

Relatório de Análise Cinematográfica para PProductions

Análise Estratégica para o Desenvolvimento de Novos Filmes

Autor: Everton Reis de Souza

Data: 03 de setembro de 2025

1. Resumo Executivo

Este relatório apresenta uma análise aprofundada do conjunto de dados cinematográficos `imdb_dataset`, com o objetivo de fornecer à PProductions insights estratégicos para o desenvolvimento de seu próximo filme. Através de uma análise exploratória, modelagem preditiva e resposta a questões de negócio, identificamos os principais fatores que influenciam o sucesso de um filme, tanto em termos de crítica quanto de faturamento. As recomendações aqui contidas visam mitigar riscos e maximizar o retorno sobre o investimento em futuras produções.

2. Análise Exploratória dos Dados (EDA)

A primeira etapa do projeto consistiu na preparação e análise exploratória dos dados. O dataset inicial continha 1000 registros de filmes com 15 atributos distintos, incluindo título, gênero, ano de lançamento, nota IMDb, faturamento (Gross) e pontuação da crítica (Meta_score).

O processo de limpeza incluiu a conversão de tipos de dados (como `Runtime` e `Gross` para valores numéricos) e o tratamento de valores ausentes nas colunas `Certificate`, `Meta_score` e `Gross` utilizando a média ou a moda, garantindo a integridade do conjunto de dados para a análise.

As principais descobertas desta fase foram:

- Alta Qualidade dos Filmes:** O conjunto de dados possui um viés para filmes de alta avaliação, com a maioria das notas IMDb concentrada entre 7.7 e 8.5. Isso indica que estamos analisando um portfólio de filmes bem-sucedidos.
- Domínio do Gênero Drama:** O gênero "Drama" é, de longe, o mais prevalente, aparecendo em mais de 700 dos filmes analisados. Gêneros como "Ação", "Aventura" e "Comédia" também são muito populares, frequentemente aparecendo em combinação com outros.
- Crescimento da Produção:** Observamos um aumento significativo no número de filmes de alta qualidade a partir dos anos 1990, com um pico nas décadas de 2000 e 2010.

