

**INSTITUTO
FEDERAL**

Sul de Minas Gerais

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO SUL DE
MINAS GERAIS – CAMPUS PASSOS**

Francisco Paulino Cuvace

**Análise do mercado de jogos digitais via raspagem de dados da plataforma
Steam, visando auxiliar desenvolvedores na compreensão e inserção no
mercado**

PASSOS, MG

2025

Francisco Paulino Cuvace

Análise do mercado de jogos digitais via raspagem de dados da plataforma Steam, visando auxiliar desenvolvedores na compreensão e inserção no mercado.

Trabalho de Conclusão de Curso apresentado ao IFSULDEMINAS como parte das exigências do curso de graduação em Bacharelado em Ciência da Computação para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Vinícius Alves Silva

PASSOS, MG

2025



Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais
IFSULDEMINAS - Campus Passos

FOAP Nº20/2025/CBCC/CURSTECSUP/CGE/DDE/DG/PAS/IFSULDEMINAS

FOLHA DE APROVAÇÃO

FRANCISCO PAULINO CUVICE

ANÁLISE DO MERCADO DE JOGOS DIGITAIS VIA RASPAGEM DE DADOS DA PLATAFORMA STEAM, VISANDO AUXILIAR DESENVOLVEDORES NA COMPREENSÃO E INSERÇÃO NO MERCADO

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Passos, como parte das exigências necessárias para a conclusão do Curso de Graduação em Bacharelado em Ciência da Computação.

Orientador: Prof. Dr. Vinícius Alves Silva

Aprovado em: 01/12/2025

COMISSÃO EXAMINADORA:

Vinícius Alves Silva - Professor Orientador

Alessandro de Castro Borges - Membro examinador

Wedson Gomes da Silveira Júnior - Membro examinador

Documento assinado digitalmente pela Comissão Examinadora.

Documento assinado eletronicamente por:

- **Vinicius Alves Silva**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 01/12/2025 15:11:42.
- **Alessandro de Castro Borges**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 01/12/2025 15:13:36.
- **Wedson Gomes da Silveira Junior**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 01/12/2025 16:49:55.

Este documento foi emitido pelo SUAP em 01/12/2025. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsuldeminas.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 613375

Código de Autenticação: 75cc74ce64



Documento eletrônico gerado pelo SUAP (<https://suap.ifsuldeminas.edu.br>)
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais

AGRADECIMENTOS

Agradeço à minha família por me dar a oportunidade de estudar e desenvolver o presente trabalho.

Agradeço também à minha namorada por todo o apoio emocional durante o desenvolvimento.

Agradeço, ainda, aos meus amigos, por tornarem a jornada descontraída e alegre, principalmente ao Guilherme por surgir com a ideia do tema do trabalho e apoiar até o fim com comentários e sugestões.

RESUMO

A grande variedade de jogos complexos nesta era digital elevou as exigências dos jogadores, tornando o desenvolvimento de jogos de sucesso um desafio maior. Para alcançar o êxito e se destacar, os desenvolvedores precisam compreender profundamente as preferências desse mercado. O presente trabalho propôs o uso da técnica de Raspagem de Dados com Python e a biblioteca Playwright para coletar e analisar dados do mercado de jogos digitais na Steam. A metodologia envolveu a coleta e a Análise Exploratória de Dados (AED) de uma amostra de 15.172 jogos lançados a partir de 2015 que possuíam, no mínimo, 200 avaliações na plataforma. Os resultados da AED permitiram identificar padrões e tendências de mercado relevantes. A funcionalidade “Um Jogador” mostrou-se dominante, presente em mais de 90% dos jogos. Verificou-se que a tag “Indie” é a mais popular no catálogo, e a análise das tags mais frequentes sugere que, embora o multiplayer competitivo gere alto engajamento, as avaliações extremamente positivas estão mais associadas a experiências simples, bem executadas e com forte componente narrativo. Além disso, jogos com faixas de preço mais elevadas apresentaram um volume mediano de avaliações superior, ou seja, metade dos jogos de preço elevado receberam mais avaliações do que metade dos jogos de preço baixo, indicando maior engajamento da comunidade para produções que custam mais. Com base na sazonalidade de lançamentos, uma recomendação estratégica foi elaborada: o desenvolvedor deve cronometrar o lançamento com desconto para uma semana antes de uma grande promoção sazonal da Steam, aproveitando o pico de usuários gerado pelo evento e maximizando a exposição. Em conclusão, a qualidade e a inovação na execução do produto são os fatores mais decisivos para a recepção crítica positiva na Steam.

Palavras-chave: Raspagem de Dados. Web Scraping. Análise de Mercado. Jogos Digitais. Steam.

ABSTRACT

The wide variety of complex games in this digital age has raised the demands on gamers, making the development of successful games a greater challenge. To achieve success and stand out, developers need a deep understanding of the preferences of this market. This work proposed the use of the Data Scraping technique with Python and the Playwright library to collect and analyze data from the digital games market on Steam. The methodology involved the collection and Exploratory Data Analysis (EDA) of a sample of 15,172 games released since 2015 that had at least 200 reviews on the platform. The results of the EDA allowed the identification of relevant market patterns and trends. The "Single Player" functionality proved to be dominant, present in more than 90% of the games. It was found that the "Indie" tag is the most popular in the catalog, and the analysis of the most frequent tags suggests that, although competitive multiplayer generates high engagement, extremely positive reviews are more associated with simple, well-executed experiences with a strong narrative component. Furthermore, games with higher price ranges showed a higher median volume of reviews; that is, half of the high-priced games received more reviews than half of the low-priced games, indicating greater community engagement for productions that cost more. Based on the seasonality of releases, a strategic recommendation was developed: the developer should time the discounted release to one week before a major seasonal Steam sale, taking advantage of the user peak generated by the event and maximizing exposure. In conclusion, quality and innovation in product execution are the most decisive factors for positive critical reception on Steam.

Keywords: Data scraping. Web scraping. Market analysis. Digital games. Steam.

LISTA DE FIGURAS

Figura 1: Processo seguido pelo script que coleta informações básicas de cada jogo.....	19
Figura 2: Processo Seguido Pelo Segundo Script que Coleta Informações Específicas de Cada um dos Jogos Lançados a partir de 2015.....	20
Figura 3: Fluxo de Execução do Trabalho.....	21
Figura 4: As 15 Combinações de Tags com Maior Média de Avaliações.....	23
Figura 5: Frequência de Conteúdos Relacionados à Classificação Indicativa.....	24
Figura 6: As 20 Tags com Maior Média de Avaliações por Jogo.....	25
Figura 7: Presença de Funcionalidades.....	26
Figura 8: As 20 Tags Mais Comuns.....	27
Figura 9: Distribuição das Principais Tags por Ano de Lançamento.....	28
Figura 10: As 10 Tags Mais Populares em Jogos de Avaliação Extremamente Positiva e em Jogos de Avaliação Negativa.....	29
Figura 11: Distribuição do Lançamento de Jogos de 2015 a 2025 por Mês e Faixa de Preço.....	30
Figura 12: Distribuição do Lançamento de Jogos Indie de 2015 a 2025 por Mês e Faixa de Preço.....	31
Figura 13: Comparação entre jogos pagos e gratuitos.....	32
Figura 14: Engajamento por faixa de preço.....	33

LISTA DE ABREVIATURAS E SIGLAS

AED	Análise Exploratória de Dados
API	<i>Application Programming Interface</i>
COVID-19	<i>Coronavirus Disease 2019</i>
CSV	<i>Comma-Separated Values</i>
DLC	<i>Downloadable Content</i>
FPS	<i>First-Person Shooter</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ID	Identificador
LGBM	<i>Light Gradient Boosting Machine</i>
Link	<i>Hyperlink</i>
PSF	Python Software Foundation
RMSE	Raiz do Erro Quadrático Médio
URL	<i>Uniform Resource Locator</i>

SUMÁRIO

1 INTRODUÇÃO.....	. 11
1.1 Objetivo Geral.....	. 12
1.2 Objetivos Específicos.....	. 12
2 REFERENCIAL TEÓRICO.....	. 13
2.1 Linguagem Python.....	. 13
2.2 Raspagem de Dados da Web.....	. 13
2.2.1 Biblioteca Playwright.....	. 14
2.3 Processamento de Dados.....	. 14
2.3.1 Biblioteca Pandas.....	. 15
2.4 Análise Exploratória de Dados.....	. 15
2.4.1 Visualização de Dados.....	. 15
3 LITERATURA CORRELATA.....	. 17
4 MATERIAL E MÉTODOS.....	. 19
5 RESULTADOS E DISCUSSÕES.....	. 23
6 CONCLUSÃO.....	. 34
REFERÊNCIAS.....	. 36

1 INTRODUÇÃO

Nesta nova era digital, jogos complexos de diferentes gêneros estão disponíveis, criando uma grande variedade de opções no mercado. 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, uma pesquisa feita por Cardoso (2023) revela que a receita estimada da indústria de desenvolvimento de jogos foi de 251,6 milhões em 2022 e cerca de 82,1% dos brasileiros afirmam que jogar jogos digitais é uma de suas principais formas de diversão. A mesma pesquisa afirma que em 2023 foram identificados 1.042 estúdios ativos que desenvolvem jogos digitais no Brasil.

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, 2025b).

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 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 a técnica de raspagem de dados para coletar informações de jogos na plataforma Steam, com o objetivo de analisar e gerar *insights* que possam auxiliar no desenvolvimento de jogos, identificando padrões em jogos bem e mal avaliados.

1.1 Objetivo Geral

Utilizar a biblioteca Playwright para executar a raspagem de dados da plataforma Steam e analisar as informações obtidas para identificar tendências de mercado úteis para o desenvolvimento de jogos.

1.2 Objetivos Específicos

- Executar a raspagem de dados de jogos lançados nos últimos 10 anos na plataforma Steam;
- Estruturar e tratar os dados obtidos;
- Analisar os dados, identificando padrões e tendências;
- Obter *insights* que ajudem a compreender a preferência por determinados gêneros;
- Elaborar recomendações que auxiliem desenvolvedores na criação e publicação de seus jogos.

2 REFERENCIAL TEÓRICO

Neste capítulo, são abordados os conceitos teóricos fundamentais para o entendimento e desenvolvimento deste trabalho. A seção 2.1 decorre sobre a linguagem Python e suas características. Na seção 2.2 é explicado o conceito de Raspagem de Dados da Web e é apresentada a biblioteca Playwright usada para a raspagem. Na seção 2.3 é discutido o conceito de Processamento de Dados e é apresentada a biblioteca Pandas. Por fim, a seção 2.4 fala sobre Análise Exploratória de Dados, a importância da Visualização de Dados e introduz as bibliotecas Matplotlib e Seaborn.

