

TEMPLATE PARA ENTREGA DO PROJETO DA DISCIPLINA
Visualização de Dados
Fase 1

Nome do estudante: Insira aqui seu nome

DESCRIÇÃO DO CONJUNTO DE DADOS ESCOLHIDO

Deve ser apresentado o dicionário de dados, ou seja, origem (link), uma descrição de cada campo do arquivo e total de registros. **IMPORTANTE:** o arquivo de dados deve ter no **mínimo 10 campos**, **mínimo 1000 registros** e **estar no formato csv**. Não podem ser usados os arquivos de dados já trabalhados nas aulas.

O conjunto de dados utilizado neste projeto foi obtido em uma base pública sobre jogos da plataforma Steam, contendo informações sobre mais de 10.000 títulos disponíveis. O arquivo já foi utilizado anteriormente em outros trabalhos, mas está atualizado para 2025, nomeado como steam_games.csv, possuindo 21 campos.

Origem (link):

<https://www.kaggle.com/datasets/srgiomanhes/steam-games-dataset-2025?resource=download>

Descrição de cada campo (o conjunto de dados deve ter no mínimo 10 campos):

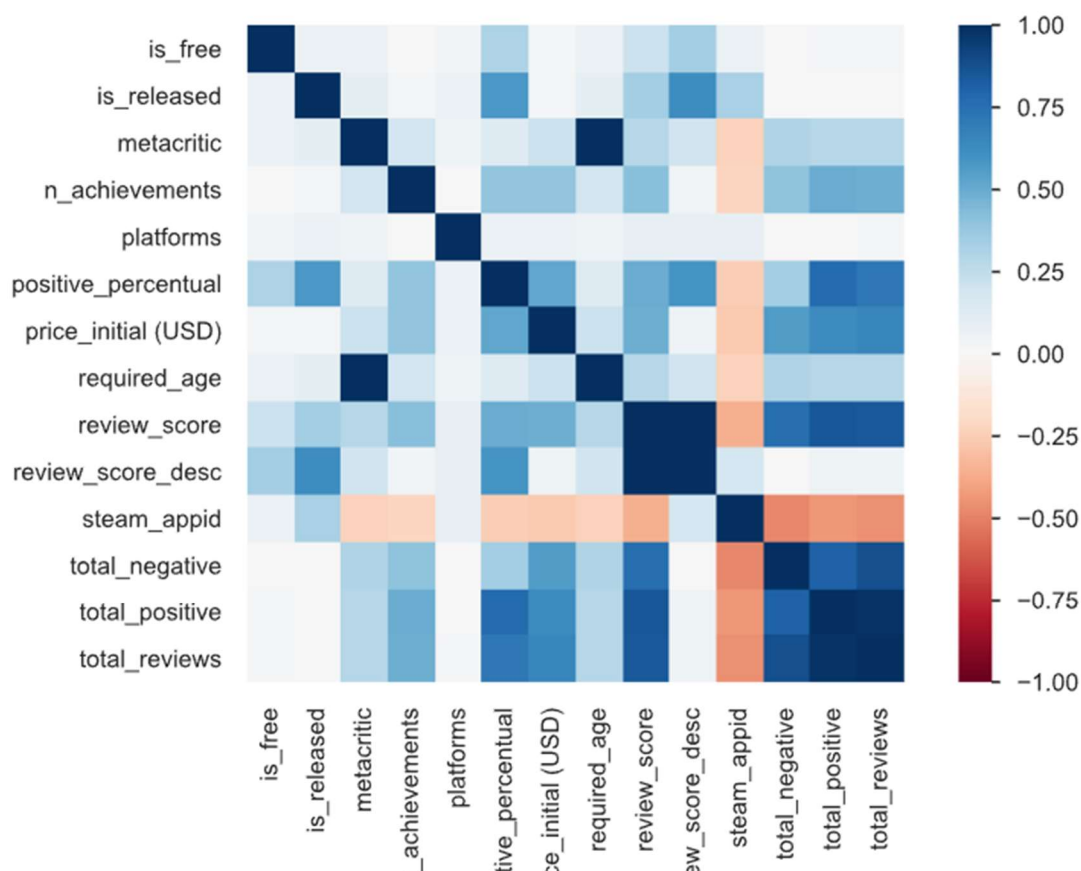
| Nome da coluna | Descrição da coluna |
|---------------------|---|
| steam_appid | Identificador único do jogo |
| name | Nome do jogo |
| developers | Nome dos desenvolvedores |
| publishers | Nome do publicador |
| categories | Categorias (Single-Player, Co-Op, etc) |
| genres | Generos do jogo (Action, Adventure, RPG, etc) |
| required_age | Faixa etária |
| n_achievements | Número de conquista do jogo |
| platforms | Plataformas compatíveis |
| is_released | Status de lançamento |
| release_date | Data de lançamento do jogo |
| additional_content | Conteúdo adicional (DLC, Skins, etc) |
| total_reviews | Total de avaliações do jogo |
| total_positive | Total de avaliações positivas do jogo |
| total_negative | Total de avaliações negativas do jogo |
| review_score | Pontuação média (0-10) |
| review_score_desc | Classificação textual associada à nota (ex: "Very positive", "Neutral", |
| positive_percentual | Percentual de avaliações positivas |
| metacritic | Nota média atribuída pela crítica especializada |
| is_free | Indica se o jogo é gratuito ou não |
| price_initial(USD) | Preço inicial do jogo em dólares |

