

Relatório de Análise de Filmes e Previsão de Nota do IMDB

1. Análise Inicial do Banco de Dados

A análise começou com uma revisão completa do banco de dados, na qual foram identificados diversos problemas, como informações incorretas, faltantes ou equivocadas. Por exemplo, no caso do filme Apollo 13 (linha 967), o ano de lançamento aparece listado como "PG" ao invés de 1995.

2. Análise Exploratória de Dados (EDA)

Durante a EDA, constatou-se que apenas três colunas apresentavam dados faltantes:

```
count  Unnamed: 0  IMDB_Rating  Meta_score  No_of_Votes
mean    500.000000    7.947948    77.969121    2.716214e+05
std     288.530761    0.272290    12.383257    3.209126e+05
min      1.000000    7.600000    28.000000    2.508800e+04
25%     250.500000    7.700000    70.000000    5.547150e+04
50%     500.000000    7.900000    79.000000    1.383560e+05
75%     749.500000    8.100000    87.000000    3.731675e+05
max     999.000000    9.200000   100.000000    2.303232e+06
Unnamed: 0      0
Series_Title     0
Released_Year    0
Certificate      101
Runtime          0
Genre            0
IMDB_Rating      0
Overview         0
Meta_score       157
Director         0
Star1            0
Star2            0
Star3            0
Star4            0
No_of_Votes      0
Gross           169
dtype: int64
IMDB_Rating
7.7    157
7.8    151
8.0    141
8.1    127
7.6    123
7.9    106
8.2     67
8.3     44
8.4     31
8.5     20
8.6     15
8.8      5
8.7      5
9.0      3
8.9      3
9.2      1
```

(Meta_score: 157 registros ausentes)

Gross: 169 registros ausentes

Certificate: 101 registros ausentes)

Além disso, a coluna Gross apresenta dados extremamente imprecisos. Por exemplo, os top 20 filmes segundo o banco mostram valores de faturamento divergentes da realidade:

Top 20 do banco (em USD):

Star Wars: Episode VII - The Force Awakens — 936.662.225
Avengers: Endgame — 858.373.000
Avatar — 760.507.625
Titanic — 659.325.379
The Avengers — 623.279.547
Incredibles 2 — 608.581.744
The Dark Knight — 534.858.444
Rogue One — 532.177.324
The Dark Knight Rises — 448.139.099
E.T. the Extra-Terrestrial — 435.110.554
Toy Story 4 — 434.038.008
The Lion King — 422.783.777
Toy Story 3 — 415.004.880
Captain America: Civil War — 408.084.349
Jurassic Park — 402.453.882
Guardians of the Galaxy Vol. 2 — 389.813.101
Harry Potter and the Deathly Hallows: Part 2 — 381.011.219
Finding Nemo — 380.843.261
The Lord of the Rings: The Return of the King — 377.845.905
Deadpool — 363.070.709

Top 20 seguindo o faturamento real (em USD):

Avatar — 2.923.710.708
Avengers: Endgame — 2.799.439.100
Titanic — 2.264.812.968
Star Wars: Episode VII - The Force Awakens — 2.071.310.218
Harry Potter and the Deathly Hallows: Part 2 — 1.342.139.727
Captain America: Civil War — 1.153.296.293
The Lord of the Rings: The Return of the King — 1.146.436.214
Jurassic Park — 1.104.379.926
The Dark Knight Rises — 1.084.939.099
Toy Story 4 — 1.073.394.593
Toy Story 3 — 1.067.316.101
Rogue One — 1.058.684.742
The Dark Knight — 1.009.195.950
The Lion King (1994) — 979.161.373
Finding Nemo — 941.637.960
Guardians of the Galaxy Vol. 2 — 863.756.051
The Avengers — 1.520.538.536
Incredibles 2 — 1.242.805.359
E.T. the Extra-Terrestrial — 797.307.407
Deadpool — 782.837.347

Outro ponto relevante é que a ordenação do banco utiliza a nota do IMDB. Apesar de confiável, seria mais adequado considerar o Metacritic, pois no IMDB qualquer pessoa pode votar, misturando votos de público e críticos. Sites como Metacritic e Rotten Tomatoes separam as avaliações de público e críticos, permitindo uma análise mais precisa da recepção do filme.

Entretanto, mesmo com metodologias diferentes, a média e mediana das notas entre público e críticos são bastante próximas ao tratar de forma geral os 999 filmes. Discrepâncias existem, como *I Am Sam*, que possui 7.7 no IMDB e apenas 2.8 no Metacritic.

3. Padrões Observados

As notas dos filmes não são a única informação que pode ser percebida com uma análise simples dos dados; também é possível identificar padrões nos gêneros dos filmes. De maneira geral, a maioria das obras possui 'Drama' como um dos gêneros (723 casos). Mesmo os filmes que não apresentam a classificação de drama, geralmente contêm elementos dramáticos em sua narrativa. Esse padrão pode ser explicado por diversos fatores "*Muitas vezes, histórias trágicas se concentram em temas de amor eterno e isso leva os espectadores a pensar sobre seus entes queridos e a contar suas conquistas*", disse a professora de comunicação Silvia Knobloch-Westerwick. Além disso, algumas obras exploram experiências compartilhadas, abordando questões universais, como pressão estética e envelhecimento, como ocorre no filme *A Substância*.

