

Análise Exploratória dos Dados (EDA):

. Análise Descritiva das Variáveis Numéricas

A análise descritiva fornece um resumo estatístico das variáveis numéricas no conjunto de dados:

- IMDB_Rating:
 - Média: 8.1
 - Mediana: 8.1
 - Desvio Padrão: 0.4
 - Mínimo: 7.0
 - Máximo: 9.3
- Meta_score:
 - Média: 77.4
 - Mediana: 79.0
 - Desvio Padrão: 10.7
 - Mínimo: 40.0
 - Máximo: 100.0
- No_of_Votes:
 - Média: 754,569
 - Mediana: 463,328
 - Desvio Padrão: 896,164
 - Mínimo: 25,000
 - Máximo: 2,303,232
- Gross:
 - Média: \$140,147,141
 - Mediana: \$94,780,789
 - Desvio Padrão: \$146,487,081
 - Mínimo: \$4,360,000
 - Máximo: \$936,662,225

2. Distribuição das Notas do IMDB

A maioria dos filmes possui notas entre 7 e 9, com um pico em torno de 8.0. Isso sugere que os filmes no conjunto de dados são geralmente bem avaliados.

3. Distribuição da Receita Bruta (Gross)

A distribuição da receita bruta é altamente assimétrica, com muitos filmes gerando receitas moderadas e alguns filmes atingindo receitas extremamente altas. Esse padrão é comum em indústrias de entretenimento, onde alguns sucessos de bilheteria impulsionam grande parte das receitas totais.

4. Distribuição dos Gêneros dos Filmes

Os gêneros mais frequentes são Drama, Crime, e Ação. Isso indica que esses gêneros são populares e provavelmente bem representados no banco de dados.

5. Correlação entre Variáveis

A análise de correlação entre variáveis mostra as seguintes relações significativas:

- IMDB_Rating vs. Meta_score:

- Correlação positiva moderada (0.57), indicando que filmes com alta pontuação no Metacritic tendem a ter boas notas no IMDB.
- No_of_Votes vs. Gross:
 - Correlação positiva baixa (0.26), sugerindo que mais votos não se traduzem diretamente em maior receita.
- IMDB_Rating vs. Gross:
 - Correlação positiva baixa (0.27), indicando que uma alta avaliação no IMDB não está fortemente associada a maiores receitas.

Hipóteses Relacionadas

1. Hipótese 1: Filmes de gêneros populares (como Drama e Ação) tendem a gerar receitas mais altas.
 - Podemos investigar essa hipótese analisando a receita bruta média por gênero.
2. Hipótese 2: O número de votos é um indicativo da popularidade do filme, mas não necessariamente de sua receita bruta.
 - A baixa correlação entre No_of_Votes e Gross sugere que a popularidade medida por votos não está diretamente ligada ao sucesso financeiro.
3. Hipótese 3: Filmes bem avaliados no Metacritic também recebem boas notas no IMDB.
 - A correlação moderada entre Meta_score e IMDB_Rating suporta essa hipótese.
4. Hipótese 4: A receita bruta não é um bom preditor da nota do IMDB.
 - A correlação baixa entre Gross e IMDB_Rating sugere que a qualidade percebida de um filme (nota no IMDB) não está diretamente ligada ao seu desempenho financeiro.

Conclusões e Insights

- Recomendação de Filme: "The Godfather" (Nota 9.2) é uma excelente escolha, sendo um clássico amplamente aclamado.
- Fatores de Faturamento: Gêneros populares, número de votos, e a pontuação no Metacritic são fatores importantes.
- Insights da Coluna Overview: Utilizando técnicas de processamento de linguagem natural (NLP), é possível inferir o gênero do filme a partir do resumo.
- Predição de Nota do IMDB: Um modelo de Random Forest foi utilizado para prever a nota do IMDB com base em características como gênero, meta_score, número de votos, e receita bruta. O modelo teve melhor desempenho em termos de RMSE e R^2 em comparação com a Regressão Linear.