



# MOVIE CATALOG DATASET



Iniciar





## I **Motivação e Contexto**

- Filmes: meio de entretenimento muito popular.
- Serviços de streaming de vídeo tentando pegar a sua fatia em um mercado crescente.
- Grande quantidade de serviços, além da popularidade das franquias e universos cinematográficos.
- Escolher quais filmes e onde assisti-los: tarefa difícil.



# I **Motivação e Contexto**

- Segundo o IMDB, a média de filmes produzidos por ano é de 2577.
- Empresas da indústria cinematográfica estão explorando maneiras de aumentar seu faturamento bruto de bilheteria.
- É difícil saber o que o público gosta antes de ouvir suas críticas.
- Produção de filmes: tarefa de risco.

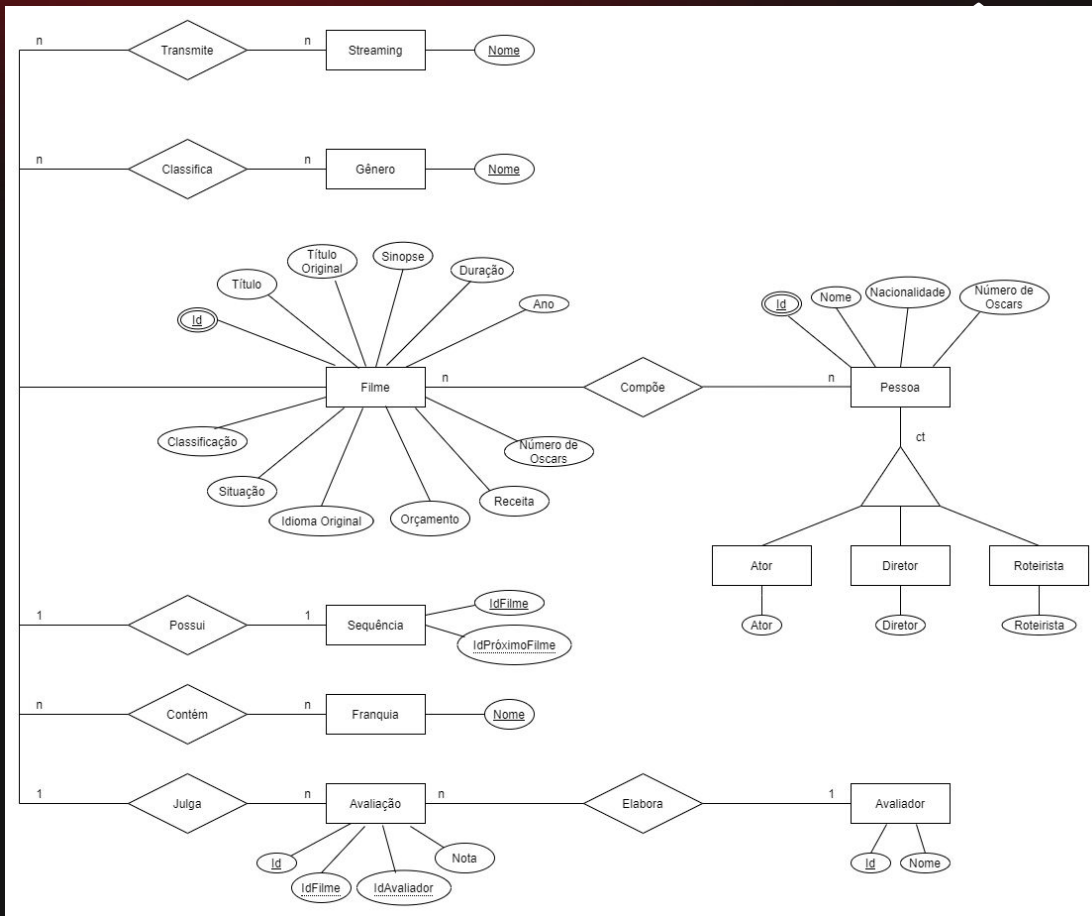


## | Descrição

O Movie Catalog Dataset, objetiva-se a ser uma base de dados sobre a indústria cinematográfica, permitindo a construção de mecanismos de busca e análise a respeito de diversos aspectos relacionados ao cinema.

Sendo que alguns deles são: gêneros, pessoas que participaram de filmes (diretores, roteiristas e atores) e os filmes por si só.