Ao analisar todos esses fatores, percebe-se que certos padrões narrativos se repetem constantemente. No entanto, o sucesso ou a qualidade de um filme não se limita apenas a esses elementos, como será discutido posteriormente.

4. Perguntas da Segunda Parte

4.1 Qual filme você recomendaria para uma pessoa que você não conhece?

The Dark Knight, por ser uma narrativa bem estruturada, com personagens conhecidos, direção equilibrada e potencial para agradar diferentes públicos.

4.2 Quais fatores estão relacionados a alto faturamento de um filme?

São uma reunião de fatores mas os principais seriam:

- Franquias ou universos conhecidos/consagrados
- Possível inovação tecnológica (ex.: 3D em *Avatar*)
- Investimento em marketing
- Efeito boca-a-boca e repercussão em redes sociais

4.3 É possível inferir o gênero a partir da coluna Overview?

-Sim, a sinopse permite inferir parcialmente o gênero e subgêneros, por exemplo, Get Out (Corra!) sugere suspense e elementos de racismo a partir da descrição textual. Esses padrões ajudam a compreender a repetição de elementos narrativos, mas o sucesso ou qualidade de um filme não depende apenas deles.

5. Previsão da Nota do IMDB

A previsão da nota do IMDB foi tratada como um problema de regressão, pois a variável alvo é numérica e contínua.

Para a modelagem, foram utilizadas apenas variáveis objetivas relacionadas ao filme, como Runtime (duração), Released_Year (ano de lançamento), Meta_score (avaliação crítica), No_of_Votes (número de votos) e Gross (faturamento bruto).

Essas variáveis foram escolhidas porque podem refletir aspectos de qualidade, relevância histórica, aceitação da crítica e do público, sem envolver fatores subjetivos como gênero, que não determinam diretamente a nota de um filme.

Os dados faltantes foram tratados substituindo-se por valores médios de cada coluna. O modelo escolhido foi o Random Forest Regressor, que captura relações não lineares, apresenta robustez contra outliers e reduz riscos de sobreajuste, embora tenha como desvantagens a menor interpretabilidade e maior custo computacional em comparação com modelos lineares.

A avaliação do desempenho foi feita com três métricas: MAE, que mostra o erro médio absoluto em pontos de nota (sendo a métrica principal, por sua interpretabilidade direta); RMSE, que penaliza erros maiores; e R^2 , que indica a proporção da variância explicada pelo modelo. Assim, o modelo conseguiu prever de forma consistente as notas do IMDB a partir de variáveis numéricas e relevantes.

O modelo conseguiu prever de forma consistente as notas do IMDB a partir de variáveis objetivas com os dados que lhe foram dados (os quais não são os referentes a 2025).

Dados do banco original:

```
{'Series_Title': 'The Shawshank Redemption',  
'Released_Year': '1994',  
'Certificate': 'A',  
'Runtime': '142 min',  
'Genre': 'Drama',  
'Overview': 'Two imprisoned men bond over a number of years, finding solace and  
eventual redemption through acts of common decency.',  
'Meta_score': 80.0,  
'Director': 'Frank Darabont',  
'Star1': 'Tim Robbins',  
'Star2': 'Morgan Freeman',  
'Star3': 'Bob Gunton',
```

```
'Star4': 'William Sadler',  
'No_of_Votes': 2343110,  
'Gross': '28,341,469'}
```

Dados atualizados aproximados:

```
{'Series_Title': 'The Shawshank Redemption',  
'Released_Year': '1994',  
'Certificate': 'A',  
'Runtime': '142 min',  
'Genre': 'Drama',  
'Overview': 'Two imprisoned men bond over a number of years, finding solace and  
eventual redemption through acts of common decency.',  
'Meta_score': 82.0,  
'Director': 'Frank Darabont',  
'Star1': 'Tim Robbins',  
'Star2': 'Morgan Freeman',  
'Star3': 'Bob Gunton',  
'Star4': 'William Sadler',  
'No_of_Votes': 3100000,  
'Gross': '73,300,000'}
```

O modelo previu 8.77, próximo da nota real de 9.3, considerando que os dados originais estavam desatualizados.

6. Recomendações para Produção de Filmes

Portanto, considerando que a PProductions deseja produzir um filme e almeja um bom faturamento, a recomendação baseada nos dados é a seguinte: buscar alguma série famosa que ainda não tenha sido adaptada e que possua certa relevância, contratar um diretor consagrado (ou já reconhecido popularmente como competente) e investir significativamente em marketing. Não é necessário que seja uma obra totalmente nova; também é possível revisitar adaptações mal-sucedidas. Um exemplo disso é Percy Jackson: os dois filmes anteriores foram mal recebidos, mas a HBO adquiriu os direitos e está produzindo uma nova série do zero. Apenas essa notícia gerou marketing suficiente para aumentar o número de assinantes da plataforma de streaming.