQUIRIDOS POR JOGOS NUNCA JOGADOS

Com a análise da distribuição (Figura 9), observa-se uma particularidade de existir uma linha no meio da plotagem do gráfico, demonstrando que grande parte das ocorrências possuem números próximos de jogos adquiridos na conta e jogos nunca jogados. O grupo analisado tem a característica de comprar jogos e não os jogar, deixando-os somente em suas bibliotecas e realizando o acúmulo dos produtos.

Figura 9 - Jogos adquiridos por jogos nunca jogados



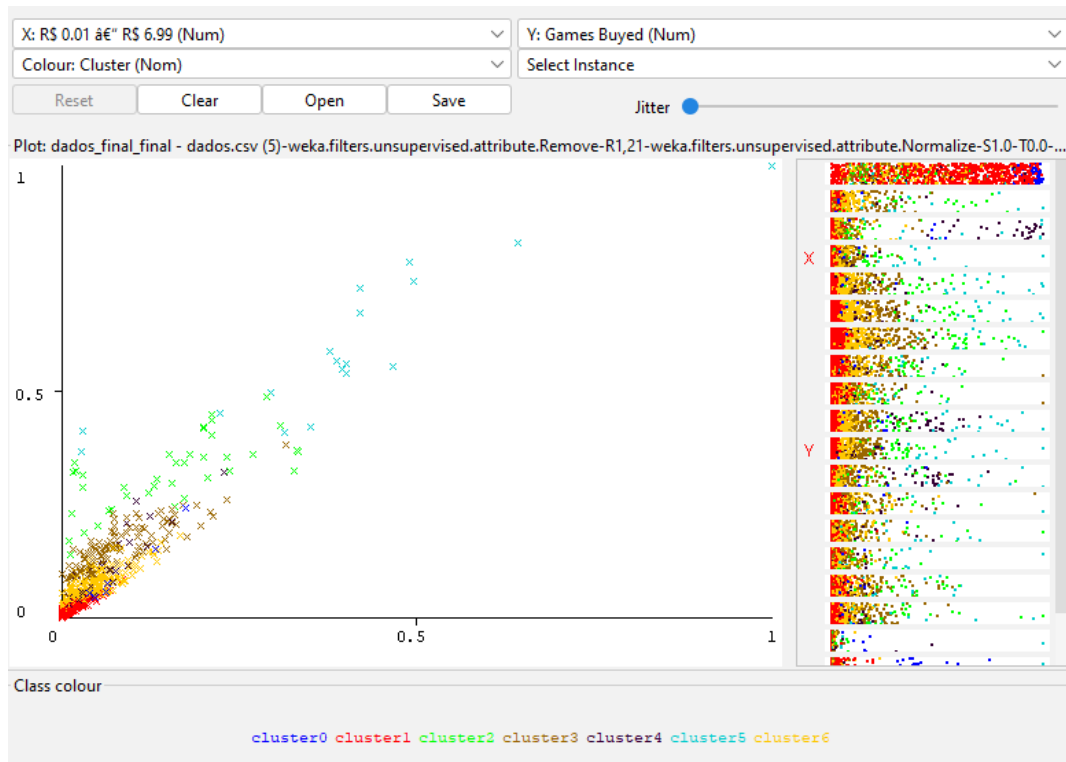
FONTE: O AUTOR (2023)

4.4 VALORES DE JOGOS POR QUANTIDADE DE JOGOS COMPRADOS

Realizando o comparativo dos gráficos(Figura 10 a 15), observamos que somente o gráfico de jogos com valores de R\$0.01 a R\$6.99 é possível visualizarmos uma concentração de dados acima da média entre X e Y, demonstrando assim que o brasileiro tem a preferência em compra de jogos nesses