2.1 Linguagem Python

Python é uma linguagem de programação interpretada de alto nível mantida pela Python Software Foundation (PSF). Foi projetada para ter uma sintaxe que favorece a legibilidade do código e que permita aos desenvolvedores expressar conceitos com menos linhas de código do que em outras linguagens, como C++ ou Java. De acordo com a PSF (2025), Python é uma linguagem fácil de aprender, altamente legível, multiparadigma, extensível, embutível e portátil. Possui uma ampla biblioteca padrão que inclui módulos para manipulação de arquivos, expressões regulares, redes, interfaces gráficas, web, entre outros. Também possui uma comunidade global muito ativa e acolhedora.

A linguagem é altamente versátil, podendo ser usada em desenvolvimento web, automação de tarefas, análise de dados e ciência de dados, inteligência artificial e machine learning, desenvolvimento de aplicações desktop e jogos, scripts para administração de sistemas, educação e ensino de programação.

O Python Package Index é o repositório oficial de pacotes de software da linguagem Python, recurso fundamental para a comunidade permitindo que desenvolvedores publiquem pacotes, compartilhem soluções, bibliotecas e ferramentas, e distribuam software de código aberto feito em Python (PSF, 2025).

A escolha da linguagem se deve pela sua facilidade de uso e popularidade, sendo uma das linguagens de programação mais populares desde seu lançamento em 1991 (McKinney, 2022). Sua vasta lista de bibliotecas para manipulação de dados também foi um fator decisivo.

2.2 Raspagem de Dados da Web

Raspagem de dados web é uma técnica utilizada para extrair dados da World Wide Web e salvá-los num sistema de arquivos ou banco de dados para análise futura, um processo que pode ser realizado manualmente, porém nos dias atuais é praticamente um sinônimo com

o uso de *bots* e *scripts* para coleta automatizada de dados devido a escala do volume de dados que as indústrias necessitam capturar e analisar (Moura *et al.*, 2024).

A raspagem requer um conhecimento profundo da estrutura e do conteúdo de um site para que os dados sejam coletados com precisão. Esse processo pode ser demorado e exigir conhecimento técnico especializado. Além disso, os dados extraídos do site podem não estar estruturados ou em um formato difícil de interpretar, o que significa que limpeza e formatação adicionais podem ser necessárias para torná-los úteis (Bansal; DAR; Bhat, 2023).

É importante utilizar técnicas de raspagem de dados de forma responsável e compreender que o uso de *bots* para realizar requisições a um domínio pode impactar negativamente a experiência de outros usuários e sobrecarregar a infraestrutura do site (Moura *et al.*, 2024).

2.2.1 Biblioteca Playwright

Playwright é uma biblioteca de código aberto desenvolvida em 2020 pela Microsoft especificamente para testes de ponta a ponta de aplicativos e serviços da web. De acordo com a Microsoft (2025), a biblioteca oferece suporte a todos os mecanismos de renderização modernos, incluindo Chromium, WebKit e Firefox, funcionando nas plataformas Windows, Linux e macOS, e também permite testar interfaces web suportando interações reais de usuário, como cliques, preenchimento de formulários, navegação, carregamento de páginas e muito mais. Com APIs projetadas para serem consistentes e resilientes, o Playwright cuida automaticamente de esperas inteligentes, sincronizando com eventos da interface para tornar os testes mais estáveis.

Apesar de ser uma ferramenta relativamente recente e pouco utilizada em trabalhos acadêmicos, a ferramenta foi escolhida devido sua robustez e suporte a múltiplos navegadores e APIs modernas.

2.3 Processamento de Dados

Ao falar de dados, imagina-se grandes conjuntos de dados com um grande número de linhas e colunas, como em um grande banco de dados, o que nem sempre é o caso. Os dados podem vir em diversos formatos, tabelas estruturadas, texto não estruturado contendo arquivos de imagens, áudio e vídeo em diversos formatos. Como as máquinas podem interpretar apenas 1s e 0s, os dados precisam ser transformados ou codificados para que a máquina possa analisá-los facilmente (Gupta; Bagchi, 2024).

O pré-processamento de dados é uma técnica usada para converter dados brutos em um conjunto de dados limpo. Inicialmente, os dados são coletados em formato bruto, que pode não ser adequado para análise. Dados do mundo real são frequentemente incompletos, inconsistentes e/ou carecem de certos detalhes que revelem comportamentos ou tendências. Além disso, é provável que contenham muitos erros. O pré-processamento de dados é um método comprovado para resolver esses problemas (Gupta; Bagchi, 2024).

2.3.1 Biblioteca Pandas

Pandas é definida como uma biblioteca de código aberto que fornece capacidade de processamento de dados de alto desempenho, métricas de desempenho aprimoradas, estruturas de dados fáceis de usar e pacotes de análise de dados, ferramentas e bibliotecas para a linguagem de programação Python (Sial; Rashdi; Khan, 2021). É uma das bibliotecas Python mais utilizadas em ciência e análise de dados, fornecendo inúmeras funções e métodos que agilizam as etapas de análise e pré-processamento de dados (Gupta; Bagchi, 2024).

O uso de Pandas com Python abrange vários campos de especialização, incluindo ciências de dados, modelagem computacional, finanças, economia, análise estatística e aprendizado de máquina (Sial; Rashdi; Khan, 2021).

Por ser amplamente utilizada e aclamada por suas funcionalidades, a biblioteca foi escolhida para tratar os dados raspados pela Playwright.

2.4 Análise Exploratória de Dados

A Análise Exploratória de Dados (AED) é uma investigação preliminar de dados para descobrir padrões, detectar anomalias, testar e validar hipóteses com a ajuda de estatísticas descritivas e representações gráficas (Neto *et al.*, 2022).

2.4.1 Visualização de Dados

Visualização de Dados é a ilustração gráfica para representar dados com o uso integrado de design ilustrado, com a única perspectiva de fornecer uma forma visualizada que seja mais fácil de entender e apresentar (Sial; Rashdi; Khan, 2021).

Matplotlib é uma das bibliotecas de visualização de dados mais utilizadas em Python. Construída por John Hunter junto a vários colaboradores, é um aspecto integral da comunidade de ciência de dados da linguagem Python, sendo usada por grande parte dos cientistas e filósofos do mundo, e é facilmente compatível com NumPy, Pandas e outras bibliotecas relevantes (Sial; Rashdi; Khan, 2021).

Seaborn é uma biblioteca de visualização gráfica construída sobre as configurações primárias da Matplotlib. Ela fornece acessibilidade aos usuários com alguns dos processos de visualização de dados mais comuns, com certas necessidades de visualização de dados, como mapeamento de cores para uma variável (Sial; Rashdi; Khan, 2021).

Além da grande popularidade das bibliotecas, ambas oferecem ferramentas robustas para criar gráficos de alta qualidade de forma simples e rápida. Enquanto a Matplotlib oferece controle total sobre o gráfico (tamanho, títulos, eixos, layout), a Seaborn fornece gráficos estatísticos com estética aprimorada e sintaxe mais simples, complementando uma a outra. Ademais, ambas são compatíveis com estruturas de dados da biblioteca Pandas, justificando assim sua escolha.

3 LITERATURA CORRELATA

A coleta de dados relacionados a jogos eletrônicos têm recebido mais destaque recentemente, com vários trabalhos utilizando raspagem e outras técnicas para coletar e analisar dados para os mais diversos propósitos.

Teja, Hanafi e Qomariyah (2023) utilizaram bibliotecas da linguagem Python para raspar dados da Steam e testar vários modelos diferentes de regressão para tentar prever a classificação recebida no site Metacritic por um dos jogos mais vendidos na Steam, utilizando vários datasets dos melhores jogos no metacritic entre 1995 a 2021. As bibliotecas utilizadas foram BeautifulSoup4, Pandas, Requests e Re. Foram testados onze modelos de regressão utilizando a métrica da Raiz do Erro Quadrático Médio (RMSE) para medir a precisão de cada modelo, sendo o melhor aquele que possui o menor RMSE. Dentre eles se destacaram LGBM (RMSE de 0,5227) e Random Forest (0,5991), ambos obtendo os menores RMSE com uma diferença negligível entre os dois, mostrando-se os mais eficientes em prever a classificação de um jogo.

Kainz e Pirker (2023) desenvolveram uma ferramenta de análise e visualização que procura, mostra e analisa informações relacionadas a jogos na Steam. Os dados gerais e o ID dos jogos são coletados usando crawlers (programas que navegam automaticamente pela web para coletar e indexar informações de páginas). As avaliações são coletadas por meio de requisições HTTP GET da Steam Web API baseando-se no ID dos jogos. A ferramenta possibilita realizar análises de dados usando os dados do jogo e, opcionalmente, avaliações armazenadas num banco de dados SQLite. Também oferece suporte a gráficos de linhas, histogramas sobrepostos, gráficos de barras agrupadas e gráficos de barras como tipos de gráficos disponíveis, criados utilizando a biblioteca Matplotlib da linguagem Python. A ferramenta oferece quatro áreas principais de análise, incluindo análise geral, análise baseada em gênero, análise baseada em tags e análise de detalhes do jogo, além de permitir que usuários gerem tarefas personalizadas por meio de consultas SQL Select. Utilizando a ferramenta, foi feito um estudo de caso sobre o aumento no número de avaliações na pandemia da COVID-19, onde foi constatado um aumento de 94% da quantidade de avaliações escritas em 2020 comparado a 2019.

Cunha, Pessa e Mendes (2024) analisaram dados organizados ao longo do tempo de popularidade de 5.840 jogos da Steam ao longo de 11 anos, identificando cinco padrões gerais de comportamento (*decreasing*, *hilly*, *increasing*, *bursty* e *valley*). O estudo mostra que a maioria dos jogos apresenta queda rápida de popularidade, enquanto poucos exibem crescimento sustentado. Embora o foco dos autores seja a dinâmica temporal da atenção dos jogadores, a pesquisa evidencia a importância de fatores que influenciam a visibilidade e o engajamento na

Steam, o que se relaciona diretamente com a análise de mercado realizada neste trabalho, apesar de tratar variáveis distintas como tags, features e janelas de lançamento.

