



MOVIE CATALOG DATASET



Iniciar





I Contexto e Motivação

- 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 Contexto e Motivação

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



| Descrição

O Movie Catalog Dataset, visa ser uma base de dados sobre a indústria cinematográfica, que permitirá a criação de mecanismos de busca por filmes, capazes de realizar consultas como:

- Procurar por franquias ou universos cinematograficos, ordenando os filmes em uma sequência topológica.
- Busca de filmes dirigidos por um certo(a) diretor(a) ou protagonizados por um determinado artista.



| Descrição

- Encontrar os filmes melhores avaliados ou que receberam alguma premiação.
- Descobrir em quais plataformas um determinado filme pode ser encontrado.
- Ranquear as plataformas com base nas avaliações dos filmes contidos nelas o tamanho do catálogo e o custo do serviço.

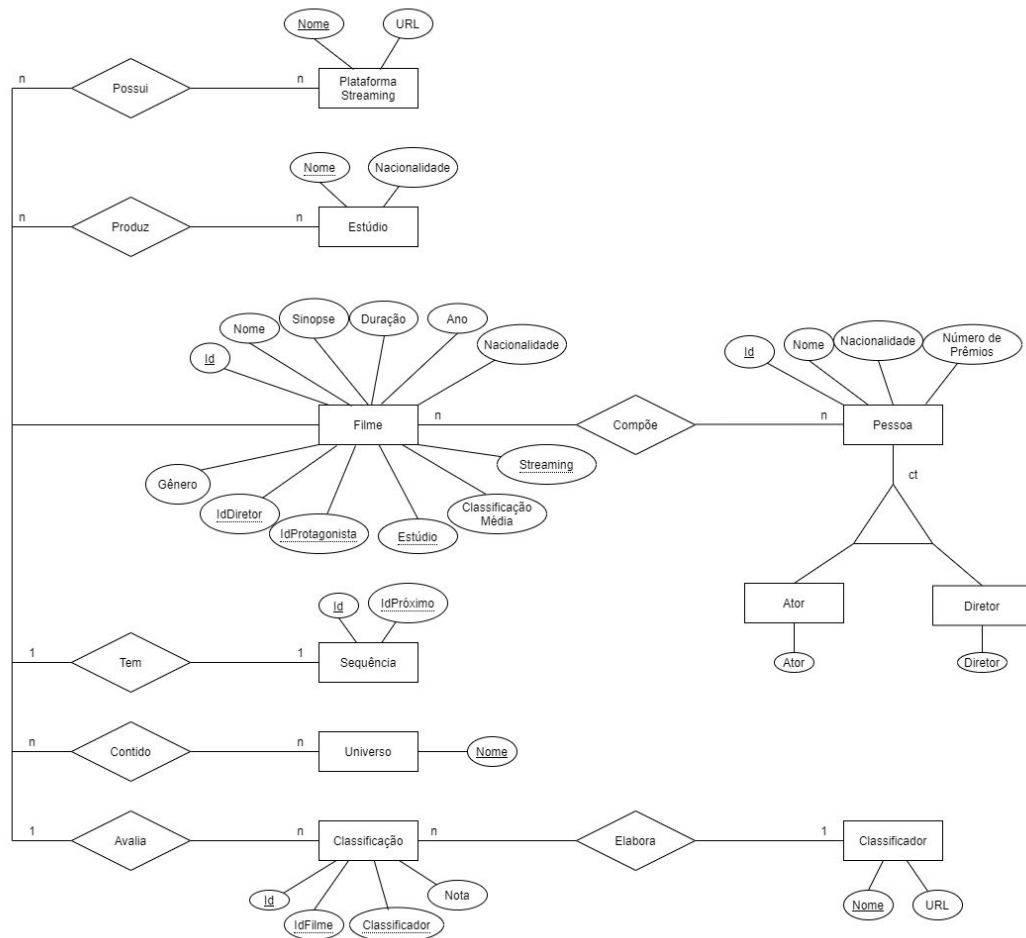


| Descrição

O Movie Catalog Dataset poderá ser usado em estudos sobre a indústria cinematográfica como, por exemplo:

- Reconhecer características em filmes que os fazem ter sucesso e com essas características, prever a bilheteria de novos filmes antes do seu lançamento.
- Encontrar os filmes e gêneros mais populares ao longo das décadas e estudar a relação entre essa popularidade e a mudança nas tendências culturais ao longo do tempo.

| Modelos Conceitual





| Modelos Lógicos

- Modelo Relacional
 - Consultas simples (filmes, diretores, artistas, etc)
- Modelo de Grafos
 - Caminhar nas relações entre filmes (franquias, universos)
 - Relações entre filmes, diretores, artistas
- Modelo de Documentos
 - Criação de catálogos
 - Formatação de dados para uso em API



I Conjunto Inicial de Perguntas

- Como se dá a distribuição de gêneros de filmes dentre os principais serviços de streaming?
- Qual a ordem correta para assistir uma franquia de filmes?
- Quais serviços de streaming melhor englobam meus gostos?
- Qual a é a avaliação média de um filme segundo os principais rankings?
- Quais filmes foram dirigidos pelo diretor X?



I Conjunto Inicial de Perguntas

- Quais filmes o ator/atriz X protagonizou?
- Quais as produtoras de filmes que mais faturam com bilheterias anualmente?
- Quais os gêneros de filmes mais populares por década?
- Quais os filmes mais populares por década?



| Fontes de Dados

IMDB - Internet Movie Database

Banco de dados com informações sobre filmes, series de tv, video games e premiações, possui uma api que permite acesso via key a várias informações em sua base.



TMDB - The Movie DataBase

Banco de dados aberto e colaborativo, que contém dados sobre milhares de filmes e series de tv. Possui uma api pública pela qual é possível consumir diversas informações sobre filmes e séries.





| Fontes de Dados

Movie Industry

Dataset com 6820 filmes lançados entre 1986-2016. Têm informações como:

- Orçamento
- Produtora
- Diretor
- Faturamento
- Tempo de duração

TMDB 5000 Movies

Dataset com 5000 Filmes. Contêm informações como:

- Título original
- Overview
- Popularidade
- Status de lançamento
- Língua Original



| Fontes de Dados

The Story of Film

Dataset com mais de 45000 Filmes (até 2017) extraídos do TMDB, contendo um total de 24 colunas, com informações como:

- Classificação de idade
- Franquia
- Generos
- Homepage
- Número de votos

The Oscar Award, 1927 - 2020

Dataset com vencedores do Oscar entre 1927 e 2020, Contêm 7 colunas, entre elas:

- Ano do filme
- Ano da premiação
- Categoria
- Nome do premiado
- Nome do filme



| Fontes de Dados

Movies on Netflix, Prime Video, Hulu and Disney+

Dataset que contém uma coleção (9515 registros) de filmes encontrados nessas plataformas de streaming. É formado por 17 colunas, com informações como:

- Nome do filme
- Nome do diretor
- Classificação de idade
- Rotten Tomatoes
- Gênero



| Fontes de Dados

Operações que serão aplicadas:

- Agregação de dados do IMDB e TMDB;
- Integração dos Dados dos Datasets e API's;
- Tratamento dos dados em todas as bases durante a integração das informações;
- Web Scrapping em sites com informações adicionais as coletadas nas fontes de dados anteriores, caso necessário;
- Transformação de dados para implementação de novas consultas;



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