

O Que os Dados Dizem Sobre o Sucesso no Cinema?



COMUNICAÇÃO ESTRATÉGICA

Publicidade que entende o agora
para projetar o amanhã.

Cenário Fictício

A **agência de publicidade fictícia Lumi Comunicação Estratégica** foi contratada por um grande estúdio de streaming e cinema para auxiliar na criação de campanhas promocionais mais impactantes.

O estúdio deseja entender **quais tipos de filmes têm maior aceitação do público**, seja por bilheteria, avaliação, gênero ou duração, a fim de usar essas informações como base para **criar trailers, slogans, artes visuais e posicionamento nas redes sociais**.

Como analista de dados contratada pela Lumi, meu papel é utilizar uma base de dados pública sobre filmes para **identificar padrões de desempenho** que possam indicar o que está “na moda” ou o que tem mais chances de engajar o público nas campanhas publicitárias.

Objetivos de Análise

1. Identificar os gêneros mais populares e melhor avaliados;
2. Analisar avaliação e gênero;
3. Detectar padrões de sucesso em filmes lançados por ano;
4. Apontar o que define um filme como “promissor” para uso estratégico em publicidade.

Fonte do Dados

Base pública retirada do site Kaggle com mais de 9.000 filmes, a partir de 1937 até 2022.

link: <https://www.kaggle.com/disham993/9000-movies-dataset>

Ferramentas utilizadas

SQL Workbench para consultas do banco de dados em questão;

Excel para organização das colunas e limpeza básica de dados;

Power BI para visualização dos resultados extraídos da base de dados;

-
- **Entregáveis finais:** Documento com comandos SQL e respostas comentadas;
Painéis visuais no Power BI com insights chave;
Slide ou PDF com resumo do projeto, conclusões e sugestões publicitárias.

Perguntas de Análise

1. Qual a média de popularidade por gênero e quais os 3 gêneros com as maiores notas?
2. Os filmes lançados nos últimos 5 anos tiveram melhor desempenho em nota média, comparado aos anos anteriores?
3. Quais anos tiveram mais lançamentos de filmes registrados na base?
4. Houve crescimento ou queda nas notas médias dos filmes ao longo dos anos?
5. Quais são os 10 filmes com maior popularidade? Quais gêneros e línguas eles representam?

Consultas SQL utilizadas em cada pergunta

Questão 1

```
SELECT genre, AVG(popularity) AS media_popularidade  
FROM entretenimento  
WHERE popularity > 0  
GROUP BY genre  
ORDER BY media_popularidade DESC  
LIMIT 3;
```

Questão 2

Ao importar a tabela para o excel, foi necessário criar uma outra coluna de data com as datas corretamente, pois o SQL não estava identificando as datas vindas do excel, deixando-as como “46607”. Tal coluna se chama “Data_certa”. Comando utilizado para a criação da mesma:

```
ALTER TABLE entretenimento ADD COLUMN Data_Certa DATE;  
UPDATE entretenimento,  
SET Data_Certa = DATE_ADD('1899-12-30', INTERVAL Release_Date DAY);
```

Consulta para representar a média entre filmes anteriores a 2020 com filmes lançados nos últimos cinco anos:

```
SELECT
CASE
  WHEN YEAR(Data_Certa) >= 2020 THEN 'Últimos 5 anos'
  ELSE 'Anteriores'
END AS periodo,
COUNT(*) AS qtd_filmes,
AVG(Vote_Average) AS media_notas
FROM entretenimento
WHERE Vote_Average > 0
GROUP BY período;
```

Os filmes anteriores a 2020 tiveram, em média, notas ligeiramente melhores do que os lançados nos últimos 5 anos.

Questão 3

```
SELECT
  YEAR(Data_Certa) AS ano,
  COUNT(*) AS qtd_filmes
FROM entretenimento
WHERE Data_Certa IS NOT NULL
GROUP BY ano
ORDER BY qtd_filmes DESC
limit 10;
```

Nesta consulta, foi colocado o limite dos TOP 10 anos que mais tiveram lançamentos, **sendo o ano de 2021 com 327 lançamentos.**

Questão 4

```
SELECT
  YEAR(Data_Certa) AS ano,
  ROUND(AVG(Vote_Average), 2) AS media_nota
FROM entretenimento
WHERE Vote_Average > 0
GROUP BY ano
ORDER BY media_nota DESC;
```

É demonstrado que ao longo dos anos, a **média caiu**.

