

Sumário

[Sumário](#)

[Github](#)

[Notebooks criados no Databricks](#)

[Camada Bronze](#)

[Camada Silver](#)

[Camada Gold](#)

[Relatórios](#)

[Coleta](#)

[Objetivo](#)

[Dicionário de dados](#)

[TITLE.AKAS.TSV](#)

[TITLE.BASICS.TSV](#)

[TITLE.CREW.TSV](#)

[TITLE.EPISODE.TSV](#)

[TITLE.PRINCIPALS.TSV](#)

[TITLE.RATINGS.TSV](#)

[NAME.BASICS.TSV](#)

[Camada BRONZE](#)

[Importação dos dados do IMDb](#)

[Processo de importação](#)

[Camada SILVER](#)

[Camada Gold](#)

[TB_TITULO_LANCAMENTO_BRASIL](#)

[TB_ATOM_TITULO](#)

[TB_ATOM_TITULO_BRASIL](#)

[Análises e Resultados](#)

Github

Segue abaixo o link e os processos realizados nos Notebooks criados para cada etapa do MVP.

Github pessoal (Thomas Abrantes): <https://github.com/hannemanbr/PUC-RIO-MVP/tree/main>

Notebooks criados no Databricks

Link dos notebooks com cada processo da MVP no GitHub.

Camada Bronze

<https://github.com/hannemanbr/PUC-RIO-MVP/blob/main/Thomas-Notebook-PUC-RIO%20-%20Camada%20Bronze.ipynb>

Camada Silver

<https://github.com/hannemanbr/PUC-RIO-MVP/blob/main/Thomas-Notebook-PUC-RIO%20-%20Camada%20Silver.ipynb>

Camada Gold

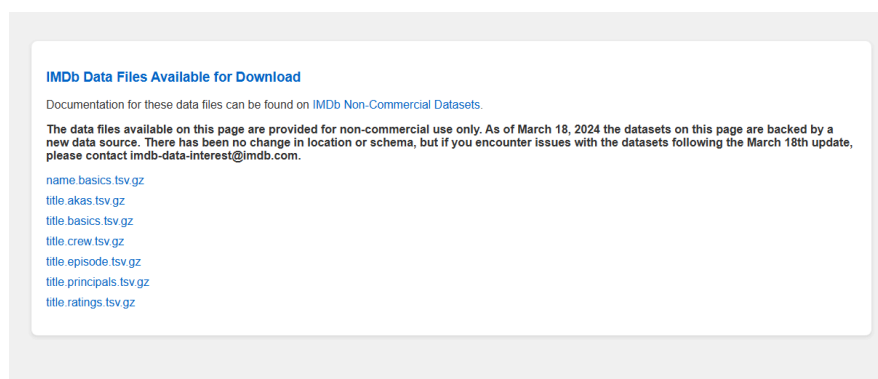
<https://github.com/hannemanbr/PUC-RIO-MVP/blob/main/Thomas-Notebook-PUC-RIO%20-%20Camada%20Gold.ipynb>

Relatórios

<https://github.com/hannemanbr/PUC-RIO-MVP/blob/main/Thomas-Notebook-PUC-RIO%20-%20Relatorios.ipynb>

Coleta

O IMDb é uma Base de Dados de Filmes na Internet, contendo informações de séries, documentários, videogames e outras mídias. Todos os dados utilizados neste trabalho foram coletados nos dados públicos disponibilizados pelo IMDb (*Internet Movie Database*) no link <https://datasets.imdbws.com/>

