

# Relatório de Análise Cinematográfica para PProductions

Gustavo Dias de Oliveira

September 2, 2025

# Contents

<b>1</b>	<b>Análise Exploratória dos Dados (EDA)</b>	<b>3</b>
1.1	Gêneros . . . . .	3
1.2	Faixa Etária . . . . .	3
1.3	Correlações . . . . .	3
1.4	Duração de Filmes . . . . .	3
1.5	Elenco de Sucesso . . . . .	4
<b>2</b>	<b>Respostas às Perguntas</b>	<b>4</b>
2.1	Qual filme você recomendaria para uma pessoa que você não conhece? . . . . .	4
2.2	Quais são os principais fatores que estão relacionados com alta expectativa de faturamento de um filme? . . . . .	4
2.3	Quais insights podem ser tirados com a coluna Overview? É possível inferir o gênero do filme a partir dessa coluna? . . . . .	4
<b>3</b>	<b>Previsão da Nota do IMDB</b>	<b>5</b>
3.1	Metodologia . . . . .	5
3.2	Previsão para 'The Shawshank Redemption'	5
<b>4</b>	<b>Recomendação Final</b>	<b>5</b>

# 1 Análise Exploratória dos Dados (EDA)

A análise inicial do conjunto de dados revelou características importantes para orientar a produção de novos filmes. O conjunto de dados é predominantemente composto por filmes recentes, lançados entre 1980 e 2020. A duração média dos filmes é de aproximadamente 124 minutos (cerca de 2 horas), com uma variação significativa, incluindo alguns filmes com mais de 3 horas de duração. Uma observação notável é que a maioria dos filmes no conjunto de dados possui avaliações altas no IMDb (entre 7.6 e 9.2) e no MetaScore (acima de 72), indicando que a base de dados consiste em produções aclamadas. A distribuição de votos e faturamento mostra uma concentração de valores mais baixos, com poucos filmes atingindo patamares de sucesso extremamente altos, o que sugere a existência de produções "blockbuster" de grande notoriedade.

## 1.1 Gêneros

O gênero **Drama** é o mais prevalente nos filmes analisados. No entanto, a análise de faturamento revela que os filmes de **Ação** e **Aventura** dominam o topo das bilheterias. Essa dualidade aponta para uma oportunidade: a combinação de um drama bem elaborado com elementos de ação e aventura pode ser uma fórmula de sucesso, unindo alta avaliação da crítica com grande retorno financeiro.

## 1.2 Faixa Etária

A categorização por faixa etária mostrou que:

- Filmes para **Adolescentes** tendem a ter o maior faturamento e um número superior de votos.
- Filmes com classificação **Livre** possuem, em média, pontuações de IMDb e MetaScore ligeiramente mais altas.

A hipótese é que um filme voltado para o público adolescente, com uma trama dramática, ação envolvente e aventura, poderia capitalizar tanto na popularidade quanto na aprovação crítica.

## 1.3 Correlações

A matriz de correlação indicou uma relação positiva moderada entre o número de votos, as avaliações e o faturamento, o que sugere que a popularidade de um filme pode ser um bom preditor de seu sucesso financeiro e de crítica.

- $R(\text{No\_of\_Votes}, \text{IMDB\_Rating}) = 0.5$  (correlação moderada).
- $R(\text{No\_of\_Votes}, \text{Gross}) = 0.6$  (correlação um pouco mais forte).

## 1.4 Duração de Filmes

A média da duração de filmes de maior faturamento, assim como de filmes com as maiores notas do IMDb e MetaScore, fica em torno de **155 minutos**. Esse dado é crucial para o planejamento da produção.

## 1.5 Elenco de Sucesso

Para maximizar as chances de sucesso, é crucial contratar diretores e um elenco com histórico comprovado de sucesso, como os listados abaixo.

