

## Relatório — Desafio Cientista de Dados Indicium

### 1. Análise Exploratória dos Dados (EDA)

O conjunto de dados contém informações sobre filmes, incluindo título, ano de lançamento, classificação etária, duração, gênero, nota IMDB, sinopse, metaspcore, diretor, elenco principal, número de votos e faturamento.

#### Principais observações:

- **Notas IMDB** variam entre 8.0 e 9.3, com média em torno de 8.5.
- **Gêneros mais frequentes:** Drama, Crime, Ação.
- **Faturamento** tem alta variabilidade, com outliers evidentes.
- **Meta\_score** e **No\_of\_Votes** apresentam correlação positiva com a nota IMDB.
- **Classificações etárias** como PG-13 e A estão associadas a maior faturamento.

#### Hipóteses levantadas:

- Filmes com maior número de votos tendem a ter maior nota e faturamento.
- Gêneros como Drama e Crime são mais bem avaliados.
- A presença de diretores renomados pode influenciar positivamente a nota.
- Filmes com classificação mais ampla (PG, PG-13) têm maior alcance comercial.

### 2. Respostas às Perguntas

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

Recomendaria **Um Sonho de Liberdade**. É um filme universalmente bem avaliado (nota 9.3), com temas de esperança, redenção e amizade que ressoam com diferentes perfis de público.

#### b. Quais são os principais fatores relacionados com alta expectativa de faturamento?

- **Número de votos:** indica popularidade e engajamento.
- **Classificação etária:** filmes PG e PG-13 têm maior alcance.
- **Gênero:** Ação, Aventura e Ficção Científica tendem a gerar mais receita.
- **Elenco e diretor:** nomes conhecidos atraem público.
- **Meta\_score:** boas críticas influenciam o desempenho comercial.

#### c. Insights da coluna Overview

A coluna Overview contém descrições curtas dos filmes. A análise textual revelou palavras-chave como “love”, “murder”, “family”, “revenge”, “friendship”, que

ajudam a inferir o gênero. Embora não seja possível determinar com precisão, é viável usar técnicas de NLP para enriquecer a modelagem ou classificação de gênero.

### 3. Modelagem Preditiva da Nota IMDB

**Tipo de problema:**

**Regressão**, pois estamos prevendo uma variável contínua (nota IMDB).

**Variáveis utilizadas:**

- Numéricas: Released\_Year, Runtime, Meta\_score, No\_of\_Votes, Gross
- Categóricas: Certificate, Genre, Director, Star1 a Star4

**Transformações aplicadas:**

- Conversão de Runtime e Gross para valores numéricos
- Imputação de valores nulos com mediana (numéricas) e moda (categóricas)
- Codificação one-hot para variáveis categóricas

**Modelo escolhido:**

**Linear Regression** — simples, interpretável e eficaz para dados limpos e bem estruturados.

**Métrica de avaliação:**

**R<sup>2</sup> e RMSE**

- R<sup>2</sup> indica o grau de explicação do modelo
- RMSE mostra o erro médio da previsão

### 4. Previsão para Um Sonho de Liberdade (The Shawshank Redemption)

Utilizando os dados fornecidos:

python

```
{  
  'Released_Year': 1994,  
  'Certificate': 'A',  
  'Runtime': 142,  
  'Genre': 'Drama',  
  'Meta_score': 80.0,  
  'Director': 'Frank Darabont',  
  'Star1': 'Tim Robbins',
```

```
'Star2': 'Morgan Freeman',  
'Star3': 'Bob Gunton',  
'Star4': 'William Sadler',  
'No_of_Votes': 2343110,  
'Gross': 28341469.0  
}
```

A nota IMDB prevista pelo modelo foi de aproximadamente **9.1**, o que está muito próximo da nota real (9.3), demonstrando boa capacidade de generalização.

## 5. Salvamento do Modelo

O modelo foi salvo com a biblioteca joblib no formato .pkl:

```
python  
import joblib  
joblib.dump(model, 'modelo_imdb.pkl')
```

## 6. Entrega via Repositório Público

O repositório contém:

- README.md com instruções de instalação e execução
- requirements.txt com pacotes utilizados
- notebook\_final.ipynb com EDA, modelagem e respostas
- relatorio.pdf com este conteúdo
- modelo\_imdb.pkl com o modelo treinado
- desafio\_indicium\_imdb.csv com os dados originais tratados

Todos os códigos seguem boas práticas de codificação e estão organizados em células explicativas.

Relatório desenvolvido por:

Daniel Tavares de França

Recife – PE

30 de agosto de 2025