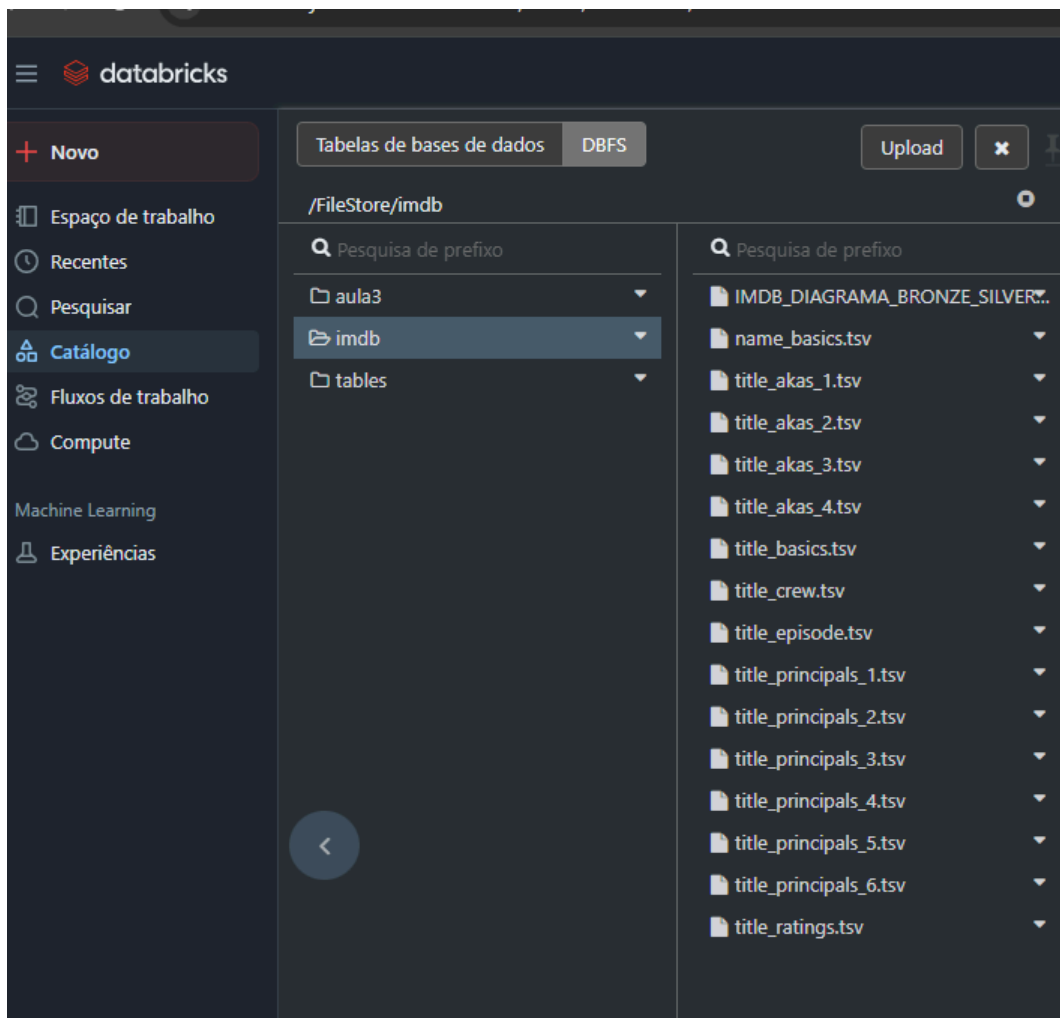


Página com arquivos utilizados na coleta de dados em <https://datasets.imdbws.com>

Utilizando a base de acesso publico no endereço <https://datasets.imdbws.com/> que disponibiliza os arquivos no formato TSV compactado no padrão GZIP:

Todos os arquivos estão compactados (formato Gnu Zip).

- title.principals.tsv.gz
- title.basics.tsv.gz
- name.basics.tsv.gz
- title.akas.tsv.gz
- title.crew.tsv.gz
- title.episode.tsv
- title.ratings.tsv.gz



Catálogo no Databricks após upload dos arquivos TSV

Após a descompactação dos arquivos estes foram enviados para o Catálogo no Databricks em uma pasta IMDB por meio do DBFS no Databricks.

Objetivo

Nesse MVP optei por analisar os dados públicos disponibilizados pelo IMDb (*Internet Movie Database*) no link <https://datasets.imdbws.com/> por se tratar de um conteúdo que familiaridade.