De Luisa *et al.* (2021) conduziu uma pesquisa focada em prever a popularidade de jogos na Steam em seus estágios iniciais após o lançamento, utilizando a contagem mediana de jogadores do primeiro mês como principal preditor. Foram coletados dados de serviços como Steam, SteamSpy e SteamDB para jogos lançados após 2015 e testaram os modelos de inferência Bayesianos normais, normais dobrados e normais dobrados hierárquicos em variantes homo e heteroscedásticas. Variáveis como preço, tamanho do jogo, idiomas suportados, data de lançamento e gêneros foram incluídas para analisar sua influência na popularidade. O modelo que demonstrou o melhor desempenho preditivo foi o Modelo Hierárquico Heteroscedástico baseado no gênero. As conclusões indicaram que o preditor mais forte foi a contagem mediana de jogadores no primeiro mês, servindo como base para previsões subsequentes, e jogos lançados no início do mês e aqueles classificados em gêneros como Esportes, Corrida e Simulação tenderam a correlações mais altas de popularidade. Os resultados reforçam a relevância do uso de dados da Steam para análises preditivas, alinhando-se à proposta deste trabalho ao demonstrar que métricas iniciais de engajamento, aliadas a características do jogo, podem atuar como fortes indicadores de desempenho futuro no mercado.

4 MATERIAL E MÉTODOS

Inicialmente, é importante ressaltar que a raspagem de dados foi feita em intervalos que respeitam a política da plataforma e não impactam negativamente a experiência de outros usuários. Dito isso, foi desenvolvido um *script* na linguagem Python utilizando a biblioteca Playwright para raspar os dados da Steam. O *script* foi utilizado para coletar dados como título do jogo, data de lançamento, ID e a URL que leva diretamente para a página do jogo, para todos os jogos presentes na plataforma em 16 de Junho de 2025. Esses dados foram armazenados num arquivo Comma-Separated Values (CSV), formando um *dataset* com todos os 134.082 jogos presentes na plataforma. O processo descrito é apresentado resumidamente na Figura 1.

Figura 1: Processo seguido pelo script que coleta informações básicas de cada jogo

```
1 INÍCIO
2     Abrir navegador automatizado
3
4     Criar arquivo CSV
5
6     Para cada página de resultados:
7
8         Acessar página e listar jogos
9
10        Para cada jogo coletar:
11            - ID
12            - Título
13            - Preço
14            - Data
15            - Link da página do jogo
16
17        Armazenar no CSV
18    Fechar navegador e encerrar
19 FIM
20
```

Fonte: Autoria própria

Em seguida, foi desenvolvido outro *script*, também utilizando Python e Playwright, que, a partir dos links coletados, raspou informações específicas de cada jogo, sendo elas o número de análises totais que um jogo possui, a classificação dessas análises (muito positivas, negativas ou mistas), marcadores populares de cada jogo (*tags*), features que apresentam (suporte para Multijogador, recurso Remote Play Together, etc.), preço atual sem descontos, quantidade de conteúdos para download (DLCs) e conteúdos de classificação indicativa (Violência, Livre, Temas sensíveis). O processo descrito pode ser observado de forma resumida na Figura 2.

Figura 2: Processo Seguido Pelo Segundo Script que Coleta Informações Específicas de Cada um dos Jogos Lançados a partir de 2015

```
1 INÍCIO
2   Ler arquivo CSV com lista de jogos
3
4   Filtrar apenas jogos lançados a partir de 2015
5
6   Para cada jogo filtrado:
7       Acessar página do jogo na Steam
8       Coletar informações principais:
9           - Título
10          - Tags
11          - Número de avaliações
12          - Preço fora de promoção
13          - Número de DLCs
14          - Conteúdos de classificação indicativa
15
16       Armazenar os dados coletados
17
18       A cada 10 jogos, salvar progresso em arquivo
19
20   Exibir total de jogos processados e ignorados
21 FIM
```

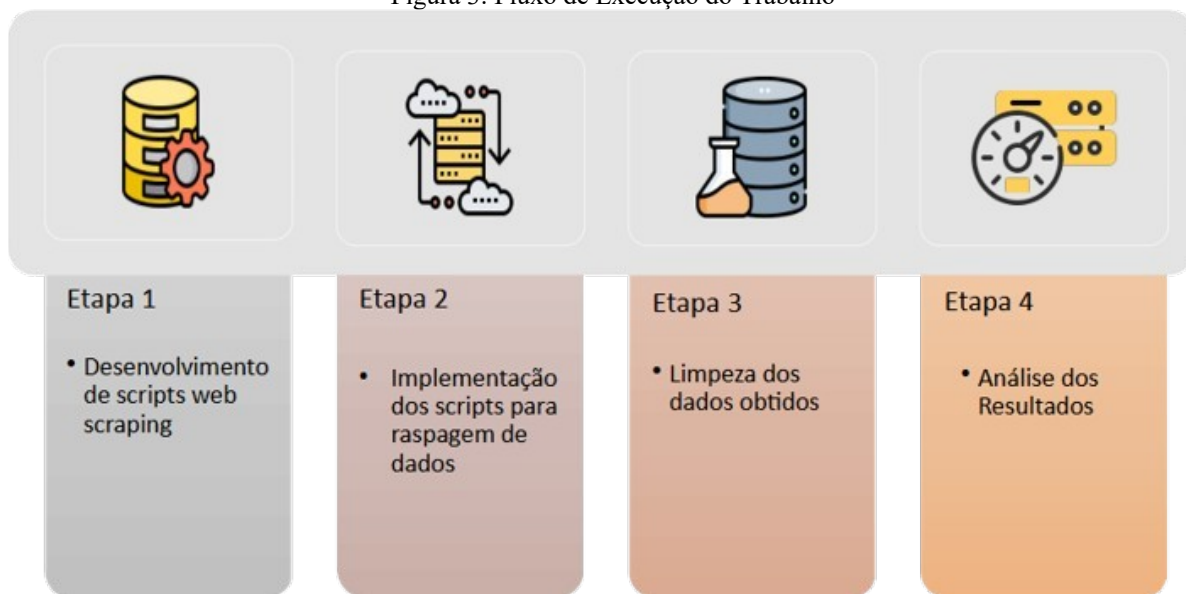
Fonte: Autoria própria

A raspagem de dados foi iniciada no dia 17 de Junho de 2025, e foi finalizada somente no dia 22 de Junho de 2025 devido à implementação de um intervalo de segurança entre requisições. A medida buscou evitar o bloqueio por um volume excessivo de requisições, visto que o Acordo de Assinatura do Steam proíbe ações que possam danificar ou interferir no funcionamento da plataforma (Valve Corporation, 2025a). Foram coletados detalhes apenas de jogos lançados a partir de 2015. Ao tentar raspar o número total de avaliações dos jogos notou-se que, em alguns casos, o carregamento assíncrono do JavaScript da página falhava,

retornando valor nulo. Foi optado, então, por obter a informação por meio de requisições feitas à própria API disponibilizada pela Steam utilizando o ID dos jogos. Durante a coleta, foram ignorados cerca de 6 mil jogos que requisitaram login na plataforma, em sua maioria jogos de conteúdo adulto explícito, impossibilitando a coleta dos dados dos mesmos. Os dados raspados, unidos aos dados do outro CSV, foram armazenados em um novo arquivo CSV, formando outro dataset, desta vez com informações de 91.311 jogos.

Foi feita a limpeza dos dados removendo dados duplicados, tratando valores ausentes, corrigindo inconsistências e convertendo tipos de dados para os desejados. Foi feita também uma filtragem, de forma que foram mantidos apenas jogos com 200 ou mais avaliações totais. Jogos com pouquíssimas avaliações podem gerar resultados extremos (como 100% de avaliações positivas) que não representam a tendência geral. Este recorte funciona como uma tentativa de minimizar o viés, mantendo somente títulos que alcançaram um nível mínimo de visibilidade e engajamento da comunidade. O valor mínimo de 200 avaliações foi escolhido porque existe uma estimativa amplamente citada na comunidade de que apenas 3% a 10% dos jogadores da Steam deixam uma avaliação. Embora 100 avaliações já represente um número razoável, a escolha do mínimo de 200 avaliações totais é uma tentativa de garantir um nível mais alto de credibilidade. Após a filtragem, sobraram 15.172 jogos, os quais foram analisados para identificar padrões de mercado e preferências dos jogadores. Na Figura 3 é possível observar o fluxo das etapas do trabalho.

Figura 3: Fluxo de Execução do Trabalho



Fonte: Autoria própria

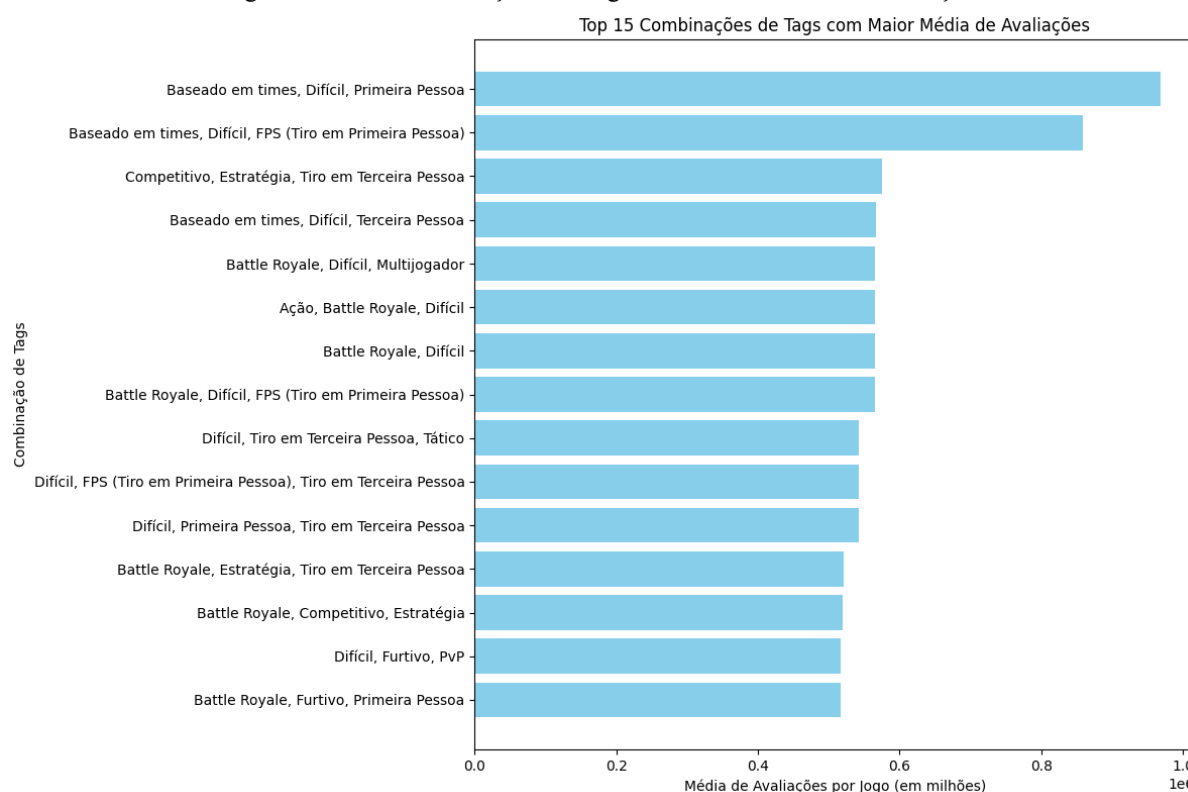
Foram conduzidas várias tentativas de agrupamento dos dados utilizando algoritmos como K-means e DBScan, no entanto, os dados se mostraram muito esparsos. Mesmo usando

apenas preço, número de avaliações totais e tags para agrupar, filtrando apenas por tags que mais aparecem, usando one-hot encoding e Análise de Componente Principal, a quantidade de ruído é muito alta, a maior parte do dataset não forma agrupamentos densos, inviabilizando o agrupamento efetivo e a possível mineração de regras de associação posterior.

