

Memorial Descritivo do Projeto - BeeView

Nome do Projeto e Integrantes do Grupo

Nome do Projeto: BeeView

Integrantes:

- 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

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, abordando uma vasta gama de temas, desde questões históricas e sociais até tendências comportamentais e avanços tecnológicos. 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.

Bases de Dados Utilizadas

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

- **The Movie Database (TMDb):** Uma vasta base de dados colaborativa de filmes, programas de TV e celebridades, acessada via API oficial.
 - **Link de Referência:** <https://www.themoviedb.org/>
- **IMDb Datasets:** Contém dados de filmes, séries, elencos e equipes do Internet Movie Database (IMDb), adquiridos por download direto.
 - **Link de Referência:** <https://developer.imdb.com/>
- **Rotten Tomatoes:** Oferece informações sobre filmes, incluindo críticas de críticos e do público, com coleta de dados realizada via web scraping para informações complementares.

- **Link de Referência:** <https://www.rottentomatoes.com/>
- **Kaggle - Full TMDb Movies Dataset 2024 (1M Movies):** Este abrangente dataset consolidou informações de mais de um milhão de filmes do TMDb, incluindo metadados detalhados, elenco, equipe técnica, gêneros, orçamentos, receitas e popularidade. Sua vasta dimensão permitiu análises mais robustas e a identificação de padrões em larga escala.
 - **Link de Referência:** <https://www.kaggle.com/datasets/prathmc/full-tmdb-movies-dataset-2024-1m-movies>
- **AdoroCinema:** Utilizado para informações gerais sobre filmes, lançamentos e avaliações no contexto brasileiro, com coleta de dados pontual para filmes específicos.
 - **Link de Referência:** <https://www.adorocinema.com/filmes/filme-3023/> (e outros filmes, com busca direta na plataforma)
- **Wikipedia (versão em português):** Fonte para contexto de premiações (como o Oscar de Melhor Filme Estrangeiro) e informações gerais sobre filmes e personalidades, com consulta focada em dados históricos.
 - **Link de Referência:** https://pt.wikipedia.org/wiki/Oscar_de_melhor_filme_internacional (e outros filmes, com busca direta na plataforma)
- **Papo de Cinema:** Site de notícias e críticas de cinema brasileiro, utilizado para análises complementares.
 - **Link de Referência:** Recomenda-se buscar por "Emilia Pérez" no site.
- **G1 (Globo):** Cobertura de notícias sobre o Oscar e filmes de destaque em festivais, utilizada para informações contextuais.
 - **Link de Referência:** Recomenda-se buscar por "Emilia Pérez Cannes" ou "Emilia Pérez Oscar".
- **Mídia Internacional (The Hollywood Reporter, Variety, Le Monde, Libération):** Fontes para cobertura detalhada de festivais (como Cannes) e notícias do Oscar.
 - **Link de Referência:** Recomenda-se buscar por "Emilia Pérez review Cannes" ou "Emilia Pérez Oscar".

Tecnologias e Ferramentas Aplicadas

Para o desenvolvimento e a análise deste projeto, foram utilizadas as seguintes tecnologias e ferramentas:

- **Linguagem de Programação:** Python (versão 3.9)
- **Bibliotecas Python para Análise de Dados:**
 - Pandas: Manipulação e análise de dados tabulares.
 - NumPy: Suporte a operações numéricas.
 - Matplotlib e Seaborn: Visualização de dados.
- **Ambiente de Desenvolvimento:** Google Colab
- **Versionamento de Código:** Git
- **Hospedagem de Repositório:** GitHub

Metodologia Utilizada

A metodologia empregada neste projeto seguiu um pipeline de dados estruturado e iterativo, visando garantir a robustez e a confiabilidade das análises e dos modelos desenvolvidos. As etapas foram concebidas para transformar dados brutos em insights acionáveis e, quando aplicável, em sistemas preditivos.

1. Coleta e Aquisição de Dados

A fase inicial consistiu na coleta de dados de diversas fontes, primariamente aquelas focadas no universo cinematográfico. As principais bases de dados utilizadas foram o The Movie Database (TMDb), com aquisição via API oficial, e os IMDb Datasets, baixados diretamente. O Kaggle - Full TMDb Movies Dataset 2024 (1M Movies) foi adquirido por download direto, servindo como uma base ampla e pré-consolidada. Complementarmente, dados do Rotten Tomatoes foram coletados via web scraping para informações de crítica, enquanto o Wikipédia e o AdoroCinema foram utilizados para contexto histórico e detalhes sobre premiações, com coleta manual ou web scraping pontual para dados mais estruturados, quando necessário.