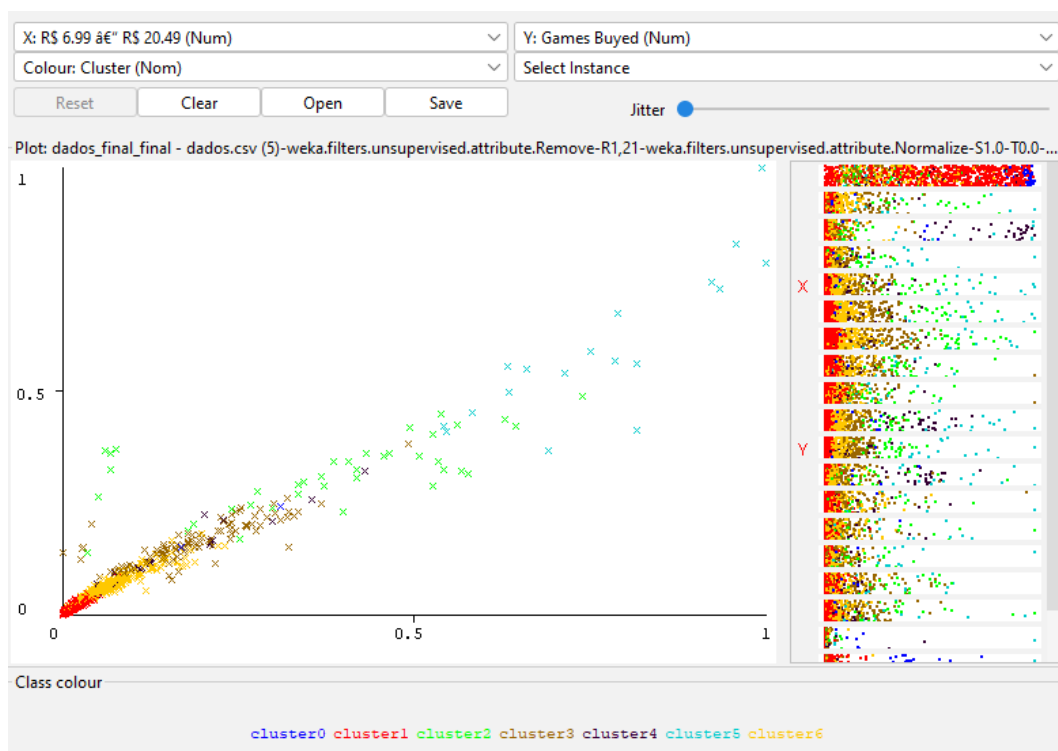
valores, e ao observarmos os outros valores, é possível perceber que quanto maior o valor, mais concentrado ao ponto 0.0 o gráfico se torna. Podemos assim definir que grande parte do grupo analisado tem preferência por jogos de valores mais baixos, visto que esses correspondem a uma parcela maior de suas bibliotecas.

Figura 10 - Jogos de R\$0.01 a R\$6.99 por jogos comprados



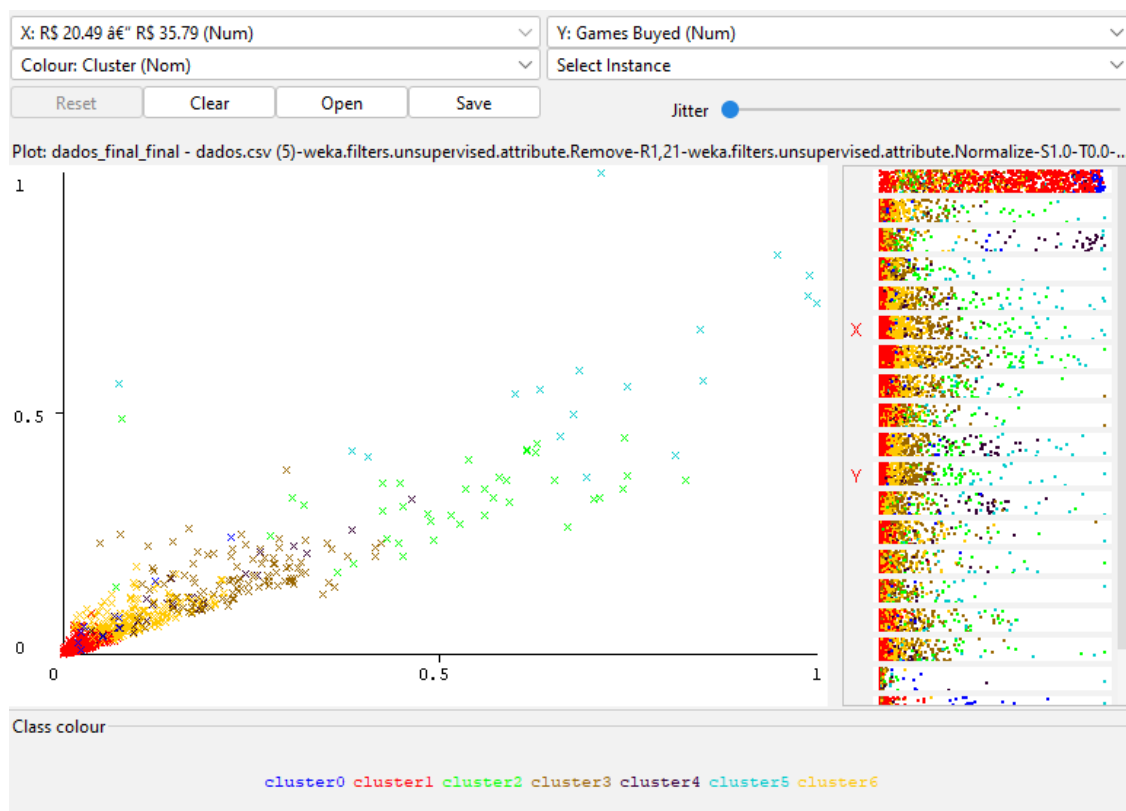
FONTE: O AUTOR (2023)