# | Modelo Conceitual





# I Modelo Lógico Relacional

Avaliador(id, nome)

Avaliacao(id\_avaliador\*, id\_filme\_TMDB\*, nota)

Franquia(nome)

FranquiaFilme(nome\_franquia\*, id\_filme\_TMDB\*)

Genero(nome)

GeneroFilme(nome\_genero\*, id\_filme\_TMDB\*)

Filme(id\_TMDB, id\_IMDB, titulo, titulo\_original, sinopse, duracao, ano, classificacao, situacao, idioma\_original, orcamento, receita, num\_oscars)

Pessoa(id\_TMDB, id\_IMDB, nome, nacionalidade, num\_oscars)

PessoaFilme(id\_pessoa\_TMDB\*, id\_filme\_TMDB\*, ator, diretor, roteirista)

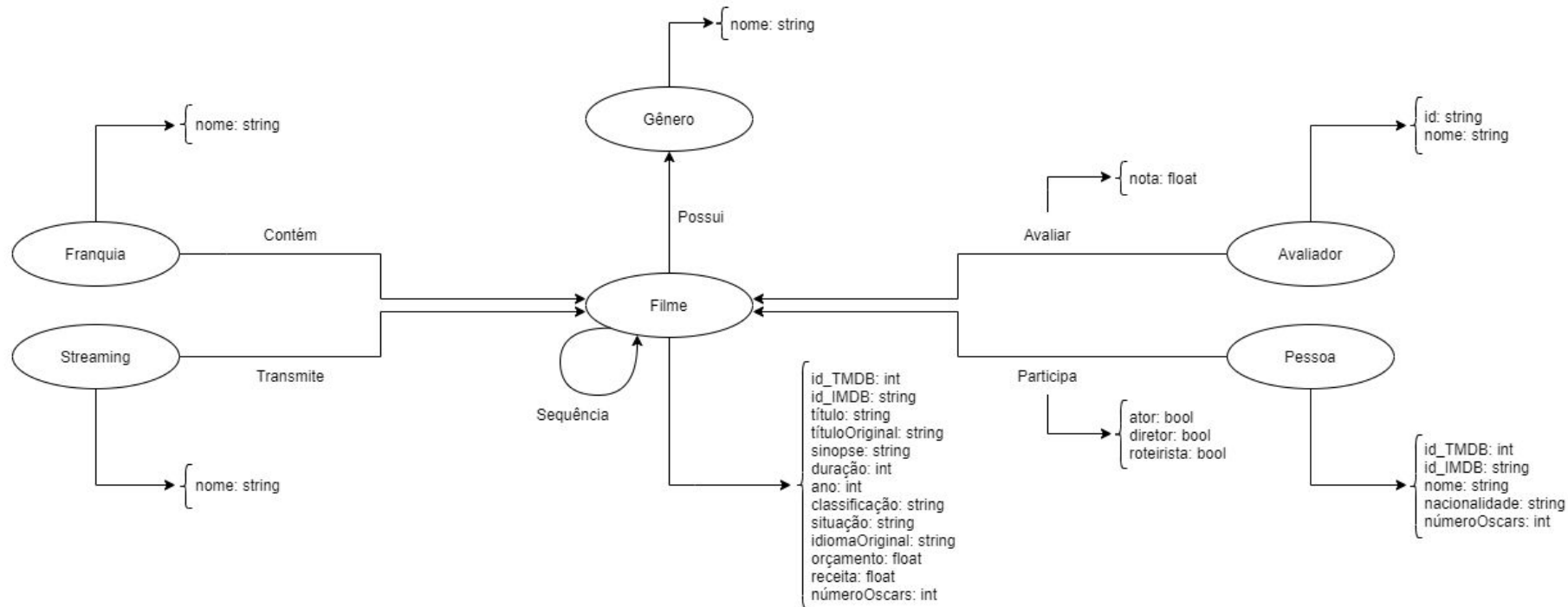
Sequencia(id\_filme\_TMDB\*, id\_filme\_sequencia\_TMDB\*)

Streaming(nome)

StreamingFilme(nome\_streaming\*, id\_filme\_TMDB\*)



# Modelo Lógico de Grafos



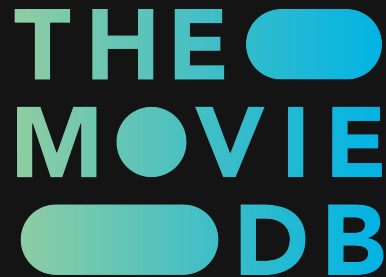


## | Fontes de Dados

Banco de dados que conta com informações sobre filmes, séries, video games e premiações, possui uma API que permite acesso via key aos dados de sua base.



Banco de dados aberto e colaborativo que contém dados sobre milhares de filmes e séries. Possui uma API públicas pela qual é possível acessar os dados de sua base.