5 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados e discutidos os principais resultados obtidos a partir da análise dos dados coletados na plataforma Steam. Os gráficos foram organizados de forma a evidenciar as características mais relevantes dos jogos em termos de tags, combinações de características, classificações indicativas, funcionalidades, preço e engajamento, permitindo identificar padrões e tendências significativas do mercado. A Figura 4 apresenta as 15 combinações de tags com maior média de avaliações por jogo.

Figura 4: As 15 Combinações de Tags com Maior Média de Avaliações



Fonte: Autoria própria

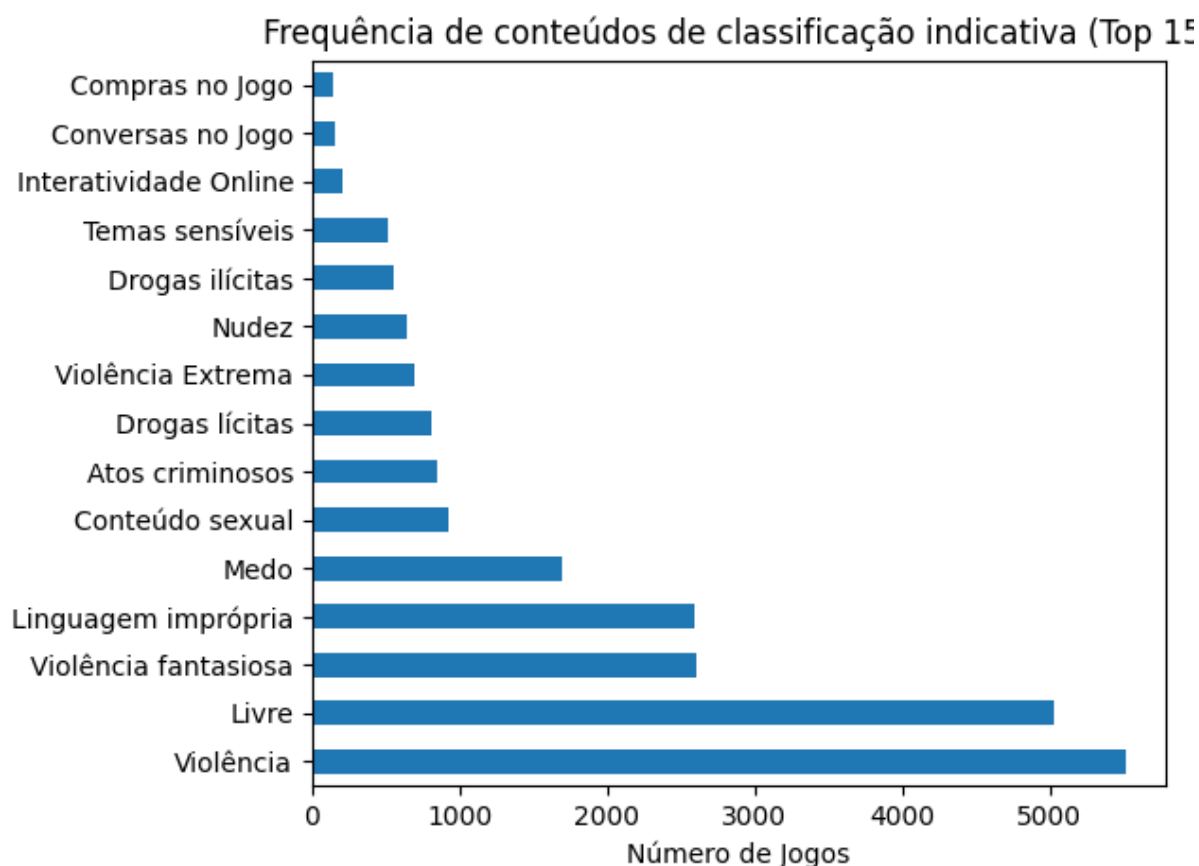
Nota-se que os conjuntos relacionados a dificuldade, perspectiva em primeira pessoa e jogabilidade em equipe se destacam, alcançando médias próximas ou superiores a 900 mil avaliações. Além disso, observa-se a recorrência de gêneros de jogos de tiro em diferentes combinações, incluindo “*Battle Royale*”, subgênero que foca na sobrevivência do último jogador em um mapa que se contrai progressivamente, “Tiro em Terceira Pessoa”, que apresenta a câmera posicionada atrás ou acima do personagem, e “FPS (Tiro em Primeira Pessoa)”, caracterizado pela perspectiva que simula os olhos do personagem.

Esse resultado evidencia a popularidade de jogos que proporcionam desafios competitivos e experiências cooperativas intensas, o que pode explicar o elevado número de avaliações. O destaque de “*Battle Royale*” confirma sua consolidação como um dos

subgêneros mais atrativos da última década, associado ao sucesso de títulos como Fortnite e PlayerUnknown's Battlegrounds.

A Figura 5 apresenta os conteúdos mais frequentes relacionados à classificação indicativa dos jogos.

Figura 5: Frequência de Conteúdos Relacionados à Classificação Indicativa



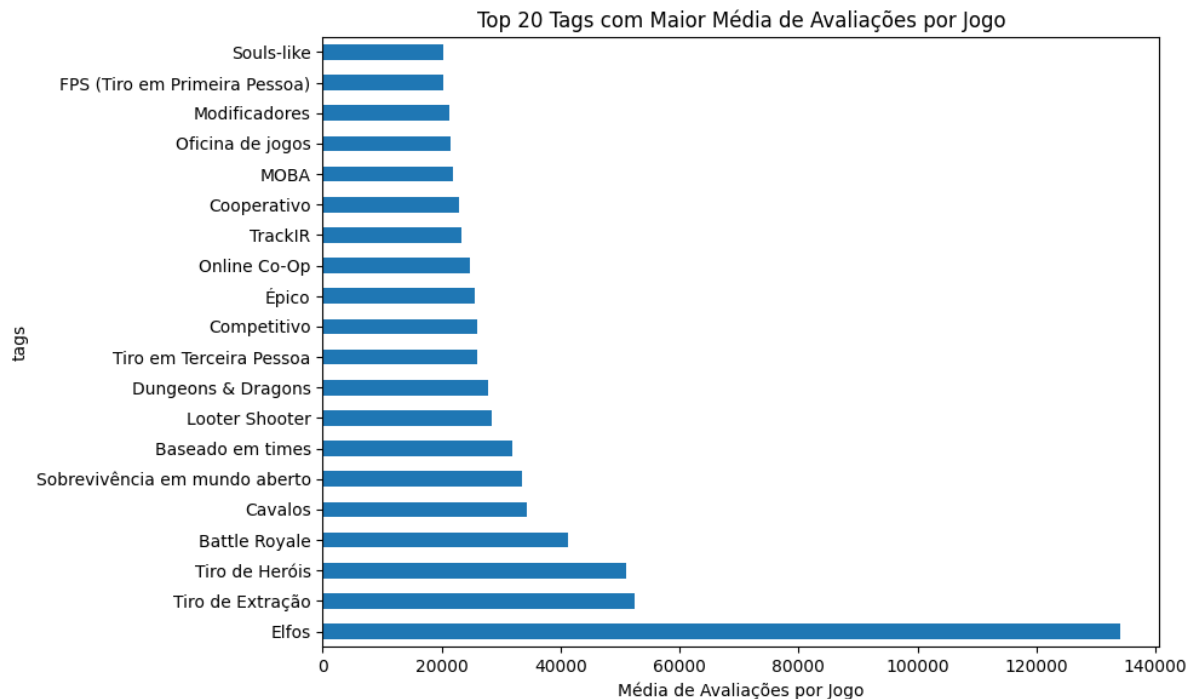
Fonte: Autoria própria

Os fatores mais recorrentes são “Violência” e a categoria “Livre”, que juntos superam 10 mil ocorrências. Outros elementos, como “Linguagem imprópria” e “Violência fantasiosa”, também aparecem com destaque, enquanto conteúdos mais específicos, como “Compras no jogo” (compras dentro do jogo com dinheiro real) e “Conversas no jogo” (comunicação dentro do jogo por texto ou voz), são menos frequentes.

Esse resultado sugere que a violência permanece como o principal tema regulatório do setor, refletindo a forte presença de combates nos jogos. A grande quantidade de jogos classificados como “Livres” também mostra a diversidade do catálogo da Steam, que abriga desde títulos casuais até produções voltadas a públicos mais velhos.

Na Figura 6, observam-se as 20 tags com maior média de avaliações por jogo.

Figura 6: As 20 Tags com Maior Média de Avaliações por Jogo



Fonte: Autoria própria

Entre as principais, destacam-se “Elfos” (personagens clássicos em jogos de fantasia), “Tiro de Extração” (subgênero de jogos de tiro que foca na coleta de recursos em um mapa perigoso, onde o objetivo é sair do mapa sem ser eliminado) e “Tiro de Heróis” (subgênero de jogos de tiro que se destaca por ter personagens pré-definidos, cada um com um conjunto único de habilidades especiais e função tática em equipe), com médias superiores a 50 mil avaliações por título. Também aparecem tags já consolidadas no mercado, como “*Battle Royale*”, “Sobrevivência em mundo aberto” (gênero que combina exploração de um mapa vasto, gestão de necessidades básicas e coleta de recursos para construção de itens e estruturas) e “Baseado em times” (conceito que se aplica a diversos gêneros, onde o sucesso depende essencialmente da colaboração e coordenação entre os jogadores da mesma equipe).

