



ANÁLISE EXPLORATÓRIA DO MERCADO DE JOGOS DIGITAIS VIA RASPAGEM DE DADOS DA PLATAFORMA STEAM

Francisco P. CUVICE¹; Vinícius A. SILVA²

RESUMO

Este estudo em andamento investiga o mercado de jogos digitais na plataforma Steam por meio da análise de dados coletados via Web Scraping. Foram obtidas informações de 91.311 jogos com o uso de scripts Python (Playwright) e da API da Steam. Após limpeza e filtragem, o conjunto de dados foi reduzido para 15.172 jogos. A análise exploratória inicial identificou padrões relacionados ao sucesso ou à insatisfação dos usuários, revelando que a chave para um jogo bem-sucedido está na entrega de experiências de qualidade voltadas a nichos específicos.

Palavras-chave:

Avaliações de Jogos; Web Scraping; Desenvolvimento de Jogos; Análise de Mercado; Tags de Jogos.

1. INTRODUÇÃO

Nesta nova era digital, jogos complexos de diferentes gêneros estão disponíveis, criando uma grande variedade. Atualmente, os usuários podem acessar jogos que combinam os gêneros de ação e aventura com um cenário aberto e livre. Ademais, a possibilidade de jogar online abriu portas e criou novas fronteiras, permitindo compartilhar a experiência com até mesmo milhares de jogadores ao mesmo tempo (Ponce-Blandón *et al.*, 2020).

No contexto brasileiro de jogos digitais, estima-se que o comércio relacionado a jogos digitais movimentou mais de 2,3 bilhões de dólares em 2021 e que quase 75% da população joga com frequência. Em relação à quantidade de estúdios no país, em 2022 foram identificados 1.009 estúdios (Fortim *et al.*, 2022). Estima-se que a quantidade de estúdios tenha crescido ainda mais com o crescimento do mercado de jogos e do acesso à tecnologia.

Desenvolver jogos de sucesso tornou-se mais difícil devido à grande escala da indústria e à exigência dos jogadores, cujas preferências evoluem com o tempo (Lin *et al.*, 2019). Compreender essas preferências é essencial para que os desenvolvedores alcancem sucesso e se destaquem no mercado de jogos digitais.

A Steam, lançada pela Valve em 2003, é uma plataforma digital com mais de 100.000 jogos, cada um com página própria contendo informações, imagens, vídeos e avaliações dos usuários, que indicam se recomendam ou não o jogo, mostrando a porcentagem de recomendação de forma numérica e informal (Valve Corporation, 2025).

Como uma das maiores plataformas digitais de distribuição de jogos do mundo, a Steam representa uma fonte riquíssima de dados para os desenvolvedores que desejam entender melhor o mercado atual e alinhar seus produtos às expectativas dos jogadores. A coleta e análise de dados de

¹Discente do Bacharelado em Ciência da Computação, IFSULDEMINAS — *Campus* Passos. Email: francisco.cuviced@alunos.ifsuldeminas.edu.br

²Orientador, IFSULDEMINAS — *Campus* Passos. Email: vinicius.silva@ifsuldeminas.edu.br

avaliações e gêneros aclamados possibilita um melhor entendimento das tendências emergentes no mercado e dos elementos que contribuem para a aceitação ou rejeição de um jogo, permitindo decisões mais embasadas durante o processo de concepção, design e lançamento do produto.

Apesar de possuir uma API própria para acessar dados da plataforma, nem todas as informações da loja são diretamente acessíveis pela API oficial. Para coletar uma grande quantidade de dados de páginas online, é comum utilizar a técnica de *Web Scraping* (Raspagem de Dados) para extrair informações estruturadas de sites por meio de scripts automatizados.

Este trabalho propõe utilizar da técnica de raspagem de dados para coletar informações de jogos e suas avaliações na plataforma Steam, com o objetivo de analisar e gerar insights que possam auxiliar no desenvolvimento de jogos, identificando e destacando padrões em jogos bem avaliados e mal avaliados.

2. MATERIAL E MÉTODOS

O projeto está em andamento; as etapas de coleta e pré-processamento de dados foram concluídas, fornecendo a base para as análises iniciais, cujas metodologias são detalhadas a seguir.