Através da análise dos arquivos e da documentação disponibilizada no link <https://developer.imdb.com/non-commercial-datasets/>, procuro responder às seguintes perguntas:

- Quantidade de títulos lançados no Brasil por categoria e década (80, 90, etc.)?
- Qual categoria teve mais lançamentos no Brasil por década?
- Todos os títulos que possuem nome exclusivo no Brasil (Com nome diferente do nome original ou comercial).
- Qual a porcentagem de pessoas que atuaram como Ator ou Atriz em Títulos do tipo Filmes, Curtas ou Vídeo, lançados no Brasil?
- Porcentagem de atores do sexo feminino e masculino que atuaram nos títulos do tipo Filme em cada década existente na base.
- Filmes lançados no Brasil, onde o Diretor(a) atuou como Ator/Atriz.
- Ranking dos diretores com títulos lançados no Brasil.

Dicionário de dados (Camada Bronze / Silver)

Esta base consiste nos arquivos abaixo no formato *tsv* e de acordo com a documentação disponibilizada pelo IMDBWS, A primeira linha em cada arquivo contém cabeçalhos que descrevem o que há em cada coluna. Um 'N' indica que um campo específico está faltando ou é nulo para aquele um título/nome.

TB_TITLE_AKAS (TITLE.AKAS.TSV)

Arquivo contendo os títulos AKAs, que são Títulos Alternativos.

São os nomes diferentes que um filme pode ter, em diferentes idiomas e países.

O arquivo consiste nas colunas:

- **titleId** (string): Identificador no formato alfanumérico exclusivo do título.
- **ordering** (inteiro) – É um número para identificar exclusivamente as linhas para um determinado **titleId**. Neste arquivo há mais de uma linha com mesmo **titleId**.
- **title** (string) – Nome do título.
- **region** (string) - Descreve a região, com dois caracteres.
Exemplo: US para United States, JP para Japão na versão do título.
- **language** (string) - o idioma do título.
- **types** (string) - Conjunto enumerado em forma de string de atributos para este título alternativo.

Na documentação é indicado como matriz mas na importação esses valores vão como uma string contendo todos os valores da coluna e caso seja usado será tratado como uma lista.

De acordo com a documentação pode conter os valores abaixo mas **destaca que** pode haver novos valores podem ser adicionados no futuro::

- alternative
- dvd
- festival
- tv
- video
- working
- original
- imdbDisplay
- **attributes** (string) - Termos adicionais para descrever este título alternativo.
Na documentação é indicado como matriz mas na importação esses valores vão como uma string contendo todos os valores da coluna e caso seja usado será tratado como uma lista.
- **isOriginalTitle** (inteiro) – O Valor 0 (false) indica que o título não original e 1 (true) que o título é original.

TB_TITLE_BASICS (TITLE.BASICS.TSV)

Arquivo contendo as informações básicas dos títulos.

O arquivo consiste nas colunas:

- **tconst** (string) - Identificador no formato alfanumérico exclusivo do título.

- **titleType** (string) – Informa o tipo/formato do título.
Exemplo: filme, curta, série de TV, episódio de TV, vídeo, etc.
No arquivo esses valores estão registrados em inglês como movie, short, tvseries, tvepisode, video.
- **primaryTitle** (string) – Nome do título popularmente conhecido ou usado pelos cineastas em materiais promocionais no seu lançamento.
- **originalTitle** (string) - Título original, no idioma original.
- **isAdult** (boolean) - O valor 0(false) informa que o título não é somente para o público adulto e 1(true) informa que título é destinado ao público adulto.
- **startYear** (inteiro com 4 caracteres) - Ano de lançamento do título. Caso seja uma série de TV o valor é o ano da primeira temporada da série.
- **endYear** (inteiro com 4 caracteres) - Ano de término para série de TV. Para outros títulos esse valor vem preenchido com “N”.
- **runtimeMinutes** - Tempo, em minutos, de execução do título principal.
- **genres** (matriz de strings) – Pode conter até três gêneros associados ao título. Esse campos quando houver mais de uma gênero, será exibido separado por uma vírgula.
Exemplo: Documentary, Short.