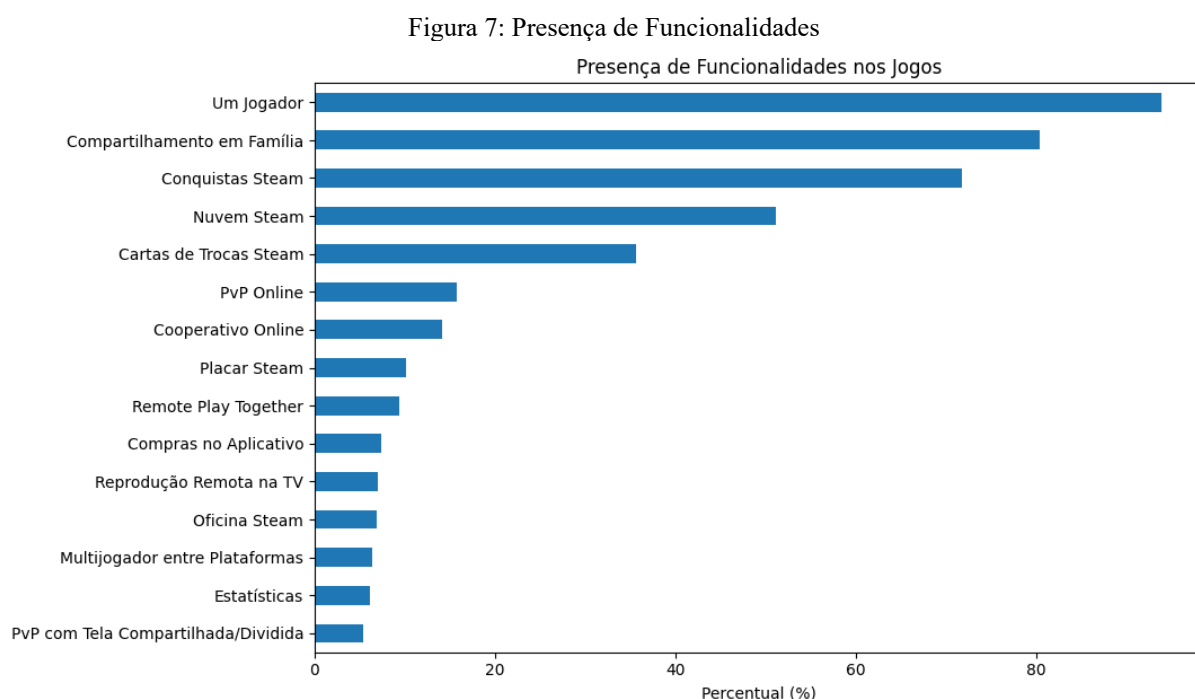
O destaque de tags menos convencionais, como “Elfos” ou “Cavalos” (títulos onde a presença de cavalos é essencial para a exploração ou mecânica central do jogo), pode estar relacionado a casos específicos de jogos muito populares que carregam essas classificações. Já os gêneros competitivos “Tiro de Heróis” e “*Battle Royale*” reforçam a tendência de maior engajamento em jogos com forte apelo *multiplayer* e competitivo.

Nota-se que algumas das tags com maior média de avaliações por jogo, como “Tiro de Heróis”, “Tiro de Extração” e “*Battle Royale*” coincidem com as combinações mais populares da Figura 4, reforçando o vínculo entre competitividade e engajamento. Esse alinhamento

sugere que os jogos mais comentados compartilham não apenas o gênero, mas também dinâmicas sociais e desafios que estimulam a interação entre jogadores.

Além disso, a presença de tags como “Mundo aberto” (exploração de um mapa vasto) e “Sobrevivência” (gestão de necessidades básicas) estabelece uma ponte com a Figura 9, em que o aumento dessas categorias nos últimos anos evidencia uma tendência de mercado voltada à imersão e liberdade de exploração, sendo assim possível perceber uma evolução gradual da preferência dos jogadores.

A Figura 7 apresenta o percentual de jogos que oferecem determinadas funcionalidades relacionadas à plataforma.



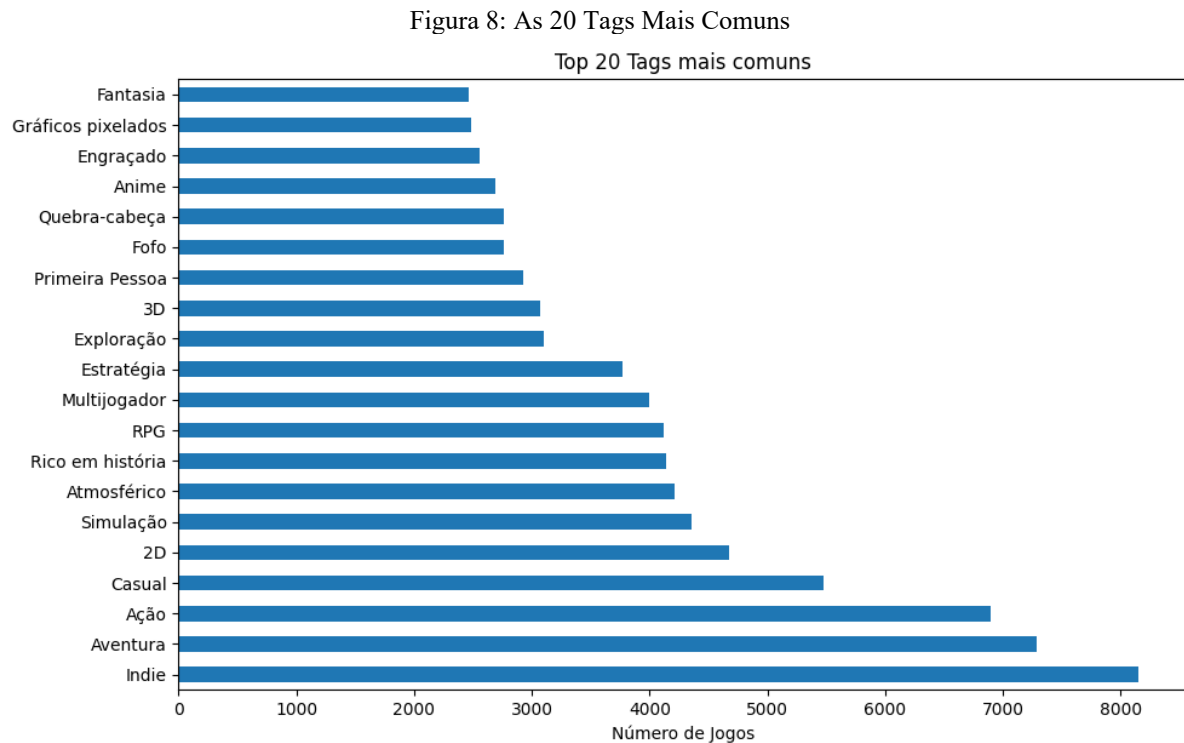
Fonte: Autoria própria

A presença de modo “Um jogador” é predominante, atingindo mais de 90% dos títulos, reforçando a raridade de jogos apenas multijogador. Em seguida, aparecem recursos como “Compartilhamento em Família” (permite que os usuários compartilhem sua biblioteca de jogos Steam com membros da família), “Conquistas Steam” (Metas e desafios que são registrados permanentemente no perfil do jogador) e “Nuvem Steam” (Serviço de armazenamento em nuvem que sincroniza e salva o progresso do jogo e configurações do usuário nos servidores da Steam), presentes em mais da metade do catálogo analisado.

Funcionalidades voltadas para a interação *online*, como “PvP Online” (modo *online* onde o foco é o combate direto e competitivo entre jogadores) e “Cooperativo Online” (modo *online* onde os jogadores trabalham juntos para alcançar um objetivo em comum), estão presentes em cerca de 15% a 20% dos jogos, evidenciando que, embora a indústria tenha forte

apelo competitivo, a maioria dos títulos publicados na Steam ainda se orienta para experiências individuais. Essa tendência sugere que o público da plataforma mantém grande interesse em jogos *single-player*, mesmo diante da popularização de experiências *multiplayer*.

A Figura 8 mostra as 20 tags mais comuns entre os jogos da plataforma.

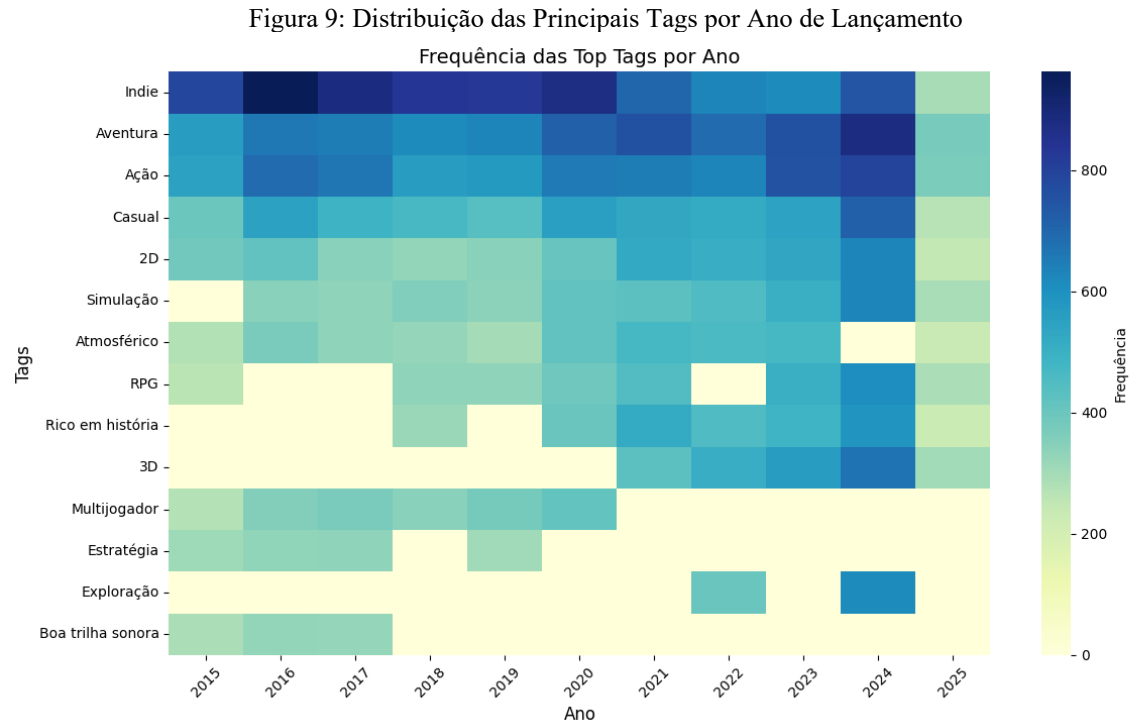


Fonte: Autoria própria

As tags “*Indie*” (se refere ao tipo de produtor do jogo, título desenvolvido por uma empresa ou indivíduo independente), “Aventura” (gênero focado em narrativa, exploração e resolução de quebra-cabeças) e “Ação” (gênero focado em desafios físicos, com alta dependência de combate, reflexos rápidos e superação de obstáculos) lideram com maior número de ocorrências, seguidas por “Casual” (gênero focado na facilidade de aprendizado e em sessões curtas de jogo), “2D” (se refere à perspectiva visual e à geometria do jogo, onde o *gameplay* se desenvolve em apenas duas dimensões) e “Simulação” (gênero focado em replicar experiências do mundo real). Esse padrão demonstra a força da cena independente na Steam, responsável por grande parte do catálogo, além da popularidade contínua de gêneros clássicos como ação e aventura.