Total de registros (mínimo 1000):
71.430 registros, com cabeçalho.

DESCRIÇÃO DA ANÁLISE DOS DADOS

Devem ser apresentadas **informações e gráficos sobre o conteúdo e a qualidade dos dados**, isto é, *insights* obtidos com a análise exploratória visual (incluindo gráficos), e uma descrição sobre o enriquecimento dos dados (se realizado). **IMPORTANTE:** para esta apresentação, use uma das ferramentas trabalhadas em aula, como o **ydata-profiling** (acesse as aulas 1 e 2 para saber mais sobre as ferramentas).

A análise dos dados da Steam demonstra uma diferente distribuição entre os dados, mantendo correlações (como demonstra o Treemap) facilitando a análise e montagem das questões que apresentei. A Steam é uma plataforma de amplo sucesso e demonstra organização dentro de seu dataset, mas ainda foi necessário a formatação de algumas colunas, como a “release_date”, possuindo valores não muito distintos, mas com formatação inadequada para os códigos, como mostra na última imagem.



release_date

Text

| | |
|--------------|---------|
| Distinct | 4346 |
| Distinct (%) | 6.1% |
| Missing | 0 |
| Missing (%) | 0.0% |
| Memory size | 5.1 MiB |



DESCRIÇÃO DAS QUESTÕES DE NEGÓCIO

Devem ser definidas, **no mínimo, 3 questões de negócio**, ou seja, o que se quer descobrir com o *dashboard* que será desenvolvido.

Questão de negócio 1:

Quais as empresas com jogos mais bem avaliados?

Descrição:

O objetivo é identificar gêneros com as melhores avaliações médias, refletindo o domínio do mercado em termos de qualidade percebida pelo público. Essa análise permite compreender quais gêneros mantêm padrões de excelência e fidelizam a base de jogadores, oferecendo uma visão global sobre quais as preferencias dos usuarios na plataforma Steam.

Campos utilizados: *genres, review_score, steam_appid*.

Questão de negócio 2:

Quais os jogos mais vendidos lançados entre 2020 e 2025?

Descrição:

Essa questão visa observar a relação entre **preço** e **volume estimado de vendas**, com o intuito de identificar tendências de precificação no mercado recente. A análise auxilia na compreensão do comportamento dos consumidores, permitindo inferir faixas de preço que não prejudicam o desempenho comercial dos títulos mais recentes.