Figura 11 - Jogos de R\$6.99 a R\$20.49 por jogos comprados



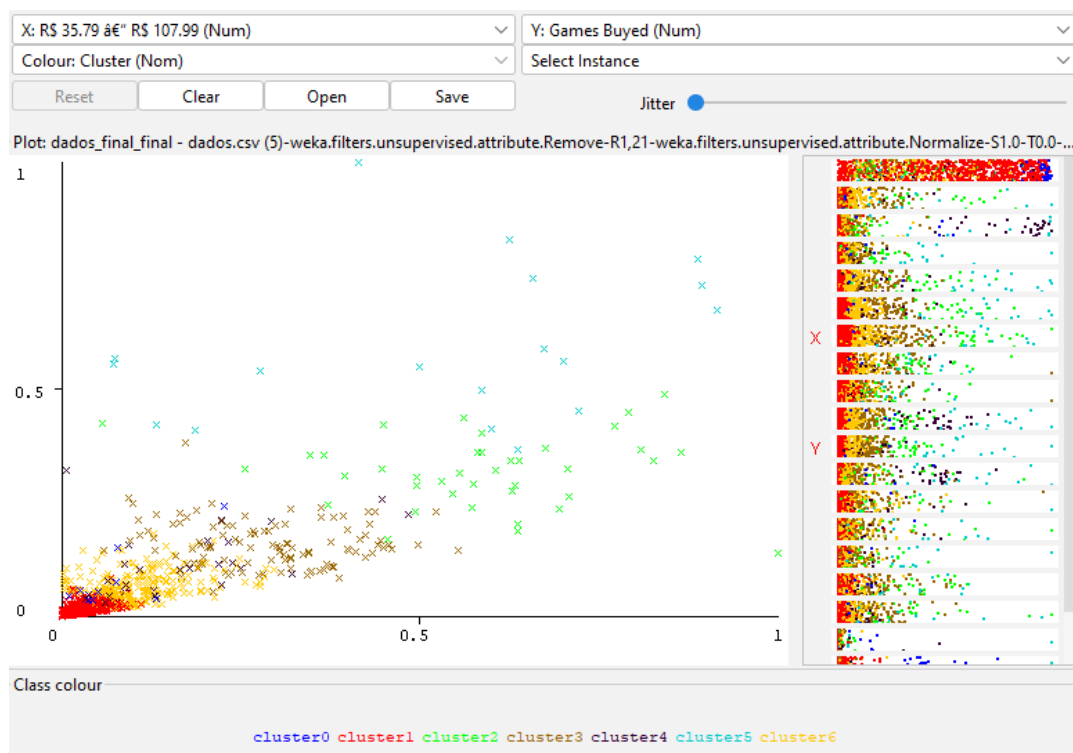
FONTE: O AUTOR (2023)