2. Pré-processamento e Limpeza de Dados

Após a coleta, os dados foram submetidos a um rigoroso processo de pré-processamento e limpeza para garantir sua qualidade e consistência. Essa etapa foi crucial para evitar vieses e otimizar a performance dos modelos. A grande escala do Full TMDb Movies Dataset do Kaggle (1M de filmes) exigiu a aplicação de técnicas otimizadas para o tratamento de grandes volumes de dados, como o uso de tipos de dados eficientes no Pandas para reduzir o consumo de memória. As principais ações incluíram: tratamento de valores ausentes, remoção de duplicatas, padronização e normalização de formatos e tipos de dados, tratamento de dados categóricos, e a consolidação de bases de diferentes fontes para criar um dataset unificado.

3. Análise Exploratória de Dados (EDA)

A Análise Exploratória de Dados (EDA) foi fundamental para compreender a estrutura, as tendências e os padrões presentes nos datasets. Esta etapa permitiu a formulação de hipóteses e a identificação de insights preliminares, utilizando estatísticas descritivas e visualizações de dados como histogramas, gráficos de dispersão, gráficos de barras e mapas de calor de correlação.

4. Engenharia de Features

Com base nos insights da EDA, a engenharia de features foi aplicada para criar novas variáveis e transformar as existentes, potencializando a capacidade preditiva dos modelos. Isso incluiu a criação de features temporais (ex: ano de lançamento, idade do filme no momento da análise, ou década de lançamento), cálculo de proporções (ex: "lucratividade" - receita/orçamento,

"relação crítica-público" - nota do Rotten Tomatoes vs. nota do público), e agregação de informações (ex: contagem de filmes por diretor, média de notas para um determinado gênero).

5. Modelagem e Análise

Nesta etapa, os modelos e as abordagens foram empregados para responder às perguntas de pesquisa. O pipeline de análise e modelagem envolveu a definição do problema, seleção de modelos, divisão dos dados em conjuntos de treinamento e teste (e, se aplicável, validação), treinamento e ajuste de hiperparâmetros, e avaliação dos modelos.

- **Modelos Estatísticos:** Foram utilizadas análises de correlação para quantificar a relação entre variáveis (ex: orçamento e receita, número de estrelas e número de votos), e testes de hipótese para comparar grupos (ex: diferença significativa entre a nota de filmes de um determinado gênero e a média geral).

Resultados Obtidos e Considerações Finais

Visão Geral da Análise

A coleta e análise dos dados, majoritariamente via API do TMDb e complementada por um dataset adicional construído pela equipe com base em pesquisa qualitativa, foram cruciais para aprofundar a compreensão sobre o desempenho dos filmes brasileiros no Oscar de Melhor Filme Estrangeiro. As informações brutas iniciais mostraram-se insuficientes para uma análise aprofundada, o que motivou a criação de métricas adicionais (como cinematografia, narrativa, impacto, elegibilidade, entre outras) para uma avaliação mais completa e assertiva.

O Potencial do Cinema Brasileiro no Oscar

A análise revelou que o cinema brasileiro consistentemente apresenta um elevado potencial artístico e técnico, evidenciado pelo desempenho de filmes como:

- **"O Pagador de Promessas" (1963):** Recebeu avaliações superiores às do vencedor do Oscar naquele ano, "Les dimanches de Ville d'Avray", especialmente em cinematografia (8.5) e narrativa (8.8) em nossa avaliação. Sua Palma de Ouro em Cannes já demonstrava o reconhecimento inicial de sua força e competência.
- **"Central do Brasil" (1999):** Obteve notas próximas à perfeição, quase empatando com o vencedor "La vita è bella". A pequena diferença na métrica "Elegibilidade" (9.5 contra 10.0 do vencedor) sugere que critérios formais ou o alinhamento com as regras da premiação podem ter sido um fator decisivo.
- **"O Menino e o Mundo" (2016):** Apresentou desempenho excelente (Cinematografia 9.8, Impacto Narrativo 9.5), mesmo competindo com uma produção de grande porte como "Divertida Mente", que se beneficiou de um investimento massivo em marketing e ampla repercussão.

De modo geral, as médias das métricas definidas no dataset adicional para os filmes brasileiros indicados se enquadram consistentemente na categoria "Excelente", confirmando que o Brasil submete ao Oscar obras de reconhecido valor artístico e técnico. Contudo, nuances competitivas parecem criar uma barreira para a vitória final.