TB_TITLE_CREW (TITLE.CREW.TSV)

Arquivo que contém informações da equipe de filmagem de um filme ou programa de TV. O arquivo consiste nas colunas:

- **tconst** (string) - Identificador no formato alfanumérico exclusivo do título.
- **directors** (matriz de **nconsts** no formato string) - Identificador no formato alfanumérico identificando o(s) diretores do título.
- **writers** (matriz de **nconsts** no formato string) - Identificador no formato alfanumérico identificando o(s) escritores do título.

TB_TITLE_EPISODE.TSV (TITLE.EPISODE.TSV)

Arquivo que contém informações de um episódio de uma série de um programa de TV. O arquivo consiste nas colunas:

- **tconst** (string) - identificador alfanumérico do episódio

- **parentTconst** (string) - identificador alfanumérico da série de TV de Origem (pai).
- **seasonNumber** (inteiro) – Número da temporada do episódio.
- **episodeNumber** (inteiro) – Número do episódio com valor **tconst** em uma série de TV.

TITLE.PRINCIPALS.TSV

Observação: Devido ao seu tamanho, o arquivo foi particionado em 6 partes, e posteriormente, unificado em um dataframe para a carga dos dados no Notebook no Databricks.

Arquivo que contém informações do título original de uma obra, na sua língua original. O arquivo consiste nas colunas:

- **tconst** (string) - Identificador alfanumérico exclusivo do título
- **ordering** (inteiro) – É um número para identificar exclusivamente as linhas para um determinado **titleId**.
- **nconst** (string) - Identificador alfanumérico exclusivo do nome/pessoa
- **category** (string) - Categoria do trabalho em que a pessoa executou na produção.
- **job** (string) - Cargo específico, se aplicável, caso contrário '\N'
- **characters** (string) - Nome do personagem interpretado, se aplicável, caso contrário '\N'

TB_TITLE)RATINGS (TITLE.RATINGS.TSV)

Arquivo que contém informações sobre as classificações e o número de votos para cada título no site. O arquivo consiste nas colunas:

- **tconst** (string) - Identificador alfanumérico exclusivo do título
- **averageRating** – Valor da média ponderada de todas as avaliações individuais dos usuários.
- **numVotes** (inteiro) - Número de votos que o título recebeu.

TB_NAME.BASICS.TSV (NAME.BASICS.TSV)

Arquivo que contém informações básicas de uma pessoa.

O arquivo consiste nas colunas:

- **nconst** (string) - identificador alfanumérico exclusivo do nome/pessoa
- **primaryName** (string) – nome pelo qual a pessoa é mais frequentemente creditada
- **birthYear** (inteiro) – Ano de nascimento no formato AAAA (4 dígitos).
- **deathYear** (inteiro) – Ano de falecimento no formato AAAA (4 dígitos). se aplicável, caso contrário '\N'.
- **primaryProfession** (matriz de strings) – Valor das três principais profissões da pessoa.
- **knownForTitles** (matriz de **tconsts**) – Valores dos títulos pelos quais a pessoa é conhecida.

Camada BRONZE

Este processo consiste em:

- Obter os dados brutos dos sete arquivos disponíveis em <https://datasets.imdbws.com/>
- Importação dos dados sem alteração, na forma como estão dispostos nos arquivos de origem.
- Carga dos dados dos arquivos e criação das tabelas no banco de dados (*schema*) IMDB_DB_BRONZE.
- Garantir a qualidade dos dados validando o volume de dados importados de acordo com a quantidade de registros dos arquivos.

Banco de Dados: IMDB_DB_BRONZE

Todas as colunas das tabelas criadas com a importação dos dados dos arquivos, são do tipo string.

Com os arquivos disponíveis no datasets do IMDB, podemos realizar a carga para o banco de dados (*schema*) IMDB_DB_BRONZE, importando os dados dos arquivos para tabelas com nomes contendo prefixo "tb_" e nome do arquivo de origem nos seus respectivos DataFrames.