Figura 12 - Jogos de R\$20.50 a R\$35.79 por jogos comprados



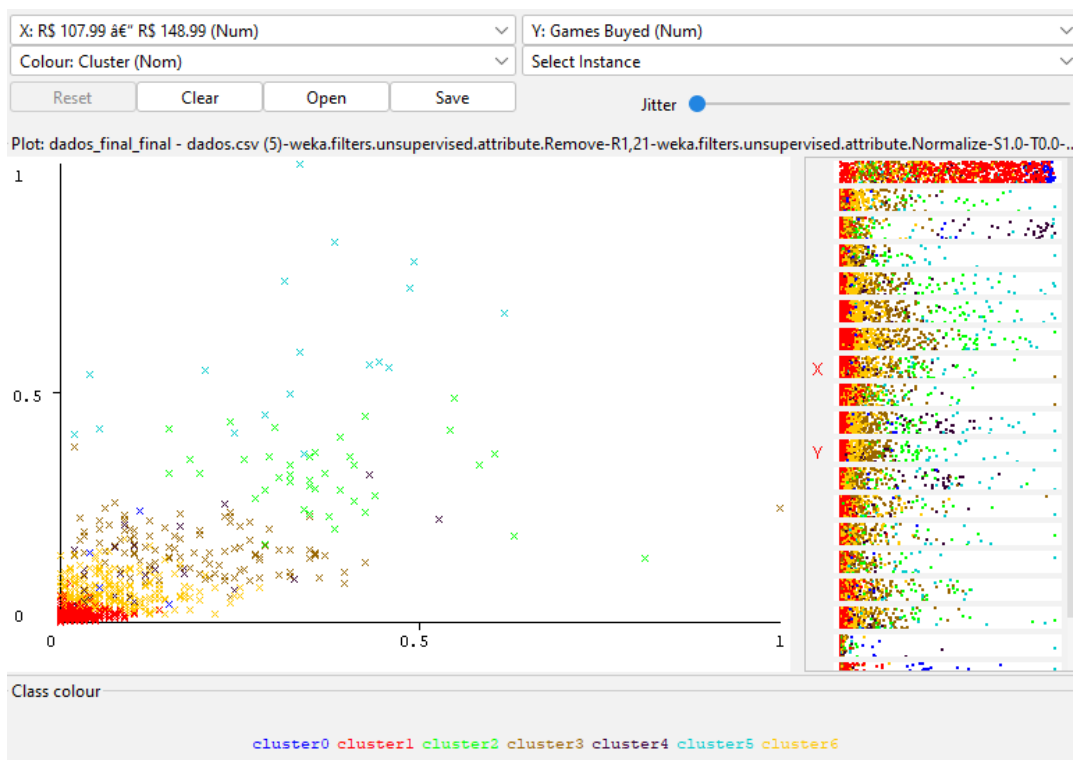
FONTE: O AUTOR (2023)

Figura 13 - Jogos de R\$35.80 a R\$107.99 por jogos comprados



FONTE: O AUTOR (2023)

Figura 14 - Jogos de R\$108.00 a R\$148.99 por jogos comprados



FONTE: O AUTOR (2023)