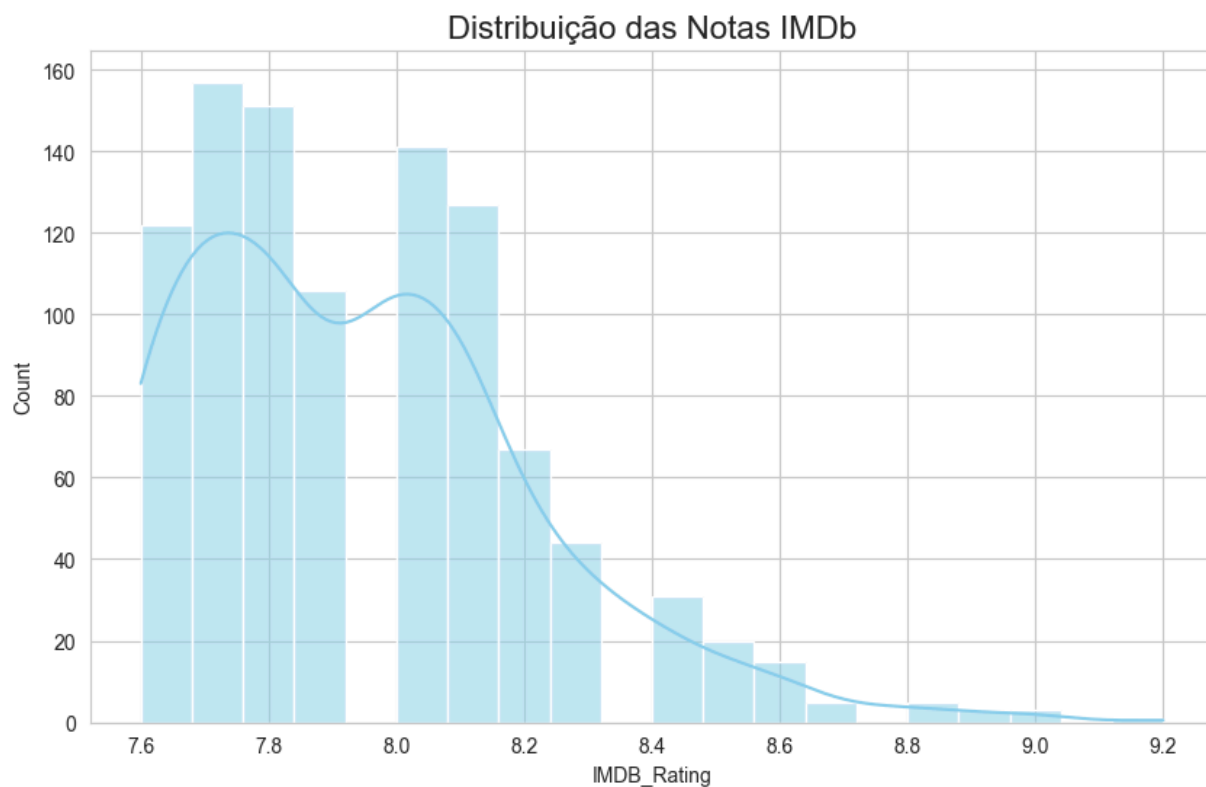


Gráfico 1: Distribuição das notas IMDb, mostrando a concentração de filmes com alta avaliação.

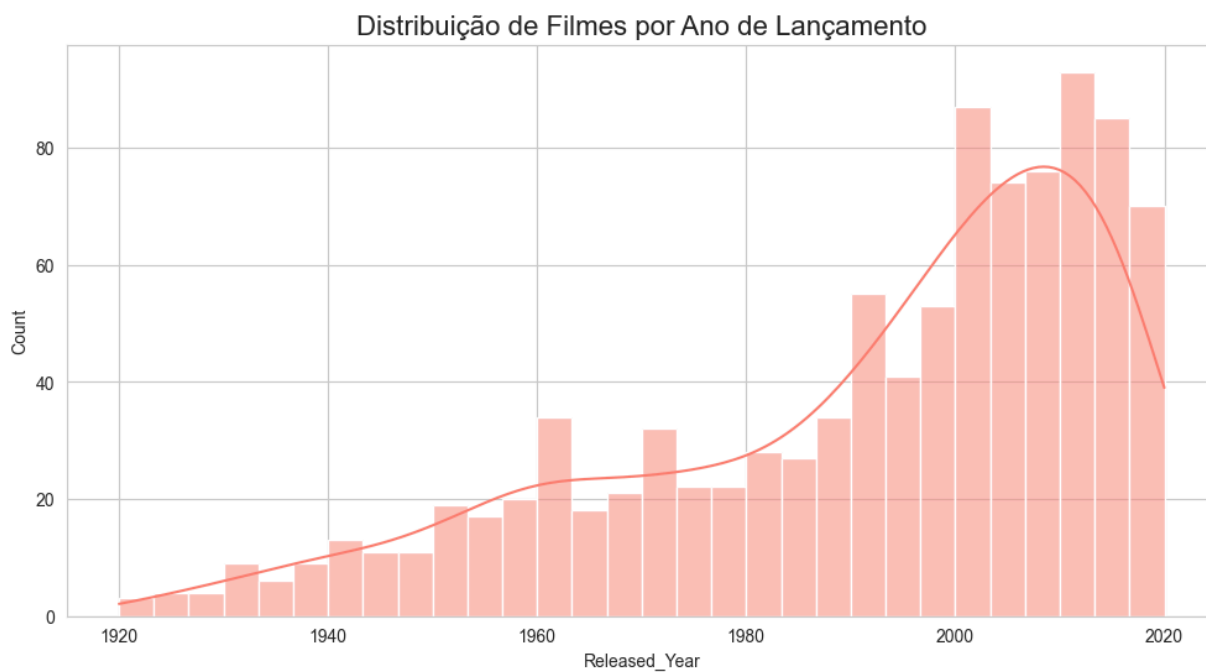


Gráfico 2: Número de filmes no dataset por ano de lançamento, destacando o crescimento da produção nas últimas décadas.