Processo de importação

- Upload dos arquivos para a pasta IMDB pelo DBFS no Databricks
- Leitura e conversão em dataframe de cada arquivo TSV com o comando **spark.read.csv('CAMINHO DO ARQUIVO NO DBFS', sep='t', header=True)** da biblioteca `pyspark.sql.session.sparksession`.

O parâmetro **header=true** foi utilizado para que o PySpark considere a primeira linha como header do arquivo (representando as colunas)

- Para cada dataframe, criação da tabela com comando **SQL** com nome das colunas iguais ao disponível na primeira linha de cada arquivo (header).
- Inserção dos dados na tabela criada com o comando **df_title_basics.write.format("delta").mode("overwrite").saveAsTable([schema].[NOME DA TABELA])**

O parâmetro “overwrite” indica que sempre que ele for executado vai sobrescrever os dados na tabela de destino.

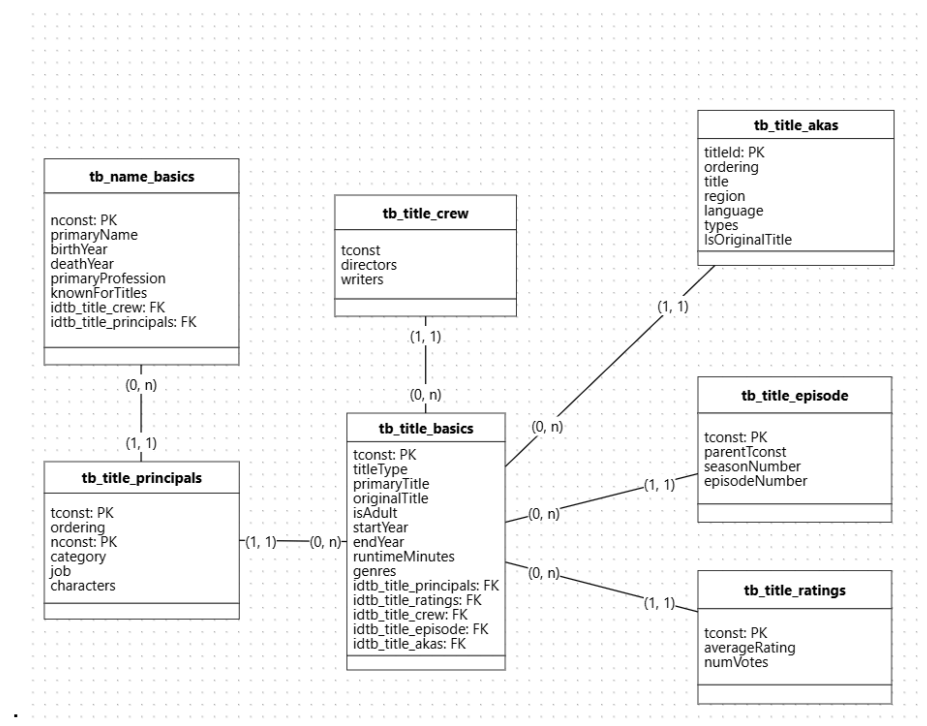


Diagrama de Entidade e Relacionamento da base de dados Bronze e Silver

Após a importação temos o seguinte diagrama da base IMDB_DB_BRONZE

Observação:

- Todas as informações das tabelas e entidades desta base de dados estão na guia “Dicionário de Dados”.

Camada SILVER

Banco de Dados: IMDB_DB_SILVER

Neste processo, ocorre a importação dos dados existentes nas tabelas do *schema* IMDB_DB_BRONZE, mantendo o mesmo nome das tabelas no *schema* IMDB_DB_SILVER.