Figura 15 - Jogos de R\$148.99 ou mais por jogos comprados

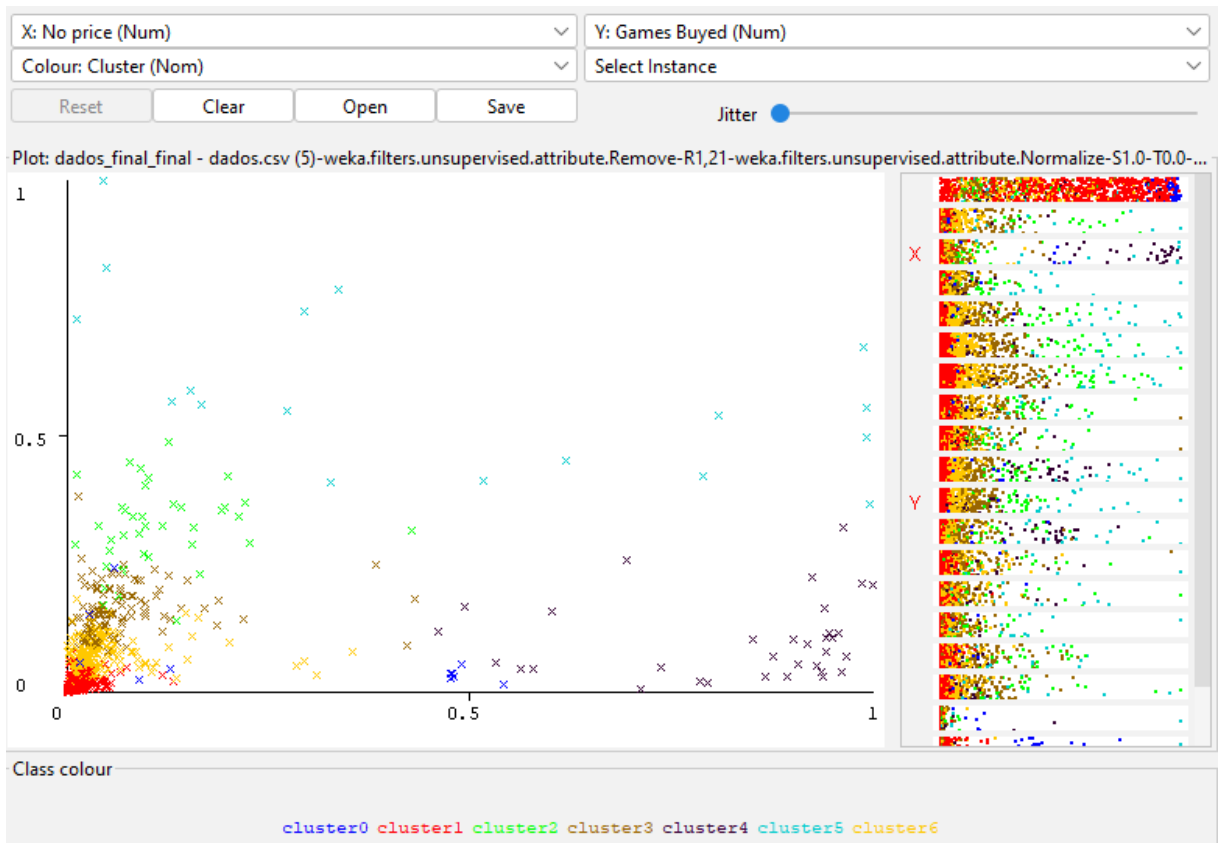


FONTE: O AUTOR (2023)

4.5 JOGOS SEM VALOR POR JOGOS COMPRADOS

A quantidade de jogos gratuitos não está relacionada com a quantidade de jogos comprados, visto a ausência de casos na área central superior do quadro(Figura 16), mas podemos notar grande acúmulo de usuários com um número de jogos comprados similar ao número de jogos gratuitos na conta, demonstrando que grande parte dos jogadores tem o padrão de jogar produtos gratuitos disponíveis na plataforma.

Figura 16 - Jogos sem valor por Jogos comprados

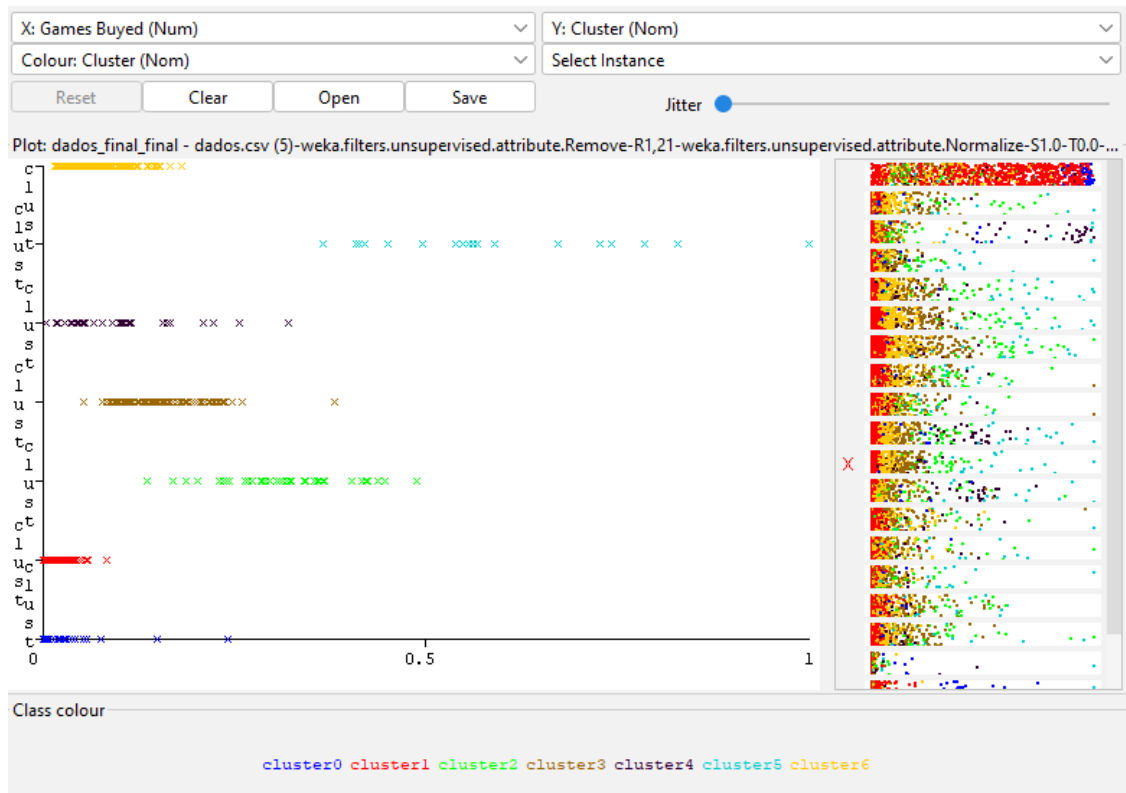


FONTE: O AUTOR (2023)