Foram desenvolvidos dois scripts na linguagem Python utilizando a biblioteca Playwright para raspar dados da plataforma Steam, em intervalos que respeitam a política da plataforma e não impactam negativamente a experiência de outros usuários. O primeiro script coletou título, data de lançamento, identificador e link da página de cada jogo presente na plataforma, os dados foram armazenados num arquivo Comma Separated Values (CSV), formando um dataset com todos os 134.082 jogos presentes na Steam em 16/06/2025.

A partir dos links contidos no CSV, o segundo script raspou informações como número de análises totais, classificação dessas análises, tags, features, preço fora de promoções, quantidade de conteúdos adicionais (DLCs) e conteúdos de classificação indicativa de jogos lançados a partir de 2015. Foram ignorados aproximadamente 6 mil jogos que requisitam login de usuário na plataforma, em sua maioria jogos com conteúdo adulto explícito. Especificamente, o número de avaliações totais de cada jogo foi obtido por meio de requisições feitas à própria API disponibilizada pela Steam. Os dados raspados, junto aos dados do outro CSV, foram armazenados num novo arquivo CSV, formando um dataset com informações sobre 91.311 jogos.

Após a limpeza, foram mantidos apenas os jogos que possuem ao menos 200 avaliações, representando um conjunto de jogos que tiveram alguma relevância na plataforma, resultando num total de 15.172 jogos, os quais foram utilizados para identificar padrões de mercado.

Para as próximas etapas do projeto, pretende-se aprofundar a análise exploratória dos dados, com a aplicação do algoritmo K-means para identificação de clusters e o algoritmo Apriori para mineração de regras de associação. Espera-se que essas análises revelem padrões significativos e

tendências de mercado no universo dos jogos da Steam.

3. RESULTADOS E DISCUSSÃO

Foram analisadas as tags mais populares entre os jogos de avaliações classificadas como extremamente positivas e jogos de avaliações classificadas como negativas, como mostrado na Figura 1.

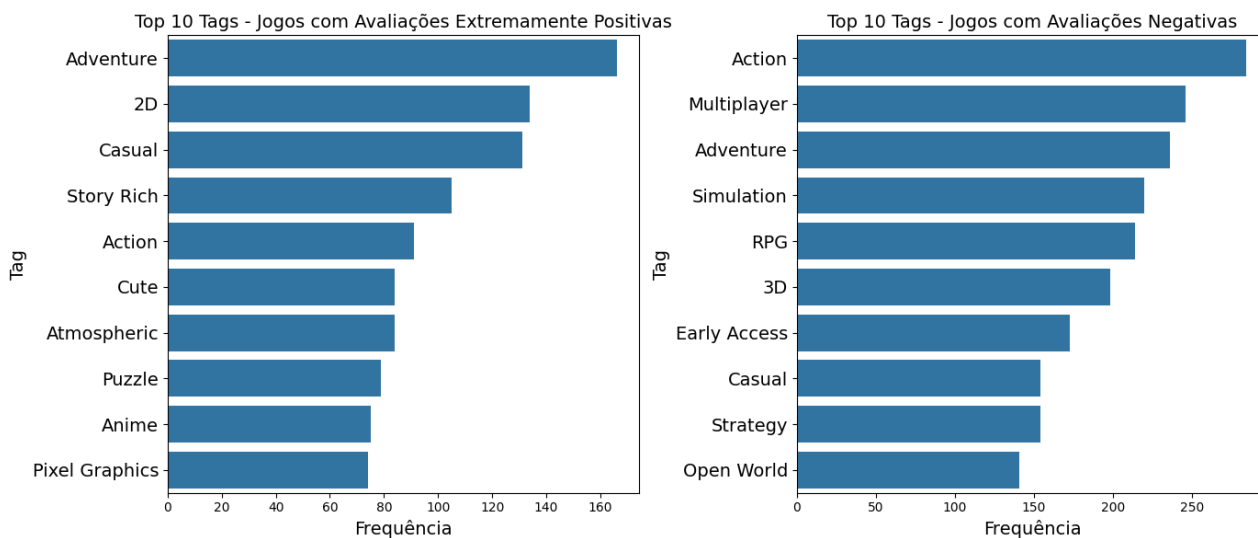


Figura 1: Popularidade de tags em jogos de avaliações negativas e extremamente positivas.