A alta frequência de tags como “Casual” e “Simulação” sugere também que muitos jogos exploram nichos específicos e experiências menos competitivas, atendendo a diferentes perfis de jogadores.

Na Figura 9 é apresentada a distribuição das tags mais frequentes para cada ano entre 2015 e 2025.



Fonte: Autoria própria

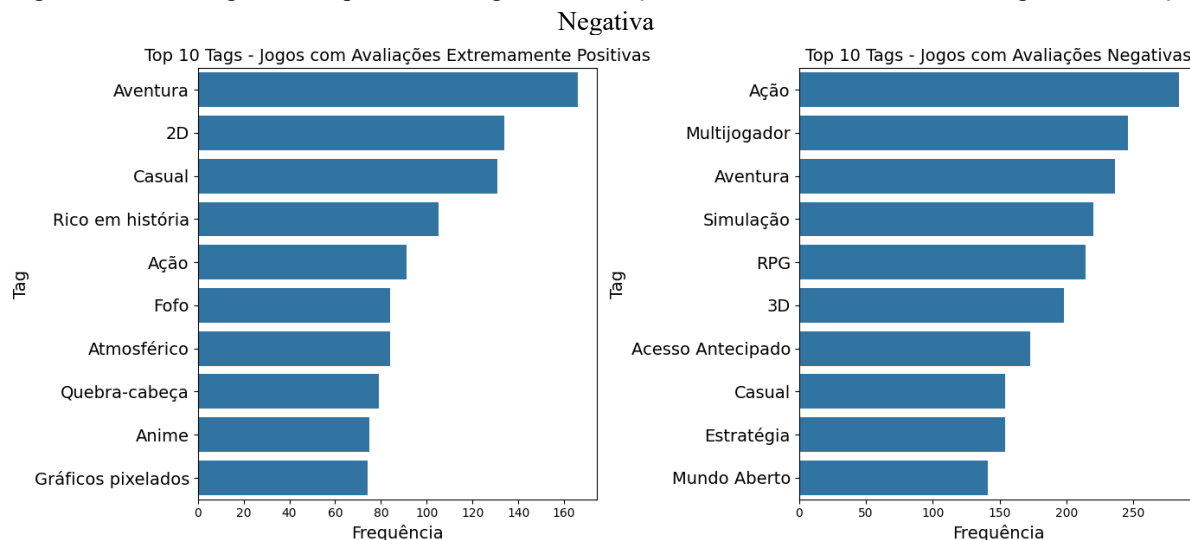
Observa-se a predominância da tag “Indie” até 2020, momento em que “Aventura” e “Ação” passam a disputar o protagonismo. Nos anos mais recentes, especialmente 2023 a 2025, as tags “Aventura” e “Ação” superam “Indie”, indicando um movimento de maior ênfase em produções com foco narrativo e interativo. Esse padrão pode estar relacionado ao amadurecimento da cena independente, em que desenvolvedores *indies* migraram de produções menores para projetos mais ambiciosos, aproximando-se de títulos de ação e aventura com maior apelo comercial.

A aparição repentina da tag “3D” (se refere à perspectiva visual e à geometria do jogo, onde o *gameplay* se desenvolve em três dimensões) nos últimos anos sugere uma mudança significativa na categorização e nas tendências de produção de jogos da Steam. Esse crescimento pode refletir tanto uma alteração na forma de rotular os jogos, com maior uso de descritores técnicos, quanto um avanço tecnológico impulsionado por motores gráficos modernos, que tornaram o desenvolvimento tridimensional mais acessível. Além disso, a ascensão de “3D” coincide com a redução da predominância da tag “Indie”, indicando um possível amadurecimento do mercado independente, que passa a investir em experiências mais complexas e visualmente elaboradas. Contudo, a ausência de “3D” entre as tags de maior média de avaliações mostra que, embora o formato tenha se popularizado, a adoção de

gráficos tridimensionais não garante maior engajamento, sendo a qualidade e a inovação os principais determinantes da boa recepção pelos jogadores.

A Figura 10 apresenta uma comparação entre as tags mais frequentes em jogos com avaliações negativas e aquelas presentes em jogos com avaliações extremamente positivas.

Figura 10: As 10 Tags Mais Populares em Jogos de Avaliação Extremamente Positiva e em Jogos de Avaliação Negativa



Nos jogos avaliados negativamente, destacam-se “Ação”, “Multijogador” e “Aventura”, gêneros que, embora populares, podem apresentar saturação de mercado ou baixa qualidade de execução em alguns títulos.

Já entre os jogos com avaliações extremamente positivas, sobressaem “Aventura”, “2D”, “Casual” e “Rico em história” (foca na imersão do jogador em uma história profunda e complexa, muitas vezes com ênfase em escolhas com consequências), sugerindo que experiências mais simples, porém com forte componente narrativo e criativo, são mais bem recebidas pelos jogadores. Esse contraste reforça a ideia de que a popularidade de uma tag não garante sucesso crítico, sendo a qualidade da execução e a adequação ao público-alvo fatores determinantes para o desempenho de um jogo.

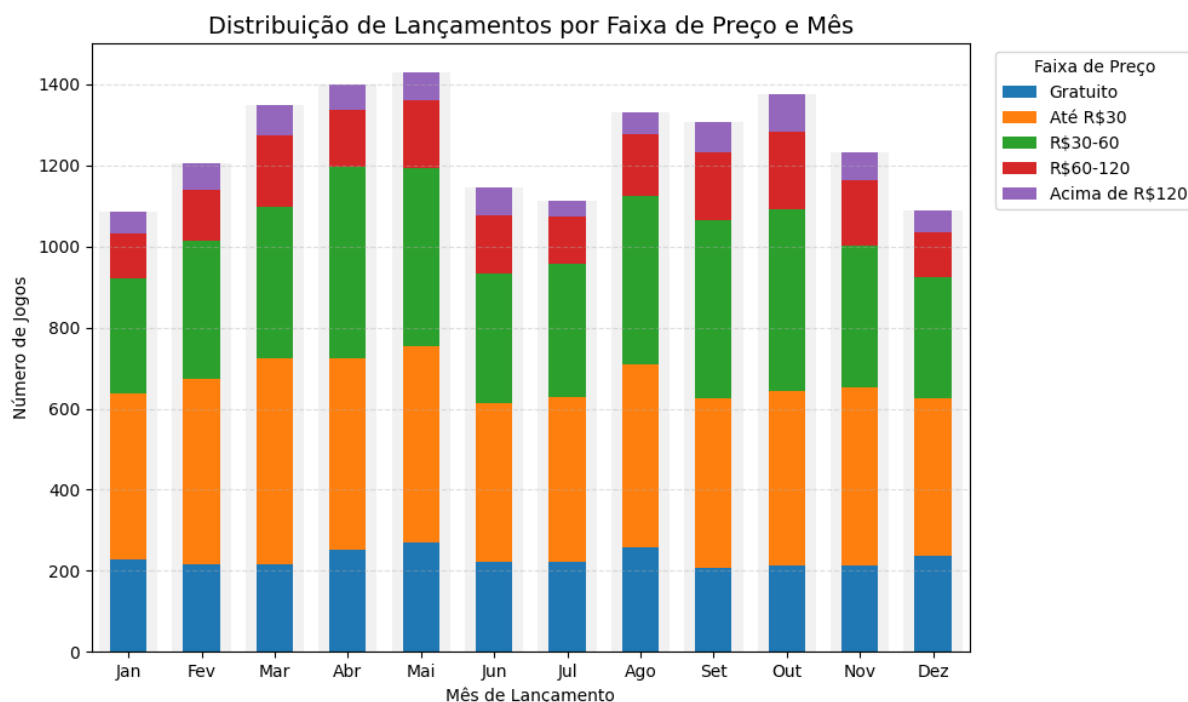
O fato de “Ação” aparecer em ambos os extremos indica que esse gênero apresenta alta saturação e grande variação de qualidade. Assim, embora o público continue interessado em experiências de ação, a execução e a inovação dentro do gênero se tornam fatores decisivos para a boa recepção crítica.

Os padrões da figura se relacionam diretamente com os padrões das Figuras 6, 8 e 9: ainda que o *multiplayer* competitivo traga alto engajamento, a avaliação positiva está mais associada à profundidade e ao design bem executado do que à categoria do jogo em si. Assim,

o sucesso crítico parece depender mais da inovação e da coesão da experiência do que da popularidade da tag.

Na Figura 11 é observada a distribuição do número de jogos lançados ao longo dos 12 meses do ano, divididos em faixas de preço.