## | Ferramentas Utilizadas





# | Operações realizadas para a construção do Dataset

Operações aplicadas:

- Agregação de dados do TMDB e IMDB;
- *Web Scraping* em páginas do Wikipédia para obtenção de informações sobre franquias de filmes;
- Transformação e tratamento de dados para obtenção das informações relevantes ao Dataset.



# | Perguntas de Pesquisa/Análise

- **Query 1.** Os filmes que fizeram sucesso com o público também fizeram sucesso com a crítica especializada?
- **Query 2.** Qual é o gênero que fez mais sucesso com a crítica (maior nota média) na década de 2000?
- **Query 3.** Qual é o gênero que fez mais sucesso com o público (maior renda média) na década de 2000?
- **Query 4.** Qual é a franquia que fez mais sucesso com a crítica?
- **Query 5.** Qual é a franquia que fez mais sucesso com o público?



# | Query 1

```
/* Relação entre sucesso com o público (receita)
e sucesso com a crítica (nota média) dos filmes */
```

```
DROP TABLE IF EXISTS FilmeReceitaNota;
```

```
CREATE VIEW FilmeReceitaNota AS
```

```
    SELECT A.id_filme, F.titulo, F.ano, F.receita, AVG(A.nota) nota_media
    FROM Avaliacao A, Filme F
    WHERE A.id_filme = F.id_TMDB
    GROUP BY A.id_filme;
```

```
-- Ordenação decrescente por receita
```

```
SELECT titulo, receita, nota_media
FROM FilmeReceitaNota
ORDER BY receita DESC;
```

```
-- Ordenação decrescente por nota média
```

```
SELECT titulo, receita, nota_media
FROM FilmeReceitaNota
ORDER BY nota_media DESC;
```

index	TITULO	RECEITA	NOTA_MEDIA
0	Avatar	2847246203	7.7
1	Vingadores: Ultimato	2797800564	8.4
2	Star Wars: O Despertar da Força	2068223624	7.5
3	Vingadores: Guerra Infinita	2046239637	8.4
4	Jurassic World: O Mundo dos Dinossauros	1671713208	6.8
5	O Rei Leão	1667635327	7

index	TITULO	RECEITA	NOTA_MEDIA
0	Batman: O Cavaleiro das Trevas	1004558444	8.8
1	O Senhor dos Anéis: O Retorno do Rei	1118888979	8.7
2	O Senhor dos Anéis: A Sociedade do Anel	871368364	8.6
3	O Senhor dos Anéis: As Duas Torres	926287400	8.6
4	A Origem	825532764	8.6
5	Vingadores: Ultimato	2797800564	8.4



## | Query 2

```
/* Gênero que fez mais sucesso com a crítica, isto é,  
   que tem os filmes com as melhores notas médias, na década de 2000 */  
  
DROP TABLE IF EXISTS GeneroCritica;  
  
CREATE VIEW GeneroCritica AS  
    SELECT G.nome, SUM(F.nota_media) nota_total, COUNT(F.id_filme) qtdade_filme  
    FROM Genero G, FilmeReceitaNota F, GeneroFilme GF  
    WHERE G.nome = GF.nome_genero  
          AND F.id_filme = GF.id_filme  
          AND ((F.ano >= 2000) AND (F.ano < 2010))  
    GROUP BY G.nome;  
  
SELECT nome  
    FROM GeneroCritica  
    WHERE (nota_total / qtdade_filme) = (SELECT MAX(nota_total / qtdade_filme)  
                                         FROM GeneroCritica);
```

index	NOME
0	Crime
1	Drama



## | Query 3

```
/* Gênero que fez mais sucesso o público, isto é,  
   que tem os filmes com a maior renda média, na década de 2000 */  
  
DROP TABLE IF EXISTS GeneroPublico;  
  
CREATE VIEW GeneroPublico AS  
    SELECT G.nome, SUM(F.receita) receita_total, COUNT(F.id_filme) qtdade_filme  
    FROM Genero G, FilmeReceitaNota F, GeneroFilme GF  
    WHERE G.nome = GF.nome_genero  
          AND F.id_filme = GF.id_filme  
          AND ((F.ano >= 2000) AND (F.ano < 2010))  
    GROUP BY G.nome;  
  
SELECT nome  
    FROM GeneroPublico  
    WHERE (receita_total / qtdade_filme) = (SELECT MAX(receita_total / qtdade_filme)  
                                           FROM GeneroPublico);
```