Barreiras para a Vitória: A Competição Internacional e o Marketing Global

A competição no Oscar de Melhor Filme Estrangeiro é intensa, com produções que frequentemente se tornam fenômenos culturais e comerciais:

- **"La vita è bella" (1999):** Além de sua alta qualidade, alcançou uma bilheteria expressiva (US\$ 230 milhões) com um orçamento (US\$ 20 milhões) significativamente maior que o de "Central do Brasil" (US\$ 2.9 milhões). Esse impacto massivo provavelmente influenciou os votantes da Academia.
- **"Divertida Mente" (2016):** Como uma produção da Disney, teve um orçamento astronômico (US\$ 175 milhões) e uma estrutura de marketing global que gerou vasta repercussão e receita, características comuns a grandes produções internacionais que concorrem.
- **"Karakter" (1998):** Superou "O Que É Isso, Companheiro?" em quase todas as métricas do nosso dataset adicional e possuía um orçamento maior, ampliando suas possibilidades de projeção internacional.

As campanhas de marketing e a visibilidade no cenário global desempenham um papel crucial. Embora filmes brasileiros apresentem pontuações favoráveis em marketing em nossa análise ("O Pagador de Promessas" - 8.5; "Central do Brasil" - 9.5), o alcance e o investimento dessas campanhas raramente se comparam aos de grandes estúdios ou produções europeias com forte apoio institucional e distribuição internacional. A descrição de "O Quatrilho" (1996), que menciona uma campanha "modesta para os padrões de grandes estúdios, mas crucial", ilustra bem esse desafio. Dados de orçamento e receita do TMDb corroboram essa diferença, indicando que maior investimento permite campanhas mais agressivas e maior penetração no mercado norte-americano, essencial para o Oscar.

Os critérios de elegibilidade são sempre altos para os filmes brasileiros indicados, demonstrando que eles cumprem os requisitos formais e, frequentemente, alinham-se com temas e estilos valorizados pela Academia em diferentes períodos. No entanto, a subjetividade da votação pode favorecer filmes que criam uma conexão emocional mais profunda com a maioria dos votantes em um determinado ano.

Conclusão: Reconhecimento e Desafios Estratégicos

Em síntese, as produções brasileiras indicadas ao Oscar demonstram, de forma crescente, uma excelência artística e técnica inegável, conforme nossa pesquisa e os dados coletados. No entanto, o principal obstáculo para a vitória final parece residir menos na qualidade intrínseca das obras e mais em uma combinação de fatores externos: a concorrência internacional acirrada, as diferenças nos recursos de marketing e distribuição, a repercussão em larga

escala (gerada por "buzz" e bilheteria), e a sintonia com as sensibilidades da Academia em cada edição.

A indicação em si já é um reconhecimento muito significativo do valor do cinema brasileiro. Para alcançar a vitória, além de manter a excelência artística, parece ser necessário um esforço estratégico conjunto, talvez um investimento em marketing e distribuição internacional mais robusto, e, em alguns casos, circunstâncias favoráveis para superar a força dos concorrentes globais no contexto específico da premiação.

O Desafio do Cinema Brasileiro no Oscar

A análise dos filmes brasileiros indicados ao Oscar, baseada em dados do TMDb e um dataset adicional, revela um cenário onde a excelência artística nacional é inegável, mas enfrenta obstáculos significativos para a vitória final.

Potencial Artístico Reconhecido

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, frequentemente superando ou se equiparando a vencedores do Oscar. A Palma de Ouro para "O Pagador de Promessas" já sinalizava esse reconhecimento internacional. As métricas do dataset adicional classificam as obras brasileiras indicadas como "Excelente", comprovando o valor técnico e artístico que o Brasil submete à Academia.

A Pesada Competição Internacional

Apesar da qualidade, filmes brasileiros competem com produções que se tornam fenômenos culturais e comerciais globais. Exemplos incluem "La vita è bella" (1999), com sua bilheteria massiva e orçamento superior ao de "Central do Brasil", e "Divertida Mente" (2016), uma produção da Disney com investimento e marketing astronômicos. Essas produções de peso se beneficiam de uma infraestrutura e alcance muito maiores, o que influencia diretamente os votantes.