- **Diretores de Destaque:**

- Lana Wachowski e Irvin Kershner (maiores notas de IMDb)
- Anthony Russo e Gareth Edwards (maiores faturamentos)

- **Atores de Destaque:**

- Elijah Wood, John Travolta e Lilly Wachowski (maiores notas de IMDb)
- Michael Peña, Mirai Shida e Maria Doyle Kennedy (maiores faturamentos)

## 2 Respostas às Perguntas

### 2.1 Qual filme você recomendaria para uma pessoa que você não conhece?

Baseando-se em critérios de alta aprovação crítica (IMDb e MetaScore), eu recomendaria **"The Godfather"**. Este filme ocupa o topo em ambas as métricas, com uma nota de 9.2 no IMDb e 100.0 no MetaScore, tornando-o uma aposta segura para uma experiência cinematográfica de alta qualidade.

### 2.2 Quais são os principais fatores que estão relacionados com alta expectativa de faturamento de um filme?

Os principais fatores são:

- **Gêneros Populares:** Filmes de Ação, Aventura, e Drama têm maior probabilidade de sucesso em bilheteria.
- **Qualidade de produção:** Filmes com MetaScore elevado tendem a ter um desempenho financeiro superior, como evidenciado na matriz de correlação.
- **Elenco e Diretores Renomados:** Atores e diretores com histórico de sucesso aumentam a visibilidade e a confiança do público.
- **Público-Alvo:** A análise por faixa etária sugere que filmes para o público adolescente têm maior potencial de faturamento.

### 2.3 Quais insights podem ser tirados com a coluna Overview? É possível inferir o gênero do filme a partir dessa coluna?

A coluna *Overview* contém os resumos dos filmes, e pode ser usada para:

- **Análise de Sentimento:** Identificar padrões emocionais nas sinopses de filmes de sucesso.
- **Extração de Palavras-Chave:** Descobrir quais temas e narrativas mais se conectam com o sucesso comercial e de crítica.

- **Previsão de Gênero:** É sim possível inferir o gênero do filme a partir desta coluna utilizando técnicas de *Processamento de Linguagem Natural* (NLP). Modelos de classificação de texto poderiam ser treinados para essa tarefa.

## 3 Previsão da Nota do IMDB

### 3.1 Metodologia

O problema de prever a nota do IMDb é tratado como um problema de **regressão**, já que a nota é uma variável contínua. Para a modelagem, as variáveis selecionadas foram:

- *Meta\_score*: Alta correlação com a nota IMDb.
- *Runtime*: Duração do filme.
- *No\_of\_Votes*: Número de votos, que reflete a popularidade.
- *Gross*: Faturamento, que também se correlaciona com a popularidade e, por tabela, com a avaliação.
- *Genre*: Gênero do filme, transformado em variáveis *dummies* (one-hot encoding) para que o modelo consiga interpretar essa informação categórica.

O modelo escolhido foi o **Random Forest Regressor** devido à sua robustez, capacidade de lidar com dados numéricos e categóricos, e por capturar relações não lineares de forma eficiente. O modelo foi avaliado com a métrica **RMSE** (*Root Mean Squared Error*), que mede a diferença média entre os valores previstos e os reais na mesma escala do IMDb, sendo de fácil interpretação. O RMSE de 0.22 obtido indica que o modelo tem um bom nível de precisão em suas previsões.

### 3.2 Previsão para 'The Shawshank Redemption'

Com base nas características do filme, o modelo previu a seguinte nota:

**Nota IMDB prevista:** 8.76

Este valor está bem próximo da nota real do filme (9.3), o que demonstra que o modelo, mesmo com as limitações de tempo para otimização, é confiável para prever a avaliação do IMDb com base nas características de entrada.

## 4 Recomendação Final

Considerando todos os fatores, a PProductions deve focar em um filme que combine os gêneros **Drama** e **Ação/Aventura**. A duração ideal para um filme que seja sucesso de público e bem avaliado é de aproximadamente **155 minutos**. O filme deve ser destinado ao público **adolescente**.

Para maximizar as chances de sucesso, é crucial contratar diretores e um elenco com histórico comprovado de sucesso.