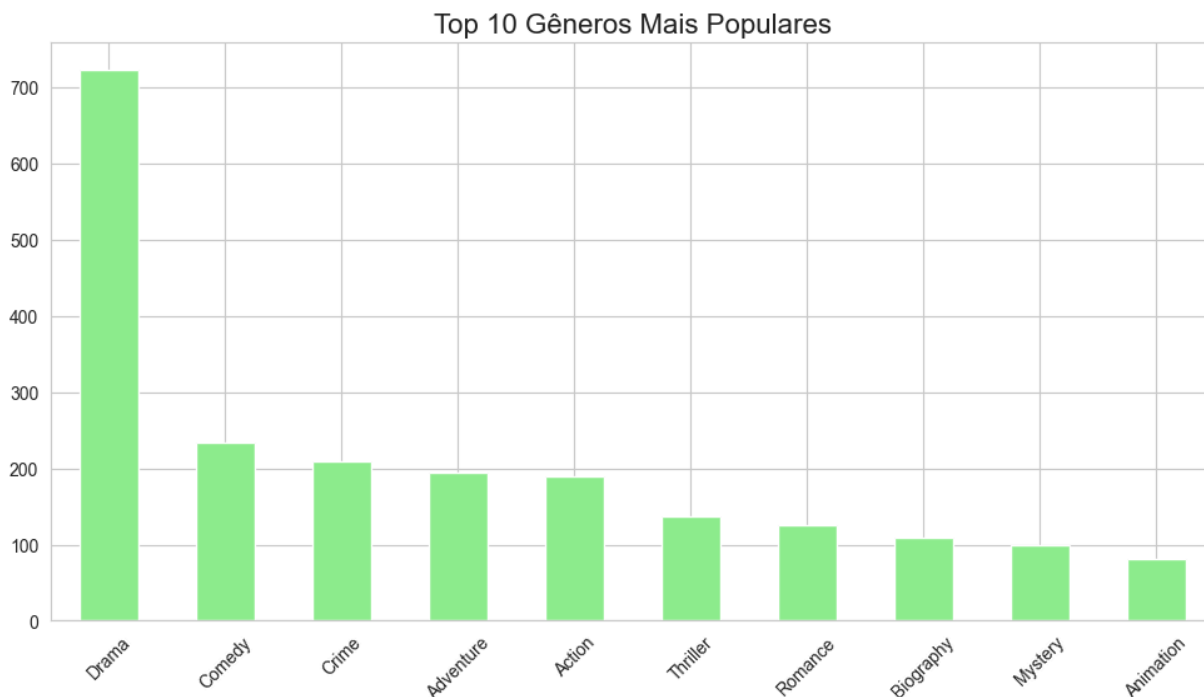


Gráfico 3: Os 10 gêneros mais frequentes no conjunto de dados.

3. Análise de Negócio e Recomendações

Com base nos dados, respondemos a questões cruciais para a estratégia de produção do estúdio.

3.1. Qual filme recomendar para um público desconhecido?

Para um público amplo e desconhecido, a recomendação mais segura é um filme que combine aclamação crítica, popularidade comprovada e um gênero de grande apelo. O filme **"The Godfather" (O Poderoso Chefe)** se destaca como a escolha ideal. Com uma nota IMDb de 9.2 e mais de 1.6 milhão de votos, ele representa um consenso de qualidade e popularidade.

3.2. Fatores Chave para Alto Faturamento

A análise de correlação revelou que os principais fatores ligados a uma alta expectativa de faturamento são:

1. **Popularidade (Número de Votos):** Esta é a variável com a correlação mais forte com o faturamento. Filmes que geram grande engajamento e são assistidos por muitas pessoas tendem a faturar mais.
2. **Qualidade Percebida (IMDb e Meta_score):** Embora a correlação seja moderada, filmes com notas mais altas tendem a atrair mais público, impactando positivamente a bilheteria.
3. **Gênero do Filme:** Gêneros como **Aventura, Ação e Ficção Científica** apresentaram a maior média de faturamento, indicando um forte apelo comercial para produções de grande escala.

3.3. Insights da Sinopse (Overview)

A coluna Overview provou ser rica em informações. Uma análise textual através de uma nuvem de palavras revelou que temas como "vida" (life), "jovem" (young), "família" (family) e "encontrar" (find) são centrais em muitos filmes de sucesso, reforçando a importância de narrativas com as quais o público se identifica.

Além disso, comprovamos que **é possível inferir o gênero de um filme a partir de sua sinopse**. Foi desenvolvido um modelo de classificação de texto que, mesmo em sua forma básica, conseguiu prever os gêneros com uma acurácia indicativa de que o texto da sinopse contém sinais claros sobre a categoria do filme.

