

Análise Exploratória de Dados (EDA) - IMDB

Este relatório apresenta a Análise Exploratória de Dados (EDA) realizada sobre a base de filmes do IMDB, após o tratamento e organização do dataset.

Estrutura do Dataset

- Total de registros: 811 filmes
- Colunas principais:
 - Series_Title → título do filme
 - Released_Year → ano de lançamento -
 - Certificate → classificação indicativa -
 - Runtime → duração em minutos -
 - primary_genre e secondary_genres → gênero principal e secundários -
 - Meta_score → nota de crítica
 - Director → diretor -
 - Star1 a Star4 → principais atores/atrizes -
 - No_of_Votes → número de votos no IMDB
 - Gross → bilheteria

Qualidade dos Dados

- Valores ausentes em: - Certificate: 47
- secondary_genres: 105
- Ajustes aplicados:
 - Conversão de Runtime, No_of_Votes e Gross para valores numéricos.
 - Separação de gêneros em coluna principal (primary_genre) e secundária
 - Retiradas linhas com No_of_Votes, mas sem Meta_score
 - Retiradas linhas com Gross menor que 6 dígitos, mas foram deixadas as linhas sem esse valor

Distribuição Geral em gênero primário

- Gênero mais frequente: Drama (239 filmes) -
- Outros gêneros relevantes: Action (147), Comedy (121), Biography (78), Crime (74),
- Gêneros minoritários: Thriller (1), Film-Noir (2), Western (4).

Estatísticas Descritivas

- Duração média (Runtime): ~123 minutos (máx. 321 min).
- Meta Score médio: 77,9 (escala de 0 a 100).
- Número médio de votos: 317 mil (máx. 2,3 milhões).
- Bilheteria média (Gross): 79 milhões, com picos próximos a 936 milhões.

Relações Importantes

- Votos x Bilheteria: forte correlação positiva (0.56).
- Filmes mais votados tendem a ter maior arrecadação.
- Meta Score x Bilheteria: correlação fraca → boa crítica não garante bilheteria.
- Runtime: pouca relação com sucesso comercial ou críticas.

Atores e Diretores

- Atores mais recorrentes: Tom Hanks, Emma Watson, Rupert Grint, Michael Caine.
- Diretores mais frequentes: Alfred Hitchcock (13 filmes), Christopher Nolan, Steven Spielberg.

Conclusões

1. Filmes de gêneros populares e comerciais (Action, Animation, Adventure) se destacam em bilheteria.
2. A popularidade medida por votos tem mais impacto financeiro que a crítica especializada.
3. O dataset mostra forte presença de dramas, mas os grandes sucessos de bilheteria vêm de gêneros mais voltados ao entretenimento de massa.