Questão 5

```
SELECT
  t.Title,
  t.Popularity,
  t.Genre,
  t.Original_Language
FROM (
  SELECT *
  FROM entretenimento
  WHERE Popularity IS NOT NULL
  ORDER BY Popularity DESC
  LIMIT 10
) AS t;
```

Recomendações

Atenção especial aos gêneros com maior média (Crime, Mystery, Thriller)

Recomendação: Investigar os elementos narrativos que tornam esses gêneros mais bem avaliados pode gerar insights para produção de conteúdo e curadoria de catálogo. É possível que esses gêneros estejam mais alinhados com as expectativas do público em termos de roteiro, ritmo e envolvimento emocional.

Analisar o boom de lançamentos em 2021

Recomendação: Verificar se esse aumento está relacionado à retomada do setor após a pandemia de COVID-19, com represamento de lançamentos de 2020. Também é interessante avaliar se a quantidade elevada impactou negativamente a qualidade percebida dos filmes.

Entender a queda gradual nas notas ao longo dos anos

Recomendação: Investigar se a queda nas notas está ligada à mudança nos critérios do público, ao aumento da produção em massa com menor curadoria, ou a transformações na forma de avaliação (ex: mais usuários, mais exigentes, ou algoritmos diferentes). Isso pode apontar para mudanças no comportamento do consumidor.

Idioma único (Inglês) como limitação

Recomendação: Diversificar a base com filmes de outros idiomas pode abrir espaço para novas comparações culturais e enriquecer a análise de tendências globais. A limitação atual impede observar, por exemplo, o crescimento de cinematografias como sul-coreana ou espanhola, que ganharam destaque nos últimos anos.

Investigar a relação entre popularidade e qualidade (nota)

Recomendação: Verificar se os filmes mais populares também são os mais bem avaliados. Isso pode revelar discrepâncias entre sucesso comercial e percepção crítica, útil para estratégias de marketing e posicionamento de filmes em catálogos.

Recomendações para visualização dos dados

Embora esta análise tenha sido desenvolvida integralmente com SQL, é importante ressaltar o papel da visualização de dados na comunicação de insights. Com o uso de ferramentas como o **Power BI**, seria possível visualizar cada consulta SQL utilizada, deixando ilustrado o que foi extraído da base de dados.

Para que isso fosse realizado, aqui vai o passo a passo dentro do Power BI:

1. Ao abrir o PBI, vá em “Obter dados” > Opção Banco de dados MYSQL (app utilizado)
2. Conecte ao banco de dados e clique em “Conectar”
3. Coloque o servidor e o banco de dados que deseja utilizar
4. Após isso, cole a sua query SQL
5. O Power BI vai carregar só o resultado da consulta após selecionar “fechar e aplicar”
6. Agora crie suas visualizações

Se você tiver salvo suas **consultas SQL como views no banco**, pode puxar direto cada view como uma tabela no Power BI — bem limpo!

Essas recomendações, partem da análise realizada e servem como sugestões para aprofundamento de oportunidades e compreensão de tendências de gênero, idioma e volume de lançamentos.

Limitações da análise

- A coluna “popularidade” é limitada a comparações internas da base e não pode ser interpretada como valor absoluto de audiência ou impacto comercial.
- *Embora o dataset cubra décadas, a distribuição de filmes por ano não é homogênea. Há concentrações em determinados anos (como 2021), o que pode enviesar as médias gerais.*
- falta de contexto externos, como bilheteria, região demográfica e estratégias de streaming.

Considerações finais

Esta análise foi realizada com o objetivo de aprendizagem contínua em relação à linguagem SQL e a extração de insights. Foi uma experiência que me permitiu explorar um tema que nunca havia visto e compreender diferentes métricas, e assim espero que alcance outros estudante de Dados ou Marketing.

*Além disso, o exercício me fez ter outros questionamentos sobre a base utilizada, mostrando que essas “dúvidas” que surgem são, na verdade, o **coração das recomendações analíticas**. Não é sobre ter todas as respostas, mas saber **quais perguntas valem ser feitas** e onde **vale cavar mais fundo**. Apesar da análise ter suas limitações, me foi possível identificar padrões e recomendações de melhoria.*