Figura 11: Distribuição do Lançamento de Jogos de 2015 a 2025 por Mês e Faixa de Preço



Fonte: Autoria própria

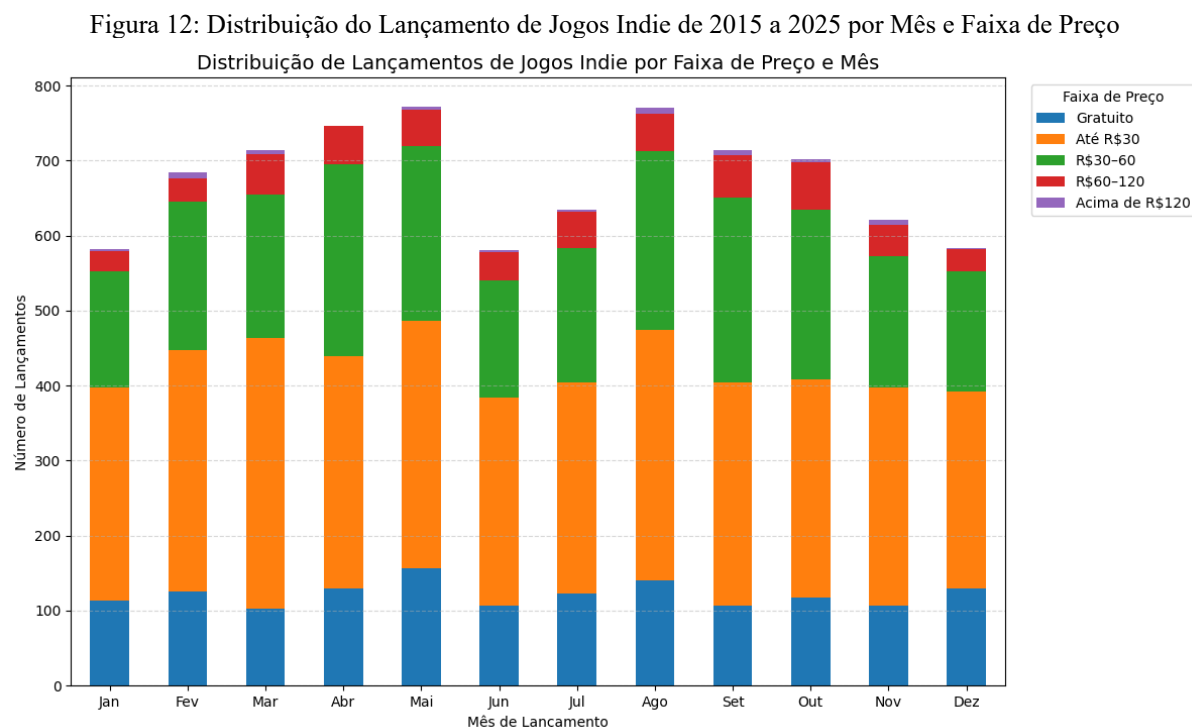
Nota-se que o maior volume de lançamentos ocorre nos meses de Abril, Maio, e Outubro, sugerindo que a indústria de jogos da Steam tende a ter seus maiores “*sprints*” de lançamento durante o Outono e a Primavera. Percebe-se, assim, que desde 2015 a indústria de jogos da Steam segue um padrão sazonal de lançamentos.

Em contrapartida, os meses de menor número são Janeiro, Junho, Julho e Dezembro, meses tipicamente de férias e feriados. Esse período coincide com o início das Promoções Sazonais da Steam, como a *Steam Summer Sale* e a *Steam Winter Sale*. Os desenvolvedores provavelmente buscam evitar a concorrência direta com as grandes promoções que dominam o comportamento de compra dos consumidores e desviam a atenção dos novos títulos de preço integral.

Em relação ao preço, a grande maioria dos jogos lançados se encontra nas faixas de preço “Grátis” e “Até R\$30”, mostrando que o mercado é principalmente impulsionado por produções de baixo custo, geralmente independentes. Esse segmento *indie*, representando 53,74% do dataset, é menos sensível à sazonalidade, lançando jogos durante todo o ano de

forma consistente. Para um desenvolvedor de jogos *low-cost*, a concorrência se mostra extremamente alta em qualquer um dos meses do ano, sendo assim, a visibilidade é um problema permanente.

A Figura 12 mostra a distribuição do número de jogos de produção independente lançados ao longo dos 12 meses do ano, divididos em faixas de preço.



Fonte: Autoria Própria

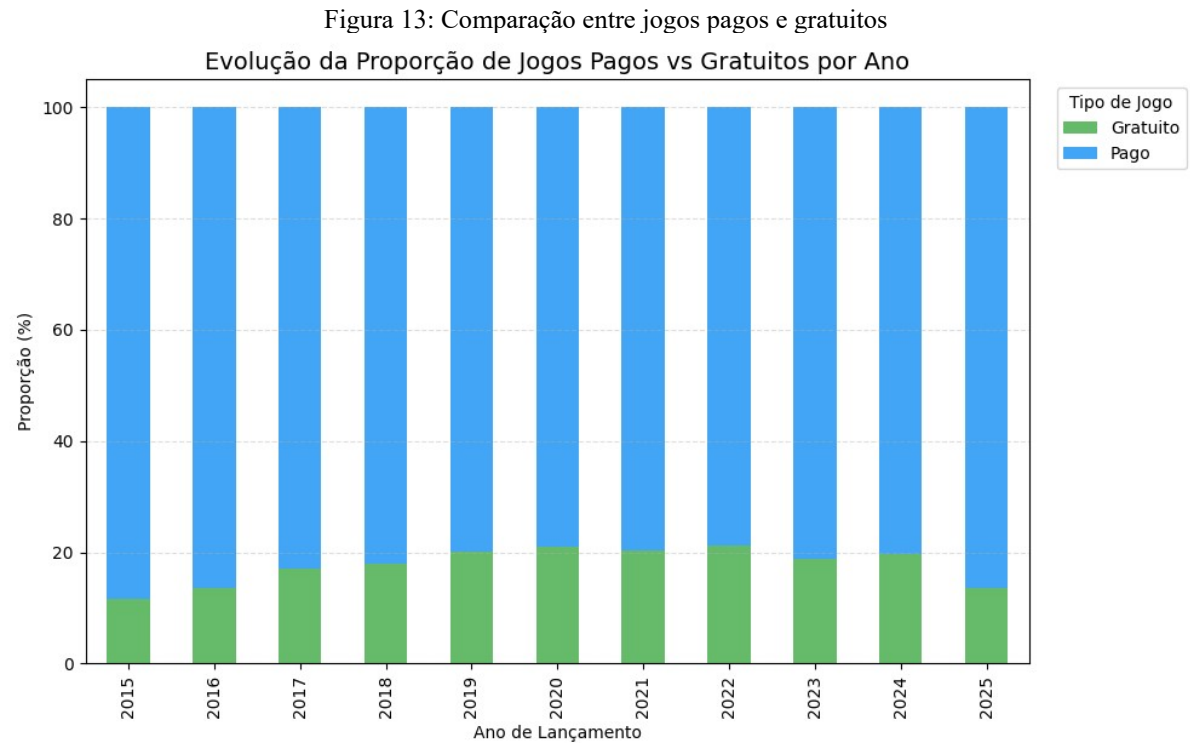
Constata-se que o mercado Indie é inundado por produções *low-cost*, evidenciado pelo volume majoritário nas faixas “Grátis” e “Até R\$30. Para jogos nesta faixa, a concorrência é máxima durante todo o ano, com picos em Maio e Agosto. Como a tag indie está presente em 53,74% dos jogos, a visibilidade na Steam é primariamente uma batalha contra outros jogos indie *low-cost*.

Identifica-se também que os períodos de queda de lançamentos parecem afetar jogos independentes apenas nas duas faixas de preço mais altas, mostrando o mesmo comportamento de fuga estratégica de produções maiores.

A análise da sazonalidade de lançamentos no mercado Steam revela um paradoxo estratégico para o desenvolvedor. Por um lado, os dados sugerem que evitar os picos de lançamento pode ser uma boa ideia para aumentar a visibilidade. Por outro lado, expõe o jogo a um ambiente dominado pela competição vertical, tornando o lançamento um risco comercial.

A Steam não permite a aplicação de descontos dentro de 30 dias de lançamento, porém a plataforma permite lançar jogos com desconto, durando até 14 dias (Valve Corporation, 2024). Uma possível estratégia para aproveitar a queda de lançamentos durante as promoções sazonais seria cronometrar o lançamento para uma semana antes do evento, garantindo que seu período de desconto de lançamento se estenda para dentro da grande promoção. Dessa forma, o jogo poderia aproveitar o grande pico de usuários gerado pela promoção sem sacrificar a visibilidade que teria como novo lançamento, transformando o período que seria um obstáculo em uma oportunidade.

Na Figura 13 é exibida uma comparação sobre o modelo de monetização dos jogos.



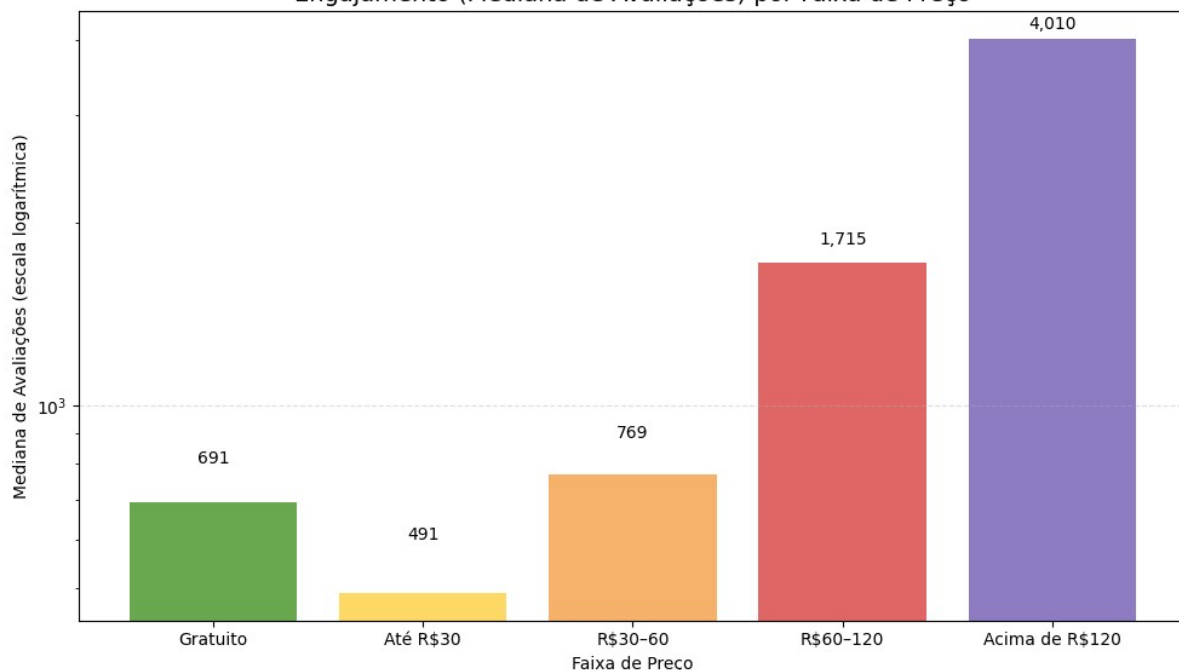
Fonte: Autoria própria

Observa-se que jogos pagos se mantiveram como modelo principal durante toda a década, sempre representando, no mínimo, 80% dos títulos anuais acima de 200 avaliações na plataforma. Isso indica que, apesar do crescimento e popularização do modelo *Free-to-Play*, a maior parte das produções reconhecidas ainda se concentra no modelo de compra única.

