

# Projeto BeeView

## 1. Introdução

**Nome do Projeto:** BeeView

**Integrantes do Grupo:**

- Christian Oliveira Jorge Moreira - 202408723946
- Rafaela de Jesus Soares - 202403610876
- Uiles dos Santos - 201951068491
- Maria Eduarda Dias Alves da Silva - 202408371446
- Antonio Carlos Sena da Conceição Junior - 202102120748

**Objetivo do Projeto:**

Este projeto de análise de dados se propõe a investigar o universo cinematográfico sob duas perspectivas principais:

1. **Análise Estratégica de Franquias:** Responder à pergunta central "Vale a pena fazer uma sequência no mercado cinematográfico?", identificando padrões de sucesso através de métricas financeiras e de recepção pública, e realizando uma análise comparativa entre os mercados do Brasil e dos EUA.
2. **Análise de Desempenho e Competitividade:** Avaliar o desempenho e os desafios do cinema brasileiro na categoria de Melhor Filme Internacional do Oscar, compreendendo os fatores que contribuem para o reconhecimento artístico em contraste com os obstáculos para a vitória final.

**Justificativa da Escolha do Tema:**

A seleção do tema "Filmes" para este projeto se justifica pela sua notável relevância cultural, social e econômica no cenário contemporâneo. O cinema, enquanto forma de arte e indústria, possui uma capacidade singular de refletir e, ao mesmo tempo, influenciar o imaginário coletivo. Essa ubiquidade e o impacto dos filmes na vida das pessoas os tornam um objeto de estudo rico e dinâmico. Adicionalmente, o interesse pessoal e a paixão pela sétima arte foram fatores motivadores para a exploração aprofundada desse universo, com o objetivo de extrair insights valiosos por meio da análise de dados.

**Público-Alvo:**

Esta documentação é destinada a outros analistas de dados, desenvolvedores, estudantes e stakeholders da indústria cinematográfica interessados em análise de mercado, tendências e padrões de sucesso no cinema.

## 2. Fonte de Dados

### Origem dos Dados:

Para a execução deste trabalho, foram empregadas diversas bases de dados, escolhidas pela sua abrangência e riqueza de informações:

- **The Movie Database (TMDB):** Vasta base de dados colaborativa, acessada via API oficial.
- **IMDb Datasets:** Dados de filmes, elencos e equipes do IMDb, adquiridos por download direto.
- **Rotten Tomatoes:** Informações de críticas de especialistas e do público, coletadas via web scraping.
- **Kaggle - Full TMDB Movies Dataset 2024 (1M Movies):** Dataset abrangente que consolidou informações de mais de um milhão de filmes do TMDB, servindo como a principal fonte de dados quantitativos.
- **AdoroCinema, Wikipedia, Papo de Cinema, G1, Mídia Internacional:** Fontes consultadas para dados qualitativos, contexto histórico, críticas e informações sobre premiações.

### Desafios dos Dados:

Durante o projeto, alguns desafios foram encontrados e tratados:

- **Dados Ausentes:** Muitos filmes, especialmente no dataset do Kaggle, possuíam valores zerados ou ausentes para **budget** (orçamento) e **revenue** (receita). Isso exigiu o uso da API do TMDB para obter dados mais precisos e confiáveis.
- **Codificação de Arquivos:** Ao carregar arquivos CSV exportados do Excel, foi encontrado um **UnicodeDecodeError**, que exigiu testar diferentes codificações (como **latin1**) para a correta leitura dos dados.
- **Consistência de Franquias:** A identificação de todos os filmes pertencentes a uma mesma franquia exigiu uma combinação de buscas por título e verificação via API, pois o dataset original nem sempre agrupava as coleções de forma completa.

## 3. Metodologia de Análise

### Ferramentas e Tecnologias Aplicadas:

- **Linguagem de Programação:** Python (versão 3.9)
- **Bibliotecas Python:** Pandas, NumPy, Matplotlib, Seaborn, Requests.
- **Ambiente de Desenvolvimento:** Google Colab
- **Versionamento e Hospedagem:** Git e GitHub

### Etapas do Processo:

A metodologia seguiu um pipeline de dados estruturado, desde a coleta até a apresentação dos resultados.

1. **Coleta e Aquisição de Dados:** Utilização da API do TMDb, download de datasets do IMDb e Kaggle, e coleta pontual de fontes complementares.
2. **Pré-processamento e Limpeza de Dados:** Esta etapa incluiu o tratamento de valores ausentes, remoção de duplicatas, padronização de formatos, e a consolidação de bases para criar um dataset unificado e de alta qualidade.
3. **Análise Exploratória de Dados (EDA):** A EDA foi fundamental para compreender a estrutura e os padrões nos dados, utilizando estatísticas descritivas e visualizações.
4. **Engenharia de Features:** Criação de novas variáveis para enriquecer a análise, como a métrica **ROI** (Retorno Sobre Investimento), a classificação de filmes como **Original** ou **Sequência**, e a definição de um critério de "Sucesso Numérico". Para a análise do Oscar, foram criadas métricas qualitativas.
5. **Análise e Modelagem:** Foram utilizadas análises de correlação e testes de hipótese para comparar grupos e quantificar relações entre variáveis.

