

## Relatório de Análise de Dados de Filmes

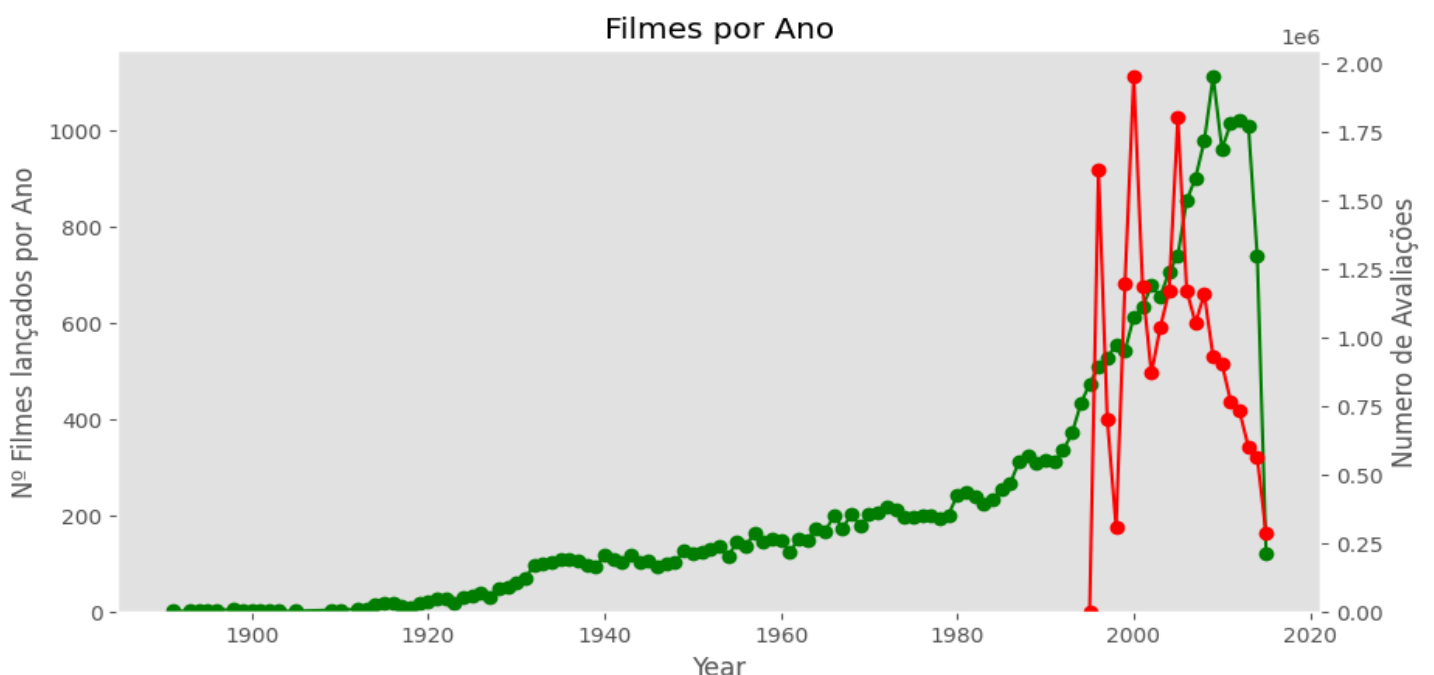
### 1. Introdução

Este relatório apresenta uma análise exploratória de dados (EDA) de um conjunto de dados contendo informações sobre filmes e avaliações de usuários. O objetivo principal é extrair insights relevantes sobre o comportamento dos usuários e tendências do mercado cinematográfico utilizando técnicas de visualização de dados e conceitos de Big Data.

### 2. Análise Exploratória de Dados (EDA) e Visualização de Dados

A análise foi iniciada com a exploração dos dados brutos, incluindo a identificação de valores ausentes e a limpeza do dataset. Em seguida, foram gerados gráficos para visualizar os dados e identificar padrões.

Gráfico 1: Filmes por Ano e Número de Avaliações:



Observa-se um aumento significativo no número de filmes lançados ao longo dos anos, com um pico por volta de 2010 e uma queda em 2020.

O número de avaliações acompanha a tendência de lançamento de filmes, com um pico no mesmo período e uma queda acentuada após 2015.

**Insight:** O mercado cinematográfico experimentou um crescimento expressivo até a década de 2010, seguido por uma possível retração nos últimos anos.