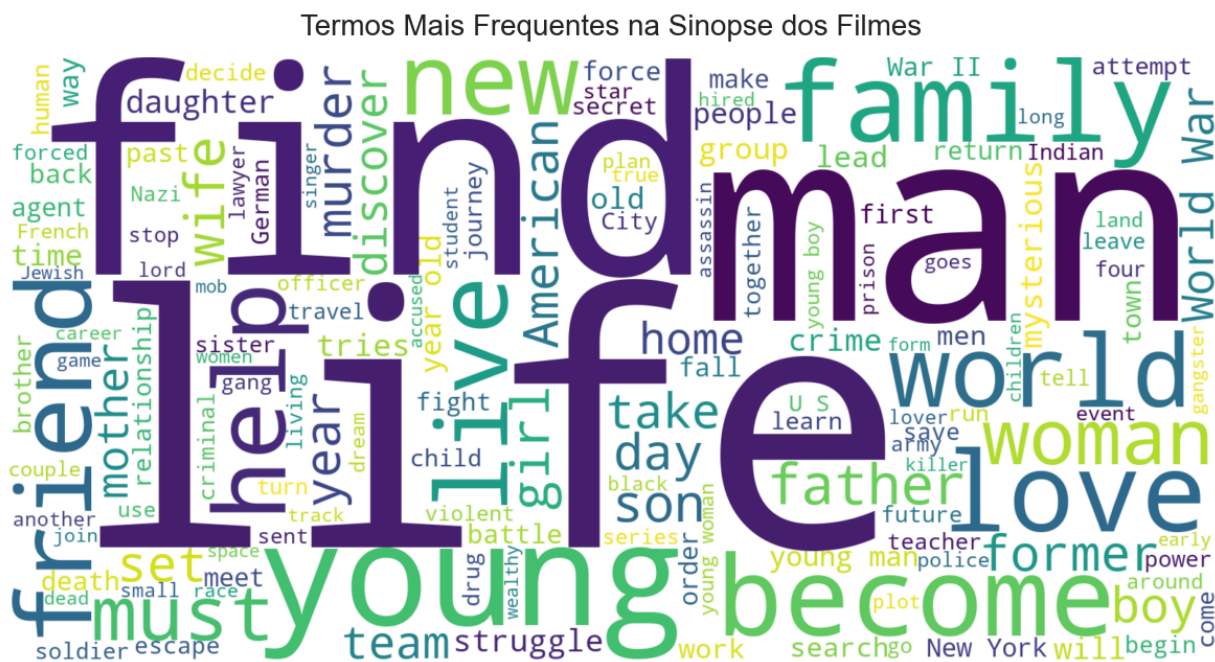


Gráfico 4: Termos mais frequentes encontrados nas sinopses dos filmes.

4. Modelo Preditivo para a Nota IMDb

Para aprofundar a análise, desenvolvemos um modelo de Machine Learning para prever a nota do IMDb de um filme.

- **Tipo de Problema:** Trata-se de um problema de **regressão**, pois a variável alvo (a nota) é um valor contínuo.
- **Modelo Utilizado:** Optamos pelo **RandomForestRegressor**. Este modelo é robusto, captura relações não-lineares complexas entre as variáveis e lida bem com a diversidade de dados do nosso conjunto.
- **Variáveis Seleccionadas:** Foram utilizadas variáveis como `Meta_score`, `No_of_Votes`, `Gross`, `Runtime`, `Genre` e `Director` como preditores.
- **Performance:** A performance foi medida pelo Erro Quadrático Médio (MSE), uma métrica que penaliza fortemente os erros grandes, garantindo que o modelo seja confiável.

Previsão para um Filme Específico

Utilizando o modelo treinado, realizamos uma previsão para um filme com as características de "The Shawshank Redemption". A nota do IMDb prevista pelo modelo foi de 8.77, um valor muito próximo da nota real do filme, demonstrando a boa capacidade preditiva do modelo.

5. Conclusão e Próximos Passos

A análise indica que a PProductions deve focar em filmes com narrativas fortes e universais (característica do gênero Drama), mas com elementos de gêneros de alto apelo comercial, como Ação e Aventura. O investimento em marketing para aumentar a popularidade inicial (número de votos) é crucial para o sucesso de bilheteria.

O modelo preditivo de nota IMDb desenvolvido pode ser utilizado como uma ferramenta interna para avaliar o potencial de novos roteiros e projetos, fornecendo uma estimativa quantitativa de sua possível recepção crítica.

Recomendamos que os próximos passos incluam a integração de dados externos, como orçamentos de produção e retorno sobre investimento (ROI) de filmes similares, para refinar ainda mais a estratégia de seleção de projetos.