A partir dos gráficos apresentados na Figura 1, podemos visualizar as tags que são comuns entre jogos criticados e aclamados, possivelmente indicando padrões de preferência entre os jogadores. As tags “Action”, “Adventure” e “Casual” estão presentes nos dois gráficos, mostrando que, embora sejam gêneros muito populares, não são indicadores fiéis de bons jogos. Por serem classificações muito abrangentes, acabam incluindo jogos de estilos e níveis de qualidade variados, o que reduz sua utilidade como indicadores confiáveis de sucesso ou fracasso.

As tags “2D” e “Pixel Graphics” presentes no gráfico de jogos com avaliações extremamente positivas indica que o estilo visual dos jogos não necessariamente necessita buscar o realismo, ainda há uma grande aprovação por estilos visuais mais simples e retrô.

Características como “Story Rich”, “Cute”, “Atmospheric” e “Puzzle” presentes no gráfico de jogos com avaliações extremamente positivas indicam uma preferência por experiências mais focadas em narrativas. Quando bem executados, esses elementos tendem a criar uma conexão emocional mais profunda com os jogadores que os buscam, levando a avaliações mais positivas.

A presença de “Early Access” (Acesso Antecipado) na lista de tags negativas é um ponto crucial, sugerindo que muitos jogos lançados em acesso antecipado não conseguem atender às expectativas dos jogadores, seja por falta de conteúdo, bugs persistentes, promessas não cumpridas ou desenvolvimento lento.

Ao comparar os dois gráficos, percebe-se que as tags exclusivas do gráfico positivo (“2D”,

“Story Rich”, “Cute”, “Atmospheric”, “Puzzle”, “Anime”, “Pixel Graphics”) tendem a ser nichos ou características mais específicas, que, quando bem executadas, agradam muito a seus públicos. As tags que aparecem no gráfico de negativos, em contraste, são mais amplas e representam gêneros com grande apelo de massa e, conseqüentemente, com maior potencial para decepção.

4. CONCLUSÃO

Este estudo empregou a técnica de raspagem de dados para realizar uma análise exploratória do vasto mercado de jogos digitais disponíveis na plataforma Steam, utilizando um dataset robusto de mais de quinze mil jogos. A investigação demonstrou que gêneros populares e mais gerais aparecem tanto em jogos bem avaliados quanto em mal avaliados, não sendo indicadores confiáveis de qualidade. As taxas de críticas negativas são particularmente elevadas em jogos classificados como “Early Access”, sugerindo um desalinhamento entre as expectativas e a entrega desses títulos. Em contrapartida, características como “2D”, “Story Rich”, “Cute”, “Atmospheric”, “Puzzle”, “Anime” e “Pixel Graphics” mostraram-se predominantemente presentes em jogos com avaliações extremamente positivas. Isso indica que estilos visuais e focos em narrativa, atmosfera e desafios específicos, quando bem executados, ressoam fortemente com nichos de público, gerando alta satisfação e engajamento.

Futuros passos incluem aprofundar a análise exploratória dos dados aplicando os algoritmos K-means para identificar clusters de jogos e Apriori para mineração de regras de associação. Os resultados deste estudo são particularmente úteis para desenvolvedores de jogos, permitindo que tomem decisões mais embasadas durante o processo de concepção, design e lançamento de seus produtos. Pretende-se apresentar os resultados finais em eventos acadêmicos.

REFERÊNCIAS

FORTIM, Ivelise et al. Pesquisa Indústria Brasileira de Games 2022. In: **TIC CULTURA: pesquisa sobre o uso das tecnologias de informação e comunicação nos equipamentos culturais brasileiros 2022**. São Paulo: Comitê Gestor da Internet no Brasil, 2023. p. 113-126. Disponível em: https://cgi.br/media/docs/publicacoes/2/20230621154638/tic_cultura_2022_livro_eletronico.pdf. Acesso em: 5 mai. 2025.

LIN, D. et al. An empirical study of game reviews on the Steam platform. **Empirical Software Engineering**, v. 24, p. 170–207, 2019.

PONCE-BLANDÓN, José Antonio et al. Videogame-related experiences among regular adolescent gamers. **Plos One**, San Francisco, CA, USA, v. 15, n. 7, p. e0235327, 2020.

VALVE CORPORATION. **Sobre a Steam**. 2025. Disponível em: <https://store.steampowered.com/about/?l=brazilian>. Acesso em: 28 abr. 2025.