Campos utilizados: *name, release_date, price_initial (USD), total_reviews*.

Questão de negócio 3:

Quais os gêneros de jogos mais lançados nos últimos 10 anos (2015-2025)?

Descrição:

Essa análise permite compreender o comportamento da indústria ao longo do tempo, identificando quais gêneros tiveram crescimento ou retração no número de lançamentos. O resultado indica tendências de mercado e auxilia na previsão de gêneros emergentes ou saturados, o que é essencial para desenvolvedores e investidores.

Campos utilizados: *genres, release_date*.

Questão de negócio 4 (OPCIONAL):

Qual o preço médio dos jogos por gênero?

Descrição:

O objetivo é investigar como os preços variam entre os diferentes gêneros de jogos, identificando quais tipos apresentam custos médios mais elevados. Essa análise permite inferir a relação entre complexidade de produção, público-alvo e posicionamento de mercado.

Campos utilizados: *genres, price_initial (USD)*.

Questão de negócio 5 (OPCIONAL):

Como se distribuem os jogos por plataforma e gênero?

Descrição: Esta análise visa determinar a matriz de suporte multiplataforma dentro dos gêneros mais frequentes da Steam, quantificando o predomínio do Windows e a extensão do desenvolvimento para macOS e Linux. Este mapeamento é fundamental para entender as estratégias de portabilidade e a concentração de mercado. O processo inclui o tratamento rigoroso de valores faltantes para garantir que a contagem de jogos não seja artificialmente reduzida, oferecendo uma visão completa da base de dados.

Campos utilizados: *genres, platforms*.

LISTA E JUSTIFICATIVA DAS REPRESENTAÇÕES VISUAIS A SEREM IMPLEMENTADAS

Devem ser listadas todas as representações visuais que farão parte do *dashboard* que será implementado na Fase 2. Além disso, para cada representação visual deve ser colocada uma justificativa para a sua escolha e devem ser listados os campos dos dados que serão utilizados para sua futura implementação. **IMPORTANTE:** o *dashboard* a ser implementado na Fase 2 deve ter de 5 a 6 diferentes técnicas de visualização interativa para ajudar a responder as questões de negócio.

Representação visual 1:

Visualização proposta:

Treemap (Gráfico de Árvore): Essa técnica permite representar visualmente a proporção de avaliações agregadas por empresa, destacando o peso de cada desenvolvedora em termos de popularidade e nota média. A hierarquia e a área proporcional de cada retângulo tornam a leitura intuitiva e facilitam comparações diretas.

Distribuição e Média de Avaliação dos Jogos por Gênero (Steam)

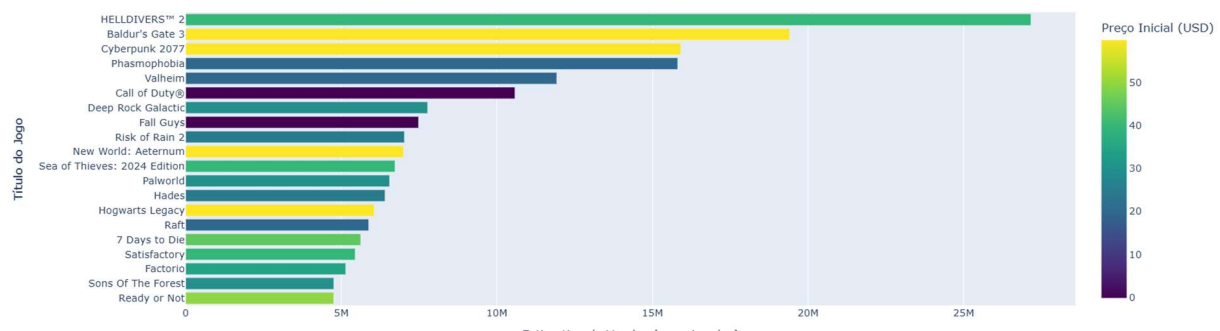


Representação visual 2:

Visualização proposta:

Gráfico de barras horizontal interativo: Utilizado para exibir o *Top 20* de jogos mais vendidos entre 2020 e 2025, com escala de cores baseada no preço inicial (USD). Essa forma de visualização facilita a comparação e evidência quais jogos combinam alto preço e bom desempenho de vendas.