Marketing, "Buzz" e Visibilidade Global

As campanhas de divulgação são um diferencial crucial. Embora os filmes brasileiros tenham pontuações favoráveis em marketing na análise (ex: "O Pagador de Promessas" - 8.5, "Central do Brasil" - 9.5), o tamanho e o alcance dessas campanhas raramente se comparam aos de grandes estúdios. Orçamentos limitados impedem campanhas "For Your Consideration" robustas, essenciais para a visibilidade entre os membros da Academia. A falta de "buzz" gerado por grandes bilheterias ou campanhas de impacto global é uma desvantagem.

Conclusão: Qualidade Brasileira, Desafios Estruturais

Em resumo, a vitória brasileira no Oscar não é uma questão de falta de qualidade intrínseca. O principal obstáculo reside em uma combinação de fatores externos: concorrência internacional acirrada, disparidade nos recursos de marketing e distribuição, menor repercussão em larga escala e a subjetividade dos votantes que favorecem filmes com grande conexão emocional em um determinado ano.

A indicação em si é um reconhecimento significativo. Para conquistar a estatueta, além de manter a excelência artística, o cinema brasileiro precisa de um esforço estratégico conjunto, possivelmente com mais investimento em marketing e distribuição internacional, e a ocorrência de circunstâncias favoráveis que permitam superar a força dos concorrentes globais.

Por que há mais franquias famosas nos EUA do que no Brasil?

A predominância de franquias nos EUA em contraste com o Brasil deve-se a fatores estruturais:

- **Tamanho do Mercado:** O mercado americano é significativamente maior, com mais salas de cinema e público, potencializando o retorno de grandes investimentos.
- **Investimento em Entretenimento:** Estúdios americanos possuem capital superior, viabilizando franquias de alta qualidade e alcance global.
- **Cultura de Franquias:** O público americano está habituado a consumir produtos em série (sequências, universos compartilhados), criando um ambiente fértil para franquias.
- **Infraestrutura de Hollywood:** Oferece produção, distribuição e marketing para sustentar franquias em escala mundial.
- **Barreiras no Brasil:** Menor investimento, capacidade limitada de exportação de conteúdo e foco em filmes isolados com temáticas locais dificultam a criação de franquias.
- **Mercado Interno Limitado:** A bilheteria nacional raramente cobre os altos custos de produção de grandes sequências.

O Ponto de Virada para a Franquia: Além dos Números

Além das métricas quantitativas de ROI, popularidade e avaliação média, é crucial entender que o "ponto de virada" para uma franquia, seja no Brasil ou nos EUA, frequentemente transcende os números frios. Ele reside na **capacidade de um filme original criar uma conexão emocional e cultural duradoura com o público.**

- **No Brasil:** O sucesso de sequências como "Minha Mãe é uma Peça" não se deve apenas à bilheteria inicial, mas à **identificação do público com os personagens, o humor e as situações cotidianas.** O apelo de Paulo Gustavo, por exemplo, foi um motor poderoso que transformou a franquia em um fenômeno cultural. As sequências aqui capitalizaram essa lealdade e o desejo de reviver essa experiência. O ROI elevado

das sequências brasileiras na sua amostra sugere que, uma vez que essa conexão é estabelecida, o retorno pode ser exponencial, pois o público já está "comprado" pela ideia.

- **Nos EUA:** Embora a indústria americana opere em uma escala diferente, com mais foco em universos compartilhados e marketing massivo, o princípio é similar. As franquias de maior sucesso, como Marvel ou Star Wars, não se baseiam apenas em efeitos visuais ou grandes orçamentos, mas na **construção de narrativas cativantes e personagens icônicos** que geram um senso de comunidade entre os fãs. Quando uma sequência não atinge as expectativas (como sugerido pela queda na avaliação média e taxa de sucesso em sua amostra), pode ser um sinal de que a conexão original foi diluída ou que o novo enredo não ressoou com a base de fãs.

Em suma, "fazer uma sequência" não é apenas uma decisão financeira, mas também **artística e estratégica**. O valor de uma franquia, e o sucesso de suas continuações, depende fundamentalmente de sua capacidade de manter e aprofundar o engajamento com o público construído pelo filme original. Sem essa base sólida de afeição e expectativa, mesmo o maior orçamento e o marketing mais agressivo podem não ser suficientes para garantir o sucesso a longo prazo.

Link do Repositório no GitHub

Todo o código-fonte desenvolvido para este projeto, juntamente com a documentação detalhada (incluindo notebooks com as análises e modelos), está disponível publicamente no seguinte repositório do GitHub:

- **Link do Repositório:** [BeeView Github](#)