## 4. Resultados e Insights

As análises foram divididas em duas frentes principais:

### 4.1 - Análise do Desempenho Brasileiro no Oscar:

A análise dos filmes brasileiros indicados ao Oscar revela um cenário onde a excelência artística nacional é inegável, mas enfrenta obstáculos significativos para a vitória final. Filmes como **"O Pagador de Promessas" (1963)**, **"Central do Brasil" (1999)** e **"O Menino e o Mundo" (2016)** demonstram consistentemente um alto nível de qualidade em métricas como cinematografia e narrativa.

Apesar da qualidade, os filmes brasileiros competem com produções que se tornam fenômenos globais, como **"La vita è bella" (1999)** e **"Divertida Mente" (2016)**, que se beneficiam de orçamentos e investimentos de marketing massivos. As campanhas de divulgação ("For Your Consideration") são um diferencial crucial, e a disparidade nos recursos financeiros cria uma barreira significativa.

### 4.2 - Análise da Viabilidade de Franquias (Brasil vs. EUA):

A análise comparativa entre 12 filmes de franquias brasileiras e 49 americanas revelou padrões distintos. As sequências brasileiras na amostra apresentaram uma taxa de sucesso de 100% (pelos critérios de ROI e avaliação) e uma melhora na avaliação média em relação aos originais, sugerindo que, uma vez que uma franquia engrena no Brasil, suas continuações têm alto potencial.

Em contraste, as sequências americanas, embora mais populares em números absolutos, mostraram uma queda tanto na taxa de sucesso quanto na avaliação média em comparação

com seus filmes originais. O gênero "Comédia" se destacou como pilar das franquias brasileiras de sucesso, enquanto nos EUA há uma maior diversidade de gêneros (Ação, Thriller, Animação). O ROI se mostrou uma métrica mais eficaz que a receita bruta para comparar mercados de escalas tão diferentes.

## 5. Conclusões e Próximos Passos

### Resumo das Conclusões:

1. **Cinema Brasileiro no Oscar:** A vitória brasileira no Oscar não é uma questão de falta de qualidade intrínseca. O principal obstáculo reside em fatores externos: a concorrência internacional acirrada, a disparidade nos recursos de marketing e distribuição, e a menor repercussão em larga escala.
2. **Viabilidade de Sequências:** A análise responde "sim, vale a pena fazer uma sequência", mas as condições para o sucesso variam. No Brasil, o êxito parece depender de um primeiro filme com forte apelo cultural e personagens cativantes (principalmente em comédias). Nos EUA, é uma estratégia industrial de alto risco e alto investimento, mais diversificada em gêneros e dependente de marketing global. O ponto de virada para qualquer franquia, em ambos os mercados, reside na capacidade de criar e manter uma conexão emocional e cultural duradoura com o público.

### Limitações:

A principal limitação da análise de franquias é o tamanho da amostra, especialmente para o Brasil (12 filmes), o que torna as conclusões indicativas. A completude dos dados financeiros também é um desafio.

### Sugestões para Trabalhos Futuros:

- Expandir o dataset com um número maior de franquias brasileiras.
- Aprofundar a análise qualitativa de estudos de caso sobre marketing e distribuição internacional de filmes brasileiros.
- Desenvolver um modelo preditivo para estimar o potencial de sucesso de uma sequência com base nas métricas do filme original.

## 6. Detalhes Técnicos e Repositório

### Estrutura do Repositório:

A organização das pastas e arquivos dentro do repositório Git segue a estrutura abaixo:

/beevview - app/	# Contém a aplicação em Python e Tkinter
/beevview notebooks/	# Contém os notebooks e datasets principais
/analise franquia/	
/datasets/	# Contém os datasets usados para a análise de franquias

/resultado/                   # Contém o arquivo Excel com o resultado da análise  
02\_analise\_Franquia.ipynb # Notebook com a análise de franquias  
/analise oscar/  
/datasets/                   # Contém os datasets usados para a análise do Oscar  
01\_analise de dados\_Oscar.ipynb # Notebook com a análise do Oscar  
/documentação/               # Contém esta documentação técnica  
/readme.md                   # Arquivo README principal do projeto

### Instruções de Execução:

1. **Pré-requisitos:** Python 3.9 ou superior.
2. **Instalação de Dependências:** `pip install pandas numpy matplotlib seaborn requests openpyxl streamlit tkinter`
3. **Execução da Análise:** A análise principal está contida nos Jupyter Notebooks (`.ipynb`) dentro das pastas `/beevview notebooks/analise franquia/` e `/beevview notebooks/analise oscar/`. Abra os notebooks no Google Colab ou em um ambiente Jupyter local e execute as células sequencialmente.
4. **Execução da Aplicação:** A aplicação em Tkinter pode ser executada a partir da pasta `/beevview - app/`.

### Link para o Repositório Git:

Repositório Git: [Github BeeView](#)

## 7. Apêndices

### Glossário:

- **ROI (Return on Investment):** Métrica usada para avaliar a eficiência de um investimento. Calculado como  $(\text{Receita} - \text{Orçamento}) / \text{Orçamento}$ .
- **EDA (Exploratory Data Analysis):** Análise Exploratória de Dados, etapa inicial de análise para resumir as principais características dos dados.
- **TMDB (The Movie Database):** Uma base de dados online colaborativa de filmes e séries de TV.

### Informações de Contato:

Email: [antoniocarlossenajunior@gmail.com](mailto:antoniocarlossenajunior@gmail.com)  
Teams: 202102120748@alunos.uniruy.edu.br