

Proposta Projeto final

Visualização de dados 2025.1

1. Título

Análise de dados da Steam

2. Membros da equipe

- Guilherme Sousa Lopes - 535869
- Lucas Rodrigues Aragão - 538390

3. Descrição do Projeto

Nossa ideia é coletar os dados da Steam, a principal plataforma de vendas de jogos para computadores, para que possamos ter insights e visualizações sobre estes. Para isso, resolvemos filtrar nossa busca de jogos, afinal a steam possui mais de 100.000 jogos em sua biblioteca, o que faria deixaria o trabalho de extração e tratamento dos dados muito custosa, por isso filtramos os jogos em três principais categorias:

- 100 mais vendidos na história
- 100 mais jogados nas últimas 24 horas (no momento da coleta)
- 100 mais bem avaliados*

*Filtramos aqueles jogos com avaliação **extremamente positiva**, e pegamos os 100 com mais avaliações.

4. Público Alvo

Usuários da plataforma que desejam ter mais informações sobre a loja.

5. Descrição dos dados

Os dados foram coletados de 3 diferentes fontes:

- Os mais vendidos foram retirados usando o SteamSpy
- Os jogos “em alta” foram retirados do SteamCharts
- Os jogos mais bem avaliados foram retirados da própria loja da Steam.

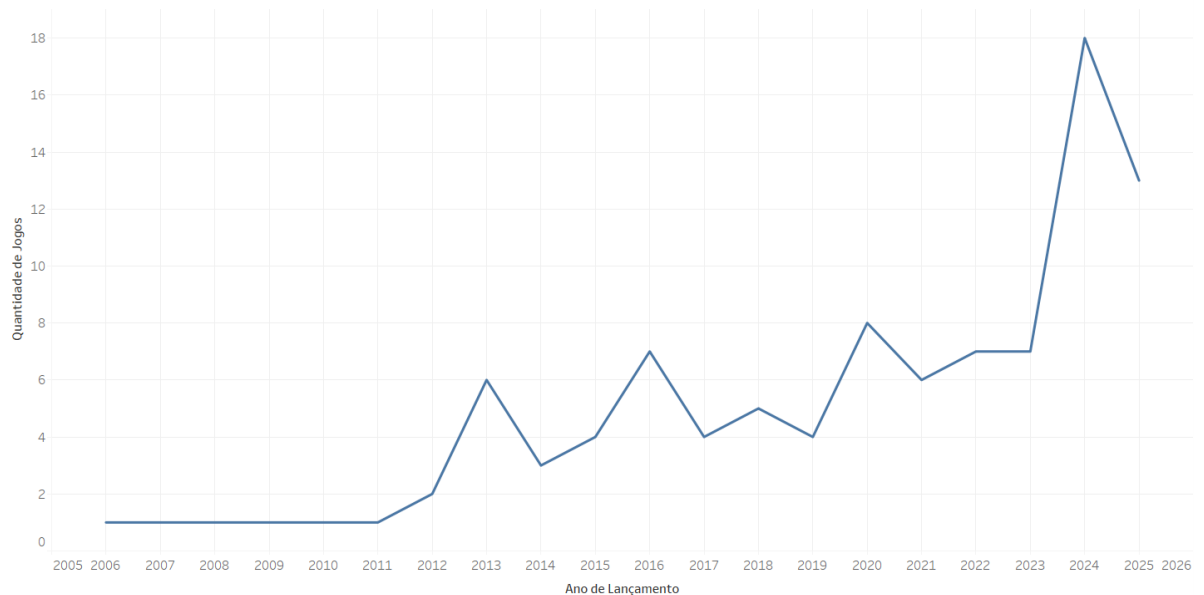
Após isso os 3 datasets foram “enriquecidos”, usando informações da Steam Store API. Com o enriquecimento temos mais informações sobre os jogos, como gênero, data de lançamento, nota do metacritic etc.

6. Perguntas de Partida

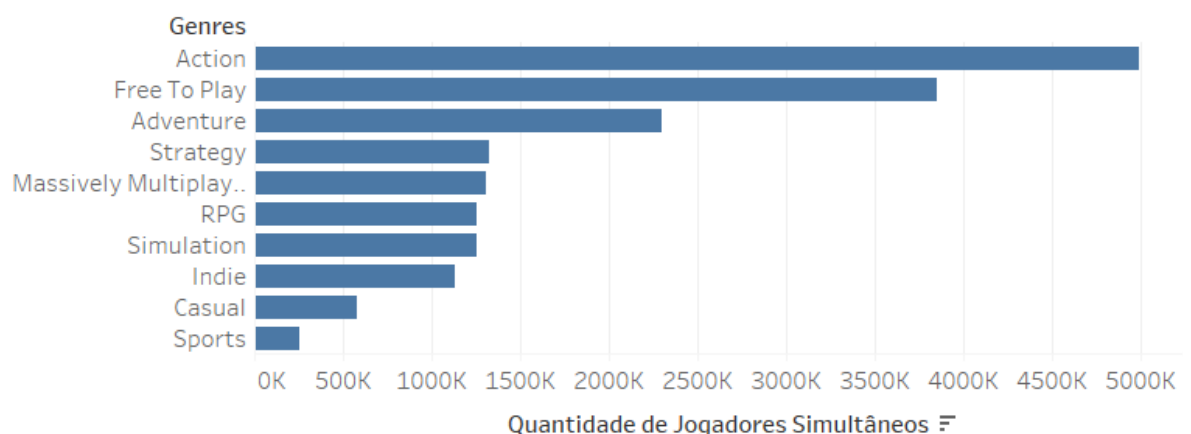
- Quais as características marcantes em cada um dos grupos?
 - Ex: Os jogos em alta, são em sua maioria, lançamentos ou multijogadores.
- Como é a relação entre os grupos?
 - Ex: Um jogo que vendeu bastante necessariamente está em alta?
- Existe algum padrão entre os jogos mais aclamados da Steam?

7. Gráficos Iniciais

Quantidade de jogos 'Em Alta' na Steam por ano de lançamento



10 Gêneros com mais jogadores simultâneos na Steam



8. Tecnologias e ferramentas utilizadas

A coleta dos jogos (nomes e ids) foi feita usando o python, por meio de requisições em APIs e web-scraping. Com os dados brutos fizemos o enriquecimento deles, coletando informações extras sobre os títulos, isso também foi feito usando requisições na API da Steam Store. A

limpeza dos dados também foi feita com o python, em que padronizamos os datasets.

Inicialmente faremos as visualizações no tableau para termos uma noção dos resultados. Após isso, migraremos para a visualização na web usando D3 e Vega-Lite API.

9. Formato do produto

A ideia é que ao final do projeto possamos entregar uma visualização web interativa, com insights sobre alguns dos principais jogos da Steam.

10. Cronograma de execução do projeto

Tarefa	Início	Fim
Escolha, extração e limpeza inicial dos dados e elaboração do projeto	29/05	01/06
Finalização de ajustes nos dados	02/06	14/06
Criação das visualizações no tableau	15/06	30/06
Criação das visualizações na Web	01/07	15/07
Ajustes finais das visualizações e elaboração dos slides de apresentação	15/07	20/07