

# Análise Preditiva para a Indústria Cinematográfica

## 1. Contexto do Projeto

Este projeto foi desenvolvido como solução para o Desafio de Cientista de Dados da Indicium (Lighthouse). O objetivo é atuar como uma consultoria para o estúdio de Hollywood "PProductions", utilizando uma base de dados de filmes para extrair insights e construir modelos preditivos que orientem a tomada de decisão sobre futuras produções. O projeto segue a metodologia CRISP-DM.

## 2. Resultados e Recomendações Estratégicas

A análise aprofundada do mercado de filmes de alta performance, enriquecida com dados externos e modelagem preditiva, revelou três pilares estratégicos para o sucesso, culminando em recomendações diretas para a PProductions.

### Pilar 1: A Fórmula do Blockbuster - Faturamento e Popularidade

A análise e o modelo preditivo de faturamento ( $R^2 = 0.616$ ) confirmaram a fórmula para um blockbuster de sucesso:

- Gênero é o Fator-Chave:** O gênero **Aventura** é o preditor de maior impacto para um alto faturamento.
- Engajamento é Essencial:** A popularidade e o "boca a boca" (medidos pelo `No_of_Votes`) são o segundo fator mais importante.
- O Poder dos Grandes Nomes:** Diretores e atores de renome continuam sendo cruciais para atrair o público e maximizar a bilheteria.

### Pilar 2: O Investimento Inteligente - Maximizando o Retorno (ROI)

Indo além do faturamento bruto, a análise com dados externos de orçamento revelou a estratégia para máxima **rentabilidade**:

- Alto ROI em Orçamentos Controlados:** Os filmes mais rentáveis (com retornos superiores a 100x o investimento) não são os de maior bilheteria, mas sim produções de gênero (como **Terror, Suspense e Drama**) com custos de produção eficientes. Isso representa uma oportunidade de investimento de baixo risco e altíssimo retorno.

### Pilar 3: A Ferramenta Preditiva - Previsão de Sucesso de Crítica

Para mitigar riscos criativos, desenvolvi uma ferramenta de previsão de alta precisão para a recepção do público.

- Modelo Preditivo de Nota IMDb:** o modelo (`modelo_nota_imdb.pkl`) prevê a nota final de um filme com um **erro médio de apenas 0.14 pontos** na escala de 1 a 10.
- Validação Prática:** O modelo previu com sucesso uma nota de **9.06** para "The Shawshank Redemption" (nota real: 9.3), validando sua eficácia.

### 3. Recomendações Finais para a PProductions

1. **Para Grandes Investimentos:** Foque em filmes de **Aventura/Ação** com talentos consagrados e marketing massivo para garantir o sucesso de bilheteria.
2. **Para Máxima Rentabilidade:** Diversifique o portfólio com filmes de gênero de **orçamento controlado** para maximizar o ROI.
3. **Uso da Ferramenta Interna:** Utilize o **modelo preditivo de nota IMDb** para avaliar e refinar projetos em fases iniciais de desenvolvimento.

### 4. Respostas às Perguntas do Desafio

- **Qual filme você recomendaria para uma pessoa que você não conhece?**
  - Um filme que combine alta avaliação de público e crítica com gêneros de grande apelo, como **Família**, **Aventura** e **Animação**. Um exemplo do dataset seria “Coco” ou “A Viagem de Chihiro”.
- **Quais são os principais fatores para um alto faturamento?**
  - O fator de maior impacto é o gênero **Aventura**, seguido pela **popularidade** ( `No_of_Votes` ).
- **É possível inferir o gênero a partir da `Overview` ?**
  - Sim, mas com baixa confiabilidade. A análise mostrou que a sinopse sozinha não é um bom preditor quantitativo dos múltiplos gêneros de um filme.
- **Como foi feita a previsão da `nota do IMDB` ?**
  - Tratei como um problema de **Regressão**. O modelo final, um `XGBoost Regressor` otimizado, utilizou todas as features numéricas e os gêneros. A métrica principal de avaliação foi o **MAE (Mean Absolute Error)**, por sua fácil interpretação. O modelo final obteve um `MAE` de **0.137**.
- **Supondo um filme com as características de “The Shawshank Redemption”, qual seria a nota do IMDB?**
  - Utilizando o modelo final e otimizado, a nota prevista foi de 9.06.

### Bônus: Aplicação Interativa para Análise de Dados

Para ir além no desafio e proporcionar uma forma interativa de explorar os dados e insights deste projeto, desenvolvi e coloquei no ar uma aplicação web com Streamlit.

➔ **Acesse a aplicação ao vivo no link abaixo:**

<https://explorador-filmes-rafael-dutra.streamlit.app/>

OBS: A aplicação é hospedada em um serviço gratuito e pode entrar em repouso. Se isso acontecer, clique no botão “Yes, get this app back up!” que surgirá na tela e aguarde alguns segundos para o carregamento completo.