Top 20 Jogos Mais Vendidos na Steam (2020-2025) com Escala de Preço

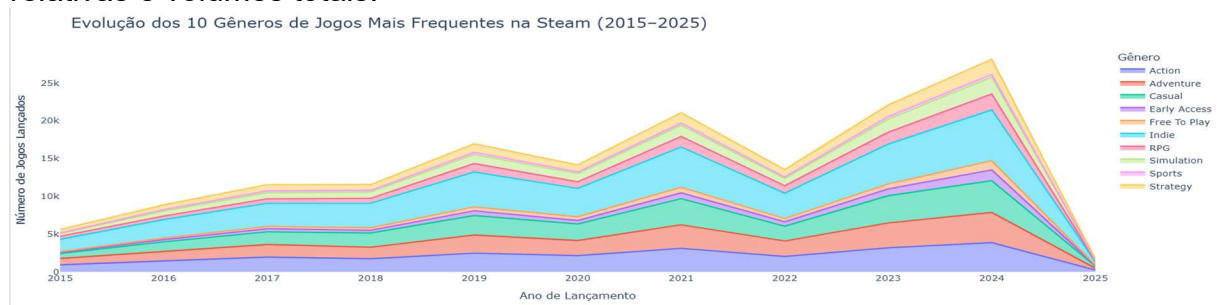


Representação visual 3:

Visualização proposta:

Gráfico de área empilhada interativo: Essa técnica mostra a evolução temporal do número de lançamentos por gênero, destacando a dominância de categorias

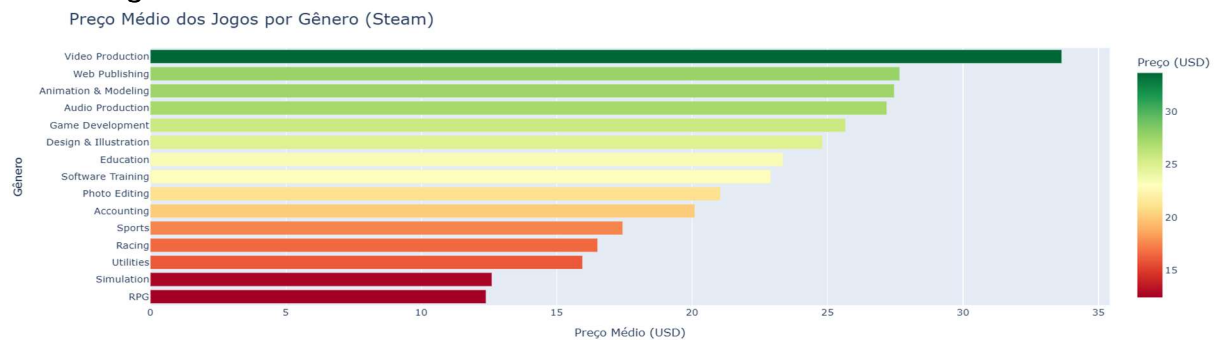
específicas ao longo dos anos. A área empilhada facilita a percepção de proporções relativas e volumes totais.



Representação visual 4:

Visualização proposta:

Gráfico de barras horizontal com escala de cores: Essa técnica apresenta os gêneros ordenados pelo preço médio, com coloração em gradiente para destacar os valores extremos. É uma forma eficiente de visualizar a diferença de precificação entre segmentos.



Representação visual 5:

Visualização proposta:

Gráfico de barras empilhadas interativo: Representa a quantidade de jogos por gênero, empilhando as diferentes plataformas. Esse formato é ideal para visualizar a distribuição relativa de plataformas dentro de cada categoria e o predomínio de lançamentos para Windows.