Gráfico 2: Total de Filmes por Gênero:

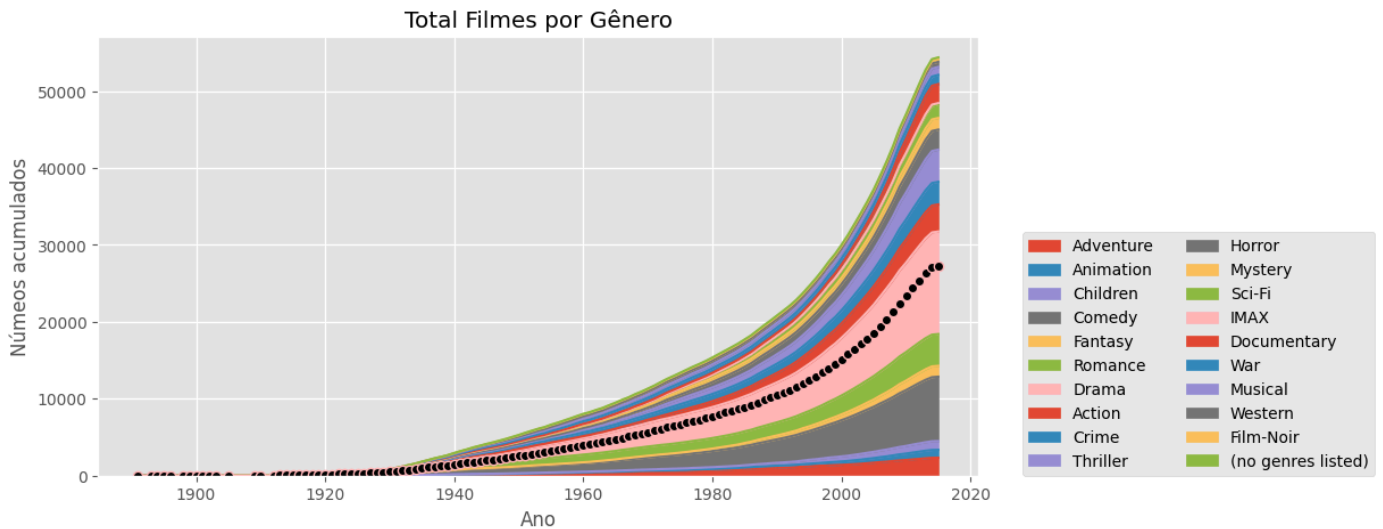


Gráfico 3: Avaliações:

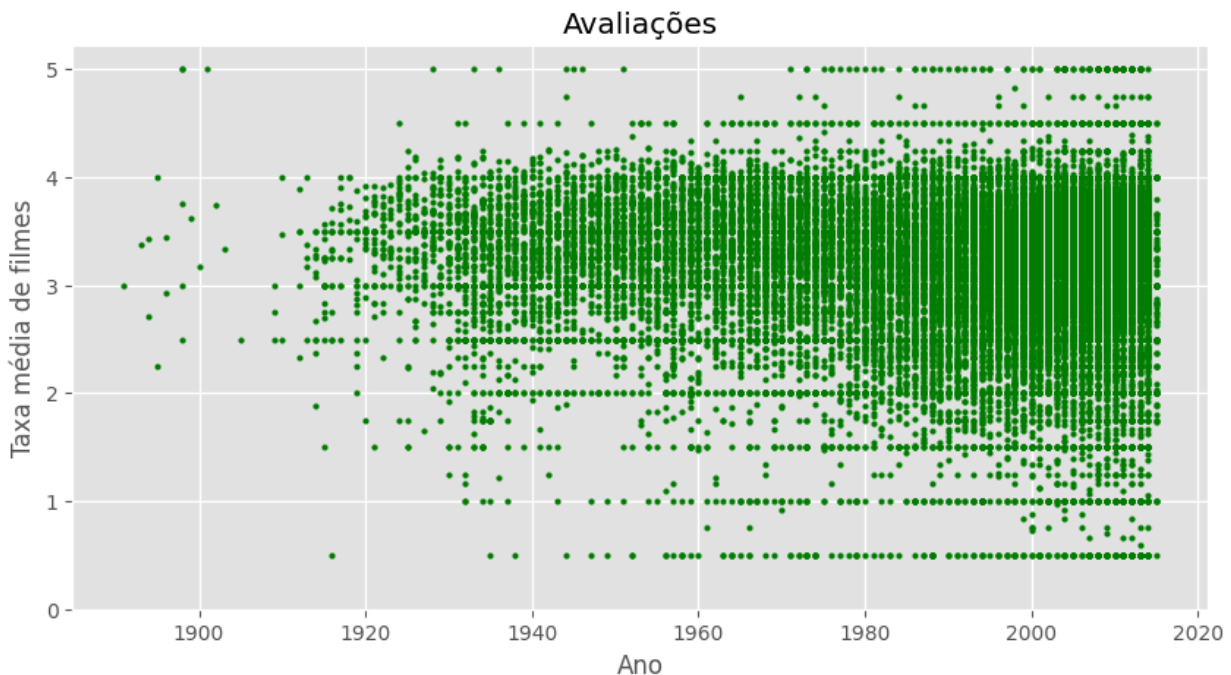
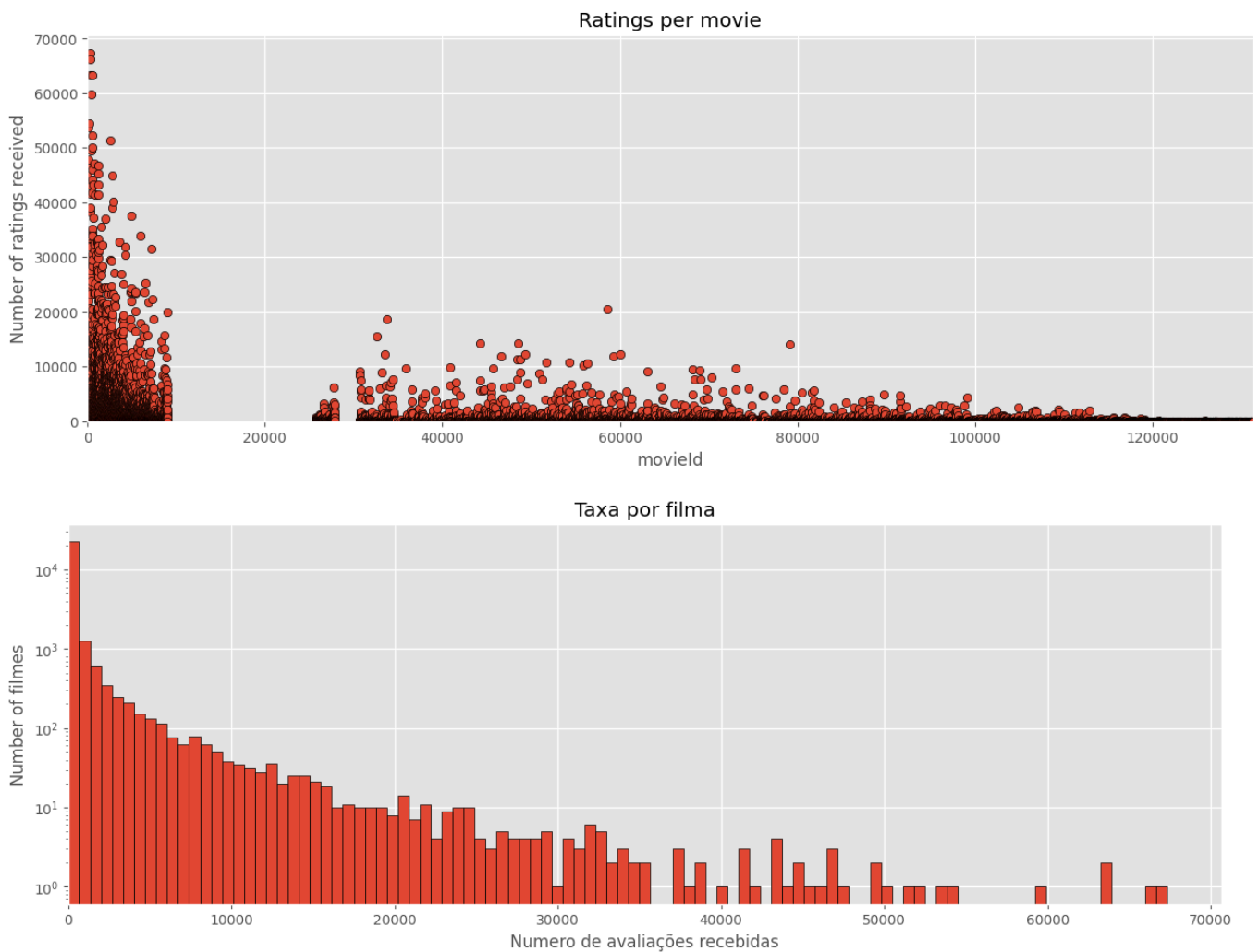


Gráfico 4 e 5: Classificações por filme e Taxa por filme:



Estes gráficos analisam a distribuição das avaliações por filme.