Durante a importação ocorre a limpeza dos dados conforme os seguintes passos:

- Criação do banco de dados (*schema*) IMDB_DB_SILVER
- Criação de todas as tabelas, com mesmo nome e nome das colunas, na IMDB_DB_BRONZE (camada bronze).
- As tabelas criadas na camada Silver, tabelas seguindo o tipo das colunas de acordo com a documentação.
- Remover o valor da legenda "N" em algumas colunas que indica que o valor “não é aplicável” e dados nulos ou inadequados..
- Tratamentos específicos, se for necessário, nas tabelas (maiores detalhes estão especificados no comentário do Notebook). Verificação na qualidade dos dados em busca de chaves duplicadas ou duplicidade, valores indevidos e nulos.
- Conferência na quantidade de dados importados
 - Quando não há condicional que remova registros, verificar se mantém a mesma quantidade de registros na camada SILVER.
 - Quando há condicional que remova registros, verificar se mantém a quantidade igual ou menor de registros na camada SILVER.

Observação:

- Todas as informações das tabelas e entidades desta base de dados estão na guia “Dicionário de Dados”.
- Na versão Community do databricks, ele não permite a criação de chave primária e estrangeira, então para garantir a qualidade dos dados, na camada SILVER os campos CHAVE da tabela são do tipo NOT NULL evitando a carga de valor nulo.

Após a importação termos o seguinte diagrama da base IMDB_DB_SILVER:

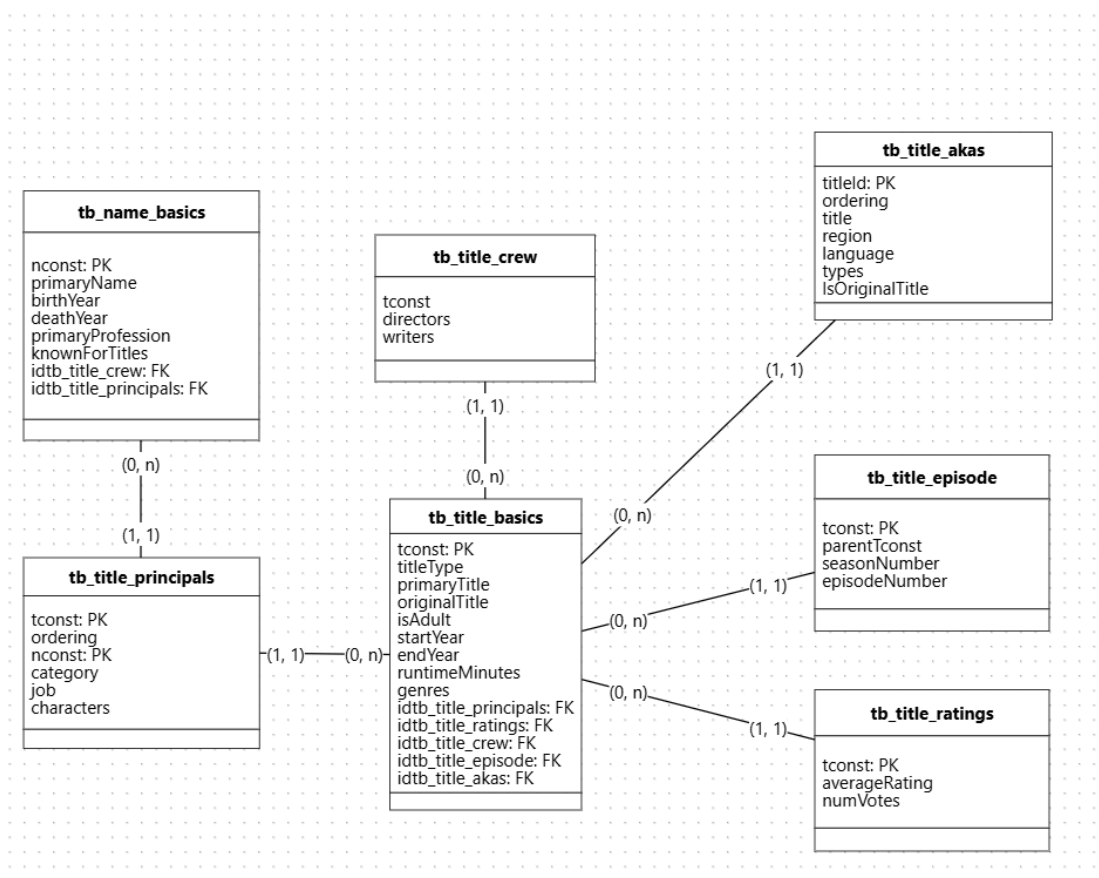


Diagrama de Entidade e Relacionamento da base de dados Bronze e Silver

Camada Gold

Banco de Dados: IMDB_DB_SILVER

Na versão Community do databricks, ele não permite a criação de chave primária e estrangeira, então para garantir a qualidade dos dados, na camada GOLD os campos CHAVE da tabela são do tipo NOT NULL evitando a carga de valor nulo. A camada GOLD consiste em:

- Tabelas e suas colunas com nomes mais amigáveis.
- Tabelas com origem dos relacionamentos de tabelas na camada SILVER, que atendem ao resultado que será consultado no Notebook Relatórios.
- Tabelas com origem dos relacionamentos de tabelas na camada GOLD, que atendem ao resultado que será consultado no Notebook Relatórios.

No diagrama abaixo podemos ver as origens de cada tabela criada na Camada Gold.

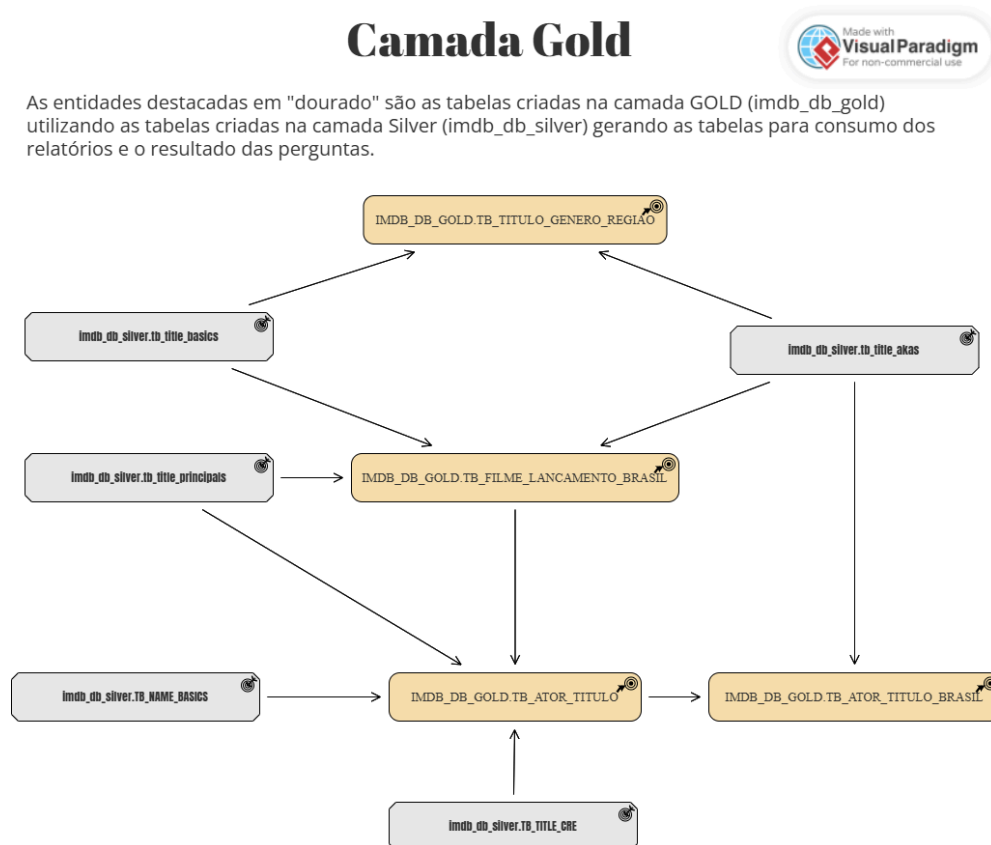


Diagrama das Entidades criadas na camada Gold e as origens na camada Silver