4.6 JOGOS COMPRADOS POR CLUSTERS

A clusterização(Figura 17) levou em conta a quantidade de jogos comprados, criando áreas em comuns entre os grupos de acordo com a quantidade de jogos comprados, mas não tendo grupos totalmente em comum entre eles. Com a análise dos outros dados cruzados com os clusters, pode-se encontrar áreas em comum de cada cluster, e assim criar grupos específicos e até mesmo nomeá-los de acordo com suas características.

Figura 17 - Jogos comprados por Clusters



FONTE: O AUTOR (2023)

5 CONCLUSÕES

Verificamos que a parcela de usuário brasileiro analisados da *Steam* busca por jogos de menor valor, e aqueles que possuem maiores valores de conta buscam variar sua biblioteca com jogos baratos, uma certa forma de colecionismo, situação provavelmente motivada pela parte do ranqueamento da plataforma quanto a conquistas e quantidade de jogos disponíveis na biblioteca. Outra característica que se pode observar é que o colecionismo favorece a questão de se ter vários jogos que nunca foram jogados, situação que compromete questões de mercado, como a análise de uso e tempo médio dos jogos.

Retornando a Teoria da Ação Racional, podemos estabelecer relação entre os temas de consumismo e colecionismo. O consumismo é um comportamento caracterizado pelo consumo excessivo e pela busca constante por novos produtos. Ele pode ser influenciado pelas atitudes das pessoas em relação aos bens de

consumo. De acordo com a TAR, a intenção de compra de um determinado produto é determinada pela atitude que o consumidor tem em relação a esse produto, bem como pela avaliação subjetiva da utilidade e das consequências dessa compra. No caso analisado, podemos relacionar a relação do consumidor com o marketing imposto pelo jogo, o que o usuário espera ao adquirir o jogo na plataforma e como ele irá jogar, como isso irá relacionar ele dentro da plataforma e se ele se ele está incluso em um grupo que esteja jogando e comentando sobre o produto. Quanto ao colecionismo, que envolve a acumulação e a exibição de objetos específicos, a TAR pode ser aplicada de forma semelhante. Os colecionadores podem desenvolver atitudes positivas em relação à atividade de colecionar, baseadas na satisfação pessoal de possuir itens raros, no prazer estético ou no sentido de identidade que a coleção proporciona. A intenção de expandir uma coleção ou adquirir novos itens pode ser influenciada pela atitude do colecionador em relação a essa atividade e pelos benefícios percebidos em adquirir novos itens. No caso estudado, podemos relacionar o ato de colecionar a exibição de uma grande biblioteca, maior variedade de itens para seu perfil, ou até mesmo o ranqueamento de sua conta em sistemas como o Steam Ladder.

Sobre os valores dos jogos, quanto maior o valor do *game*, menos compras, demonstrando assim que o usuário brasileiro não tem grande foco em comprar os chamados Triplos As, os jogos de grandes estúdios que normalmente passam o valor de R\$150 reais, dessa forma que o foco do grupo analisado é em jogos de menor custo, informação que deve ser levado em consideração para a região.

Quanto à clusterização encontrada, grande parte dos usuários se encontra em um mesmo grupo, demonstrando assim que grande parte dos usuários tem um padrão de consumo e horas jogadas em comum.

6 TRABALHOS FUTUROS

Devido a limitações das APIs utilizadas, o estudo teve que se limitar a um grupo específico, mas a situação pode ser contornada com mais tempo disponível para o desenvolvimento da base de dados, que pode ser aplicada em conjunto a outras informações disponíveis por meio de outras APIs.

Realizar análises dos clusters desenvolvidos, para que se crie grupos com definições aplicáveis para toda a plataforma, de forma que auxiliem o usuário a compreender melhor seus padrões de consumo e aproveitamento de jogos, características que auxiliam na identificação do usuário na plataforma e grupo pertencente no meio.

A análise pode ser aplicada em diversas regiões ou grupos, realizar estudos com o objetivo de criar um padrão de consumo regional para compreendermos melhor o mercado de certas regiões e assim ter melhor entendimento de como realizar o lançamento de jogos e faixas de preços aplicáveis de acordo com o público da região alvo.