Observa-se que a maioria dos filmes recebe um número relativamente baixo de avaliações, enquanto alguns filmes populares recebem uma quantidade muito maior.

**Insight:** Existe uma concentração de atenção em filmes populares, enquanto muitos filmes menos conhecidos recebem pouca visibilidade.

**Sugestões de melhorias no sistema de recomendação, com base nas preferências dos usuários e no desempenho dos filmes.**

## **1. Melhorias no Sistema de Recomendação**

### **A. Uso de Modelos Híbridos de Recomendação**

Combinar diferentes abordagens para aumentar a precisão das recomendações:

- **Filtragem Colaborativa:** Recomendar filmes com base em similaridade de preferências entre usuários.
- **Filtragem Baseada em Conteúdo:** Recomendar filmes com base em características similares (gênero, elenco, diretor, etc.).
- **Modelos de Aprendizado Profundo:** Utilizar redes neurais para capturar padrões complexos nos dados de avaliações.

### **B. Implementação de Técnicas Avançadas de Machine Learning**

- **Factorization Machines:** Melhoram a capacidade de capturar interações entre usuários e itens.
- **Graph Neural Networks:** Usar redes neurais baseadas em grafos para entender conexões complexas entre filmes e espectadores.
- **Reinforcement Learning:** Para atualizar recomendações com base no feedback contínuo dos usuários.

### **C. Personalização Dinâmica das Recomendações**

- Criar **perfis adaptativos** que ajustam recomendações conforme o usuário assiste e avalia mais filmes.
- Implementar **sistemas de recomendação sensíveis ao contexto**, considerando fatores como horário e localização do usuário.

---

## **2. Análise das Preferências dos Usuários**

### **A. Clustering para Segmentação de Usuários**

Aplicar técnicas como **K-Means** ou **DBSCAN** para identificar perfis de espectadores com base em padrões de consumo, como:

- Fãs de um gênero específico (ex: terror, ficção científica).

- Usuários que preferem filmes mais antigos vs. mais recentes.
- Usuários que assistem e avaliam filmes de forma recorrente.

### **B. Análise de Sentimento em Comentários e Avaliações**

Utilizar **Processamento de Linguagem Natural (NLP)** para interpretar avaliações textuais e entender melhor o que agrada ou desagrada os usuários.

### **C. Análise de Tempo de Engajamento**

Avaliar **tempo médio assistido por filme** para diferenciar usuários que terminam os filmes daqueles que abandonam no meio.

---

## **3. Avaliação do Desempenho dos Filmes**

### **A. Modelagem de Tendências ao Longo do Tempo**

- Aplicar **modelos preditivos** (ARIMA, Prophet) para prever a popularidade futura de um filme.
- Analisar **variações sazonais** para identificar quando determinados gêneros são mais assistidos.

### **B. Detecção de Filmes Subestimados**

Usar **análise de dispersão** para encontrar filmes bem avaliados, mas com poucas visualizações, que poderiam ser promovidos.

### **C. Avaliação Comparativa por Gênero e Diretores**

- Criar **dashboards interativos** para comparar a performance de filmes por gênero e diretor ao longo do tempo.
- Identificar padrões como diretores com histórico consistente de boas avaliações.

## Conclusão

A análise dos dados permitiu identificar padrões importantes no mercado cinematográfico, como a evolução no lançamento de filmes, preferências de gênero e a tendência de avaliações ao longo do tempo. Esses insights são essenciais para a melhoria dos sistemas de recomendação, pois possibilitam a implementação de abordagens mais eficazes, como:

- Uso de modelos híbridos de recomendação que combinam filtragem colaborativa, baseada em conteúdo e aprendizado profundo.
- Implementação de técnicas avançadas de machine learning, como Factorization Machines e Graph Neural Networks.
- Personalização dinâmica das recomendações com base no comportamento do usuário.
- Análise de sentimentos e clustering para segmentação de audiências.

Com esses aprimoramentos, é possível oferecer recomendações mais precisas e relevantes para os usuários, melhorando sua experiência e aumentando o engajamento com a plataforma de recomendação de filmes.

### Ferramentas utilizadas:

**Meu Github:** <https://github.com/Yalle5481/MEUS-PROJETOS>

**Notebook do meu Google Colab, onde executei os códigos e gerei os gráficos:**  
<https://colab.research.google.com/drive/1V59vfCyLZUKDAFIktib8djmqsGjLuSB?usp=sharing>

**Meu Kaggle:** <https://www.kaggle.com/yallefinanceiro>