Ficção científica



## | Query 4

```
/* Franquia que fez mais sucesso com a crítica,  
isto é, que tem os filmes com a maior nota média */  
  
DROP TABLE IF EXISTS FranquiaCritica;  
  
CREATE VIEW FranquiaCritica AS  
    SELECT FR.nome, SUM(FI.nota_media) nota_total, COUNT(FF.id_filme) qtdade_filme  
        FROM Franquia FR, FilmeReceitaNota FI, FranquiaFilme FF  
        WHERE FR.nome = FF.nome_franquia  
            AND FI.id_filme = FF.id_filme  
        GROUP BY FR.nome;  
  
SELECT nome  
    FROM FranquiaCritica  
    WHERE (nota_total / qtdade_filme) = (SELECT MAX(nota_total / qtdade_filme)  
                                         FROM FranquiaCritica);
```

The Lord of the Rings



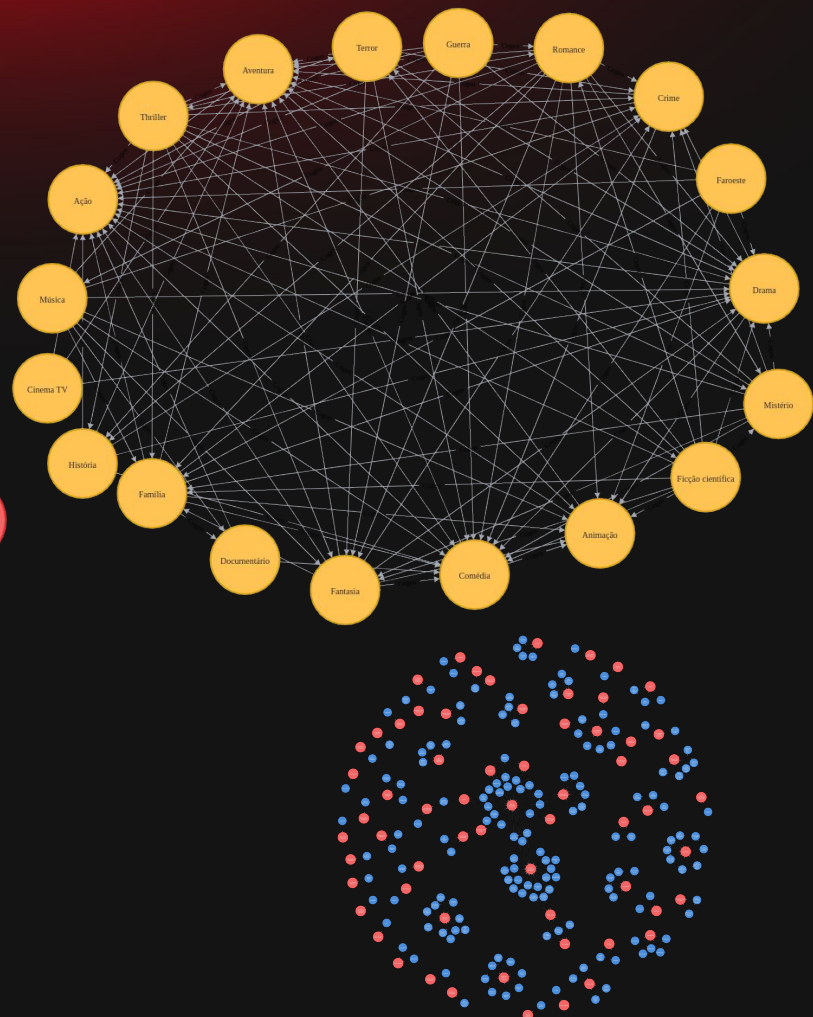
## | Query 5

```
/* Franquia que fez mais sucesso com o público,  
   isto é, que tem os filmes com a maior renda média */  
  
DROP TABLE IF EXISTS FranquiaPublico;  
  
CREATE VIEW FranquiaPublico AS  
    SELECT FR.nome, SUM(FI.receita) receita_total, COUNT(FF.id_filme) qtdade_filme  
    FROM Franquia FR, Filme FI, FranquiaFilme FF  
    WHERE FR.nome = FF.nome_franquia  
          AND FI.id_TMDB = FF.id_filme  
    GROUP BY FR.nome;  
  
SELECT nome  
    FROM FranquiaPublico  
    WHERE (receita_total / qtdade_filme) = (SELECT MAX(receita_total / qtdade_filme)  
                                             FROM FranquiaPublico);
```

Avatar Collection



**SPOILER ALERT**





# Obrigado!

## Alguma pergunta?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik

Please keep this slide for attribution