
Contextualização:

Você foi contratada(o) como cientista de dados júnior por uma empresa de análise de performance de streaming chamada **MovieScope**. Sua missão é analisar dados de filmes disponíveis em plataformas digitais, com o objetivo de gerar **insights sobre padrões de sucesso, características que influenciam a nota dos filmes e criar uma previsão de avaliação com base em dados históricos**.

Dataset Utilizado

✦ **Dataset:** [IMDb Movies Dataset - TMDb 5000 Movie Dataset](#)

Este conjunto de dados contém informações sobre milhares de filmes, incluindo:

- Nome, diretor, elenco
Receita, orçamento, gênero
- Avaliações (nota média)
- Idioma original
Data de lançamento, popularidade
- e mais.

Desafio:

Exploração Inicial

- Importar e visualizar o dataset com **pandas**.
- Corrigir problemas de codificação, valores ausentes e inconsistências.
Analisar variáveis categóricas e numéricas.

Estatística Descritiva

- Calcular média, mediana, desvio padrão e variância de atributos como **orçamento, receita, nota média, popularidade**.
- Criar uma análise interpretativa desses valores.

Visualização de Dados

- Criar visualizações com **matplotlib** e **seaborn**:
-

-
- Histogramas (ex: distribuição das notas)
 - Gráficos de dispersão (ex: orçamento vs. receita)
 - Boxplots (ex: avaliação por gênero)
 - Gráficos de barras (ex: número de filmes por idioma)

Modelagem Preditiva

- Criar um modelo de **regressão linear** para prever a **nota média (vote_average)** com base em variáveis como:
 - orçamento
 - receita
 - popularidade
 - número de votos
- Avaliar o modelo com métricas apropriadas: MSE, RMSE, R^2 .

Apresentação Profissional

Elaborar uma apresentação final que conte a história dos dados:

Cenário → Perguntas → Análises → Modelos → Conclusões → Sugestões de negócio

Resultados esperados:

1. Notebook no Google Colab com análise, gráficos e modelos.
2. Apresentação final com storytelling.

Anexos (Figuras, esquemas, desenhos, leiaute, formulários, etc):
