O presente relatório tem como objetivo realizar uma análise exploratória dos dados cinematográficos. A análise visa entender as principais características dos filmes, explorar relações entre variáveis e extrair insights que possam orientar decisões estratégicas na produção de novos filmes.

O relatório EDA realizado no Notebook descreve todas as análises iniciais para o resumo deste documento.

Link:<https://github.com/Raul-Lemelle/lighthouse_desafio_ciencia_dados/blob/main/notebooks/eda.ipynb>

## **Métodos Utilizados**

Para realizar a análise exploratória dos dados, foram utilizadas técnicas estatísticas e visualizações gráficas.

* Visão Geral dos Dados:Database do tipo “csv”, conforme descrito abaixo:

**desafio\_indicium\_imdb.csv**

Colunas:

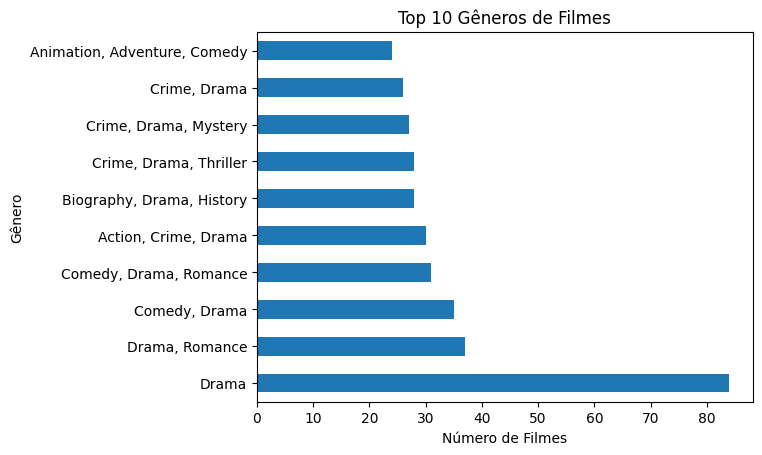
* Series\_Title => Nome do filme
* Released\_Year => Ano de lançamento
* Certificate => Classificação etária
* Runtime => Tempo de duração
* Genre => Gênero
* IMDB\_Rating => Nota do IMDB
* Overview => Overview do filme
* Meta\_score => Média ponderada de todas as críticas
* Director => Diretor
* Star1 => Ator/atriz #1
* Star2 => Ator/atriz #2
* Star3 => Ator/atriz #3
* Star4 => Ator/atriz #4
* No\_of\_Votes => Número de votos
* Gross => Faturamento
* A Análise Exploratória de Dados foi conduzida utilizando a linguagem de programação **Python**, utilizando bibliotecas como **Pandas**, **NumPy** e **Matplotlib** para análise e visualização dos dados.
* Identificação de outliers e tratamento de dados ausentes quando necessário.

## 

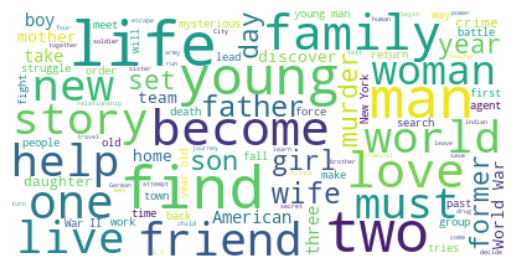
## **Conclusões**

A análise exploratória dos dados revelou diversas tendências e padrões importantes, e com base nos resultados obtidos, recomenda-se considerar:

* Investir em filmes dos gêneros mais populares como Drama, Romance e Comédia.



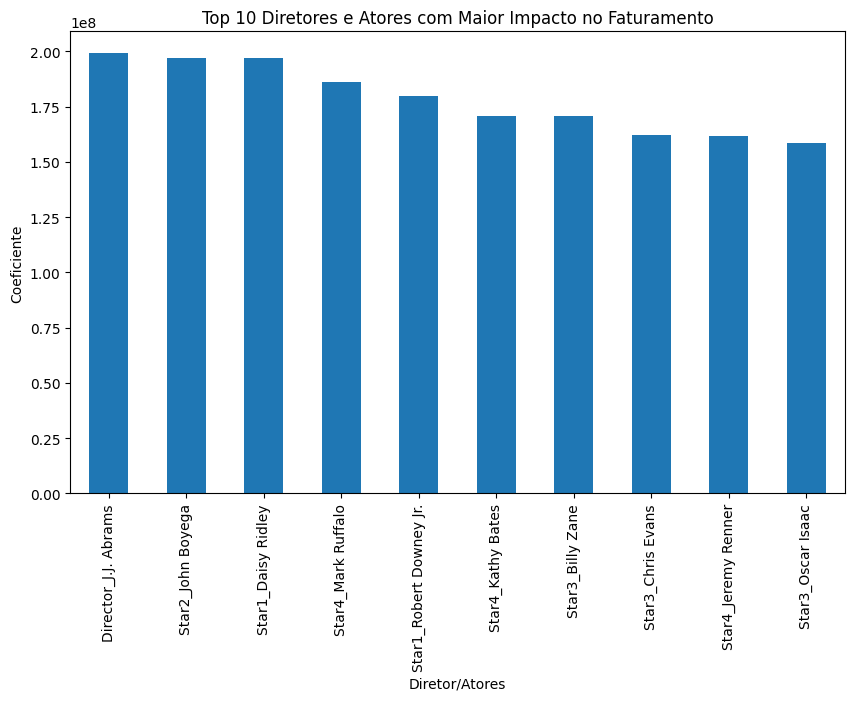
* **Coluna Overview**: Processamento de linguagem natural (NLP) para analisar a coluna. A nuvem de palavras gerada destaca temas como "amor", "família", "vida", indicando que são frequentemente explorados nos resumos dos filmes. Recomenda-se explorar filmes que enfatizem temas de família e mistério (drama), que parecem ser populares entre o público.



* **Fatores de Faturamento**:

Link:<https://github.com/Raul-Lemelle/lighthouse_desafio_ciencia_dados/blob/main/notebooks/analysis_gross.ipynb>

Foi realizado um modelo de Regressão Linear para inferir dos coeficientes quais diretores e atores têm maior impacto no faturamento. Diretores e atores com coeficientes maiores têm um impacto mais significativo.



Os coeficientes listados mostram quanto cada diretor e ator específico contribui para o faturamento previsto do filme, mantendo todas as outras variáveis constantes. Alguns insights específicos incluem:

- Director\_J.J. Abrams: Este diretor tem o maior coeficiente positivo, indicando que filmes dirigidos por J.J. Abrams tendem a ter um impacto significativo positivo no faturamento.

- Star2\_John Boyega e Star1\_Daisy Ridley: Os atores John Boyega e Daisy Ridley também têm coeficientes positivos elevados, sugerindo que filmes estrelados por eles estão associados a um aumento significativo no faturamento.

- Star4\_Mark Ruffalo, Star1\_Robert Downey Jr., Star4\_Kathy Bates, Star3\_Billy Zane, Star3\_Chris Evans, Star4\_Jeremy Renner, Star3\_Oscar Isaac: Outros atores também mostram coeficientes positivos, indicando que suas presenças nos filmes estão correlacionadas com um faturamento mais alto.

* **Previsão da Nota do IMDb**

Relatório Previsão da Nota do IMDb.

Link: IMDb - Relatório Previsão da Nota do IMDb

IMDb - Relatório Final das Análises

Link: IMDb - Relatório Final das Análises