Embora o volume absoluto de jogos gratuitos tenha aumentado, sua participação percentual continuou praticamente a mesma. A estabilidade do modelo pago sugere que, para a maioria dos estúdios, o foco principal deve ser na janela de lançamento, pois o modelo *Live Service* ou *Free-to-Play* ainda é a exceção, possivelmente por ser uma estratégia de longo prazo, oferecendo maior risco na busca de maior retorno.

A Figura 14 mostra o engajamento por faixa de preço, utilizando a mediana das avaliações em escala logarítmica, acentuando as diferenças.

Figura 14: Engajamento por faixa de preço
Engajamento (Mediana de Avaliações) por Faixa de Preço



Fonte: Autoria própria

Observa-se uma tendência clara: quanto mais alta a faixa de preço, maior o engajamento. Como jogos de custo mais alto costumam ser grandes produções, essa tendência sugere um maior orçamento de *marketing* ou que são percebidos pelo público como “de maior qualidade”, resultando em grandes números de avaliações.

Jogos abaixo de R\$60 representam a grande maioria dos lançamentos, porém, nota-se que possuem uma baixa mediana de avaliações, indicando que embora o mercado esteja saturado, a maioria desses jogos não consegue gerar grande engajamento. O grande volume de lançamentos se traduz em baixo desempenho individual, reforçando a ideia de que a saturação canibaliza a visibilidade do pequeno desenvolvedor.

O desenvolvedor na faixa "Até R\$30" enfrenta a pior combinação do mercado: possui a maior concorrência e, simultaneamente, o menor retorno mediano de engajamento, tornando este o segmento mais arriscado para lançar um título com a esperança de destaque. O gráfico mostra que, na Steam, pagar mais e entregar um produto com escopo maior se traduz diretamente em receber mais atenção e engajamento.

6 CONCLUSÃO

O presente trabalho investigou o mercado de jogos digitais na plataforma Steam por meio da análise de dados obtidos utilizando a metodologia de *Web Scraping*. Foram utilizadas as bibliotecas Playwright, Pandas, Matplotlib e Seaborn, todas da linguagem Python, para executar a raspagem, tratamento e análise de uma amostra de 15.172 jogos lançados nos últimos 10 anos na plataforma. O processo de Análise Exploratória de Dados permitiu identificar padrões e tendências de mercado relevantes para desenvolvedores que querem compreender e se inserir no mercado atual de jogos da Steam.

Os resultados mostram uma grande dominância da funcionalidade *single-player*, presente em mais de 90% dos jogos, representando um pilar de produção na plataforma. Em termos de classificação indicativa, apesar de “*Violência*” liderar em números, “*Livre*” ainda está fortemente presente no catálogo, mostrando que há um mercado aberto à todas as idades. Indie é a tag mais popular da plataforma, indicando uma grande competitividade de desenvolvedores independentes na plataforma. A distribuição anual das tags comprova que o mercado Steam é um ecossistema em constante e rápida mudança, exigindo priorizar a inovação e o posicionamento de nicho em vez de classificações genéricas. A estabilidade do modelo de jogos pagos sugere que, para a maioria, o foco principal deve ser na janela de lançamento. Verificou-se que jogos com faixas de preço mais elevadas tenderam a apresentar um volume maior de avaliações, sugerindo que produções de maior custo e escala geram um engajamento mais intenso pela comunidade. Por fim, a qualidade e a inovação, mais do que a escolha de gênero, mostram-se fatores decisivos para a recepção crítica positiva.

Com base nas análises, foram elaboradas recomendações estratégicas para os desenvolvedores. A mais crucial aponta para a importância da visibilidade em um ambiente de lançamentos extremamente saturado, onde ser notado no lançamento é o principal desafio para o sucesso de um jogo. A sugestão é cronometrar o lançamento com desconto para uma semana antes de uma grande promoção sazonal da Steam, permitindo que o período de desconto inicial do jogo se estenda para dentro do evento de grande pico de usuários, maximizando a exposição do título novo.

Apesar do sucesso na identificação de padrões por meio da AED, tentativas de agrupamento dos dados utilizando algoritmos como K-means e DBScan enfrentaram limitações devido à alta dispersão e esparsidade dos dados. Para trabalhos futuros, sugere-se a aplicação de técnicas de redução de dimensionalidade não-lineares, como t-SNE (*t-distributed Stochastic Neighbor Embedding*) ou UMAP (*Uniform Manifold Approximation and*

Projection). Estes algoritmos podem revelar agrupamentos latentes que a Análise de Componente Principal, por ser linear, não conseguiu isolar.

REFERÊNCIAS

- BANSAL, Mudit; DAR, Muhammad Ameen; BHAT, Moshin Manzoor. Data Ingestion and Processing using Playwright. **Authorea Preprints**, Authorea, 2023. Disponível em: <https://www.techrxiv.org/doi/full/10.36227/techrxiv.22699270>. Acesso em: 9 dez. 2025.
- CARDOSO, Marcos V.; GUSMÃO, Cláudio, HARRIS, Jonathan J. (Org). **Pesquisa da indústria brasileira de games 2023**. ABRAGAMES: São Paulo, 2023. Disponível em: https://www.abragames.org/uploads/5/6/8/0/56805537/2023_relatório_final_v4.3.2_-_ptbr.pdf. Acesso em: 9 dez. 2025.
- CUNHA, Leonardo R.; PESSA, Arthur A. B.; MENDES, Renio S. **Shape patterns in popularity series of video games**. [Recurso on-line]. Maringá: Universidade Estadual de Maringá, 2024. Disponível em: <https://arxiv.org/abs/2406.10241>. Acesso em: 19 nov. 2025.
- DE LUISA, Andraž; HARTMAN, Jan; NABERGOJ, David; PAHOR, Samo; RUS, Marko; STEVANOSKI, Bozhidar; DEMŠAR, Jure; ŠTRUMBELJ, Erik. **Predicting the Popularity of Games on Steam**. [Recurso on-line]. [S.l.]: [s.n.], 2021. Disponível em: <https://arxiv.org/abs/2110.02896>. Acesso em: 19 nov. 2025.
- GUPTA, Pramod; BAGCHI, Anupam. Data Manipulation with Pandas. In: **Essentials of Python for Artificial Intelligence and Machine Learning**. Singapore: Springer, 2024. p. 197–235. Disponível em: <https://link.springer.com/book/10.1007/978-3-031-43725-0>. Acesso em: 9 dez. 2025.
- KAINZ, Alexander; PIRKER, Johanna. Steamvis: a tool for collecting and analyzing data of the game distribution platform steam. In: **2023 IEEE Conference on Games (CoG)**. [S.l.]: IEEE, 2023. p. 1–8.
- LIN, Dayi; BEZEMER, Cor-Paul; ZOU, Ying; HASSAN, Ahmed E.. An empirical study of game reviews on the Steam platform. **Empirical Software Engineering**, v. 24, p. 170–207, 2019. Disponível em: <https://seal-queensu.github.io/publications/pdf/EMSE-Dayi-2019.pdf>. Acesso em: 9 dez. 2025
- MCKINNEY, Wes. **Python for data analysis: Data wrangling with pandas, numpy, and jupyter**. **O'Reilly Media, Inc.**, 2022. Disponível em: <https://www.lkhibra.ma/books/Python-for-Data-Analysis.pdf>. Acesso em: 9 dez. 2025.
- MICROSOFT. **Playwright**. 2025. Disponível em: <https://playwright.dev>. Acesso em: 27 mai. 2025.
- MOURA, Pedro Paulo de Oliveira; HOLMES, Raphael Barbosa; DE MEDEIROS, Sheyla Natália. Aprendizado de máquina na precificação de carros usados: desenvolvimento de uma base de dados para modelos de regressão. **Paulista**, 2024. Disponível em: <https://repositorio.ifpe.edu.br/xmlui/handle/123456789/1462>. Acesso em: 9 dez. 2025.
- NETO, Nelson Weber; SOARES, Raimundo C.; COUTINHO, Luciano R.; TELES, Ariel S.. Análise exploratória de dados para identificar o impacto da pandemia da covid-19 no enem dos estados do ceará, maranhão e piauí. In: **Escola Regional de Computação do Ceará, Maranhão e Piauí (ERCE-MAPI)**. [S.l.]: SBC, 2022. p. 31–40. Disponível em: <https://sol.sbc.org.br/index.php/ercemapi/article/view/21957>. Acesso em: 9 dez. 2025.
- PONCE-BLANDÓN, José Antonio; ESPEJEL-HERNÁNDEZ, Inocencio; ROMERO-MARTÍN, Macarena; LOMAS-CAMPOS, María de las Mercedes; JIMÉNEZ-PICÓN, Nerea; GÓMEZ-SALGADO, Juan. Videogame-related experiences among regular adolescent gamers. **Plos One**, San Francisco, CA, USA, v. 15, n. 7, p. e0235327, 2020. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0235327>. Acesso em: 9 dez. 2025

PYTHON SOFTWARE FOUNDATION (PSF). **Python Programming Language – Official Website**. 2025. Disponível em: <https://www.python.org>. Acesso em: 27 mai. 2025.

SIAL, Ali Hassan; RASHDI, Syed Yahya Shah; KHAN, Abdul Hafeez. Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. **International Journal**, v. 10, n. 1, p. 277–281, 2021. Disponível em: <https://www.academia.edu/download/65736020/ijatcse391012021.pdf>. Acesso em: 9 dez. 2025.

TEJA, Andreas S; HANAFI, Muhammad Lukman I; QOMARIYAH, Nunung Nurul. Predicting Steam Games Rating with Regression. In: **E3S Web of Conferences**. [S.l.]: EDP Sciences, 2023. v. 388, p. 02001. Disponível em: https://www.e3s-conferences.org/articles/e3sconf/abs/2023/25/e3sconf_icobar2023_02001/e3sconf_icobar2023_02001.html. Acesso em: 9 dez. 2025

VALVE CORPORATION. **Acordo de Assinatura do Steam**. [Recurso on-line]. [S.l.]: Valve Corporation, 2025a. Disponível em: https://store.steampowered.com/subscriber_agreement/brazilian/. Acesso em: 18 nov. 2025.

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

VALVE CORPORATION. Steamworks Development. **Launch Discount Length Now Configurable**. Steam Community, 4 mar. 2024. Disponível em: <https://steamcommunity.com/groups/steamworks/announcements/detail/4127057033586949948>. Acesso em: 10 nov. 2025.