REFERÊNCIAS

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet)**. Brasília, DF: Presidência da República; 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm . Acesso em 22 de junho de 2023.

DEAN, Brian. **Steam Usage and Catalog Stats for 2022**. Disponível em: <https://backlinko.com/steam-users> . Acesso em 18 de agosto de 2022.

FORTIM, Ivelise (Org). **Pesquisa da indústria brasileira de games 2022**. Abragames : São Paulo, 2022. Apresentação em Português: <https://www.abragames.org/pesquisa-da-industria-brasileira-de-games.html>

GALYONKIN, Sergey. **SteamSPY**. Disponível em: <https://steamspy.com/> . Acesso em 21 de junho de 2022.

GEDIGames, NPGT. **I Censo da Indústria Brasileira de Jogos Digitais, com Vocabulário Técnico IBJD**. Julho, 2014. Disponível em: https://www.abragames.org/uploads/5/6/8/0/56805537/i_censo_da_industria_brasileira_de_jogos_digitais_2.pdf . Acesso em 17 de setembro de 2022.

GUO, Yue; BARNES, Stuart. **Why people buy virtual items in virtual worlds with real money**. ACM SIGMIS Database: The DATABASE for Advances in Information Systems, [S.L.], v. 38, n. 4, p. 69-76, 28 out. 2007. Association for Computing Machinery (ACM). Disponível em: https://www.researchgate.net/publication/220628024_Why_People_Buy_Virtual_Items_in_Virtual_Worlds_with_Real_Money . Acesso em 02 de setembro de 2022.

HOYER, Wayne D. **Comportamento do consumidor** / Wayne D. Hoyer, Deborah J. Macinnis; revisão técnica Richard Vinic, Alessandra Vinic; [tradução EZ2 Translate]. São Paulo: Cengage Learning, 2011.

IBM. **What is an API?**. 2023a. Disponível em: <https://www.ibm.com/topics/api>

IBM. **Databases.** 2023b. Disponível em: <https://www.ibm.com/docs/en/db2/11.5?topic=administration-databases>

IBM. **What is unsupervised learning?**. 2023c. Disponível em: <https://www.ibm.com/topics/unsupervised-learning>

IBM. **Crawlers.** 2022a. Disponível em: <https://www.ibm.com/docs/en/watson-explorer/12.0.x?topic=explorer-crawlers>

IBM. **K-Means Clustering.** 2022b. Disponível em: <https://www.ibm.com/docs/en/db2/11.5?topic=building-k-means-clustering>

LINDGREN, Henri. **Factors Contributing to the buying decision of PC and Video games.** Disponível em: https://www.theseus.fi/bitstream/handle/10024/21183/Henri_Lindgren_.pdf?s

Acesso em 02 de setembro de 2022.

LI, Xiaozhou et al. **A statistical analysis of Steam user profiles towards personalized gamification.** In: GamiFIN Conference, 2019. Disponível em: <https://ceur-ws.org/Vol-2359/paper19.pdf> .

HOMO LUDENS. **II Censo da Indústria Brasileira de Jogos Digitais.** 2018. Disponível em: <https://censojogosdigitais.com.br/wp-content/uploads/2020/03/AF-IICenso-completo.pdf> >. Acesso em 17 de Setembro de 2022.

MONTGOMERY, D.C.; RUNGER, G.C.; HUBELE, N.F. **Estatística Aplicada e Probabilidade para Engenheiros**. LTC Editora, 2012.

MOUTINHO, Karina; ROAZZI, Antonio. **As teorias da ação racional e da ação planejada: relações entre intenções e comportamentos**. Aval. psicol. [online], vol. 9, n. 2, pp. 279-287, 2010. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000200012&lng=pt&nrm=iso . ISSN 1677-0471.

NEWZOO. **Brazilian Games Market Consumer Insights: Brazil's Mobile Players Are Likelier to Play Competitive & Midcore Games**. 2022. Disponível em: <https://newzoo.com/resources/blog/brazilian-games-market-consumer-insights-brazil-s-mobile-players-are-likelier-to-play-competitive-midcore-games>

OLSSON, Björn; SIDENBLOM, Louise. **Business Models for Video Games. Department of Informatics, Lund University**. Junho de 2010. Disponível em: <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=1672034&fileId=1672035> . Acesso em 17 de setembro de 2022.

ORLAND, Kyle. **Introducing Steam Gauge: Ars reveals Steam's most popular games**. 15 de abril de 2014a. Disponível em: <https://arstechnica.com/gaming/2014/04/introducing-steam-gauge-ars-reveals-steam-s-most-popular-games/> . Acesso em 17 de setembro de 2022.

ORLAND, Kyle. **Steam Gauge: Addressing your questions and concerns**. 16 de abril de 2014b. Disponível em: <https://arstechnica.com/gaming/2014/04/steam-gauge-addressing-your-questions-and-concerns/> . Acesso em 17 de setembro de 2022.

PUSPITASARI, B. **Consumer's Buying Decision-Making Process in E-Commerce**. Disponível em: https://www.e3s-conferences.org/articles/e3sconf/pdf/2018/06/e3sconf_icenis2018_11003.pdf . Acesso em 02 de setembro de 2022.

RIZANI, M.N.; KHALID, M.N.A.; IIDA, H. **Application of Meta-Gaming Concept to the Publishing Platform**: Analysis of the Steam Games Platform. Information, v. 14, n. 2, p. 110, 2023. Disponível em: <https://doi.org/10.3390/info14020110> .

SOUZA, Lucas Lopes Ferreira de; FREITAS, Ana Augusta Ferreira de. **Consumer behavior of electronic games' players**: a study on the intentions to play and to pay. Revista de Administração, [S.L.], v. 52, n. 4, p. 419-430, out. 2017. Elsevier BV. Disponível em: <http://dx.doi.org/10.1016/j.rausp.2017.08.004> .

SOLMON, Michael R. **O comportamento do consumidor**: comprando, possuindo e sendo. 11. ed. Porto Alegre: Bookman, 2016.

STAZIAKI. P.V. **The Birth of the Eighth Art**. Lifebar, Medium. 8 de abril de 2016. Disponível em: <https://medium.com/the-lifebar/the-birth-of-the-eighth-art-52feced27b38> . Acesso em 17 de setembro de 2022.

STATISTA. **Number of games released on Steam worldwide from 2004 to 2022**. 2023a. Disponível em: <https://www.statista.com/statistics/552623/number-games-released-steam/>

STATISTA. **Number of games available in the Epic Games Store from 2019 to 2022**. 2023b. Disponível em: <https://www.statista.com/statistics/1234037/epic-games-store-games-available/>

STATISTA. **Distribution of Steam users worldwide as of April 2018**, by country. 2022. Disponível em: <https://www.statista.com/statistics/826870/steam-distribution-country/> . Acesso em 18 de agosto de 2022.

STATISTA. **Number of peak concurrent Steam users from January 2013 to September 2021.** Disponível em: <https://www.statista.com/statistics/308330/number-stream-users/>.

STEAMDB. **Frequently Asked Questions.** 2023. Disponível em: <https://steamdb.info/faq/>

STEAMLADDER. **Frequently Asked Questions.** 2023. Disponível em: <https://steamladder.com/faq/>

SX GROUP/GO GAMERS. **Pesquisa Game Brasil 2022.** Disponível em: <https://www.pesquisagamebrasil.com.br/pt/> . Acesso em 03 de setembro de 2022.

THE INTERNET SOCIETY. **Common Format and MIME Type for Comma-Separated Values (CSV) Files.** 2005. Disponível em: <https://www.rfc-editor.org/rfc/rfc4180.txt>

TRIOLA, M.F. **Introdução à Estatística.** LTC Editora, 2017.

VALVE. **Steam, a plataforma definitiva de jogos on-line.** 2023. Disponível em: <https://store.steampowered.com/about/>

VALVE. **Documentação da Web API do Steam.** Disponível em: <https://steamcommunity.com/dev?l=portuguese> . Acesso em 18 de agosto de 2022.

VLADILAVA, M. **Factors Influencing Consumer Behavior in Gaming Industry: Evidence From Russian Market.** Disponível em: https://dspace.spbu.ru/bitstream/11701/29909/1/MT_Mashkovskaia_IB.pdf . Acesso em 02 de setembro de 2022.

WANG, Meng-Meng; WANG, Jian-Jun. **Understanding Solvers' Continuance Intention in Crowdsourcing Contest Platform:** an extension of expectation-confirmation model. Journal Of Theoretical And Applied Electronic

Commerce Research, [S.L.], v. 14, n. 3, p. 17-33, set. 2019. MDPI AG. Disponível em: <http://dx.doi.org/10.4067/s0718-18762019000300103> .

Weka. **Weka 3: Machine Learning Software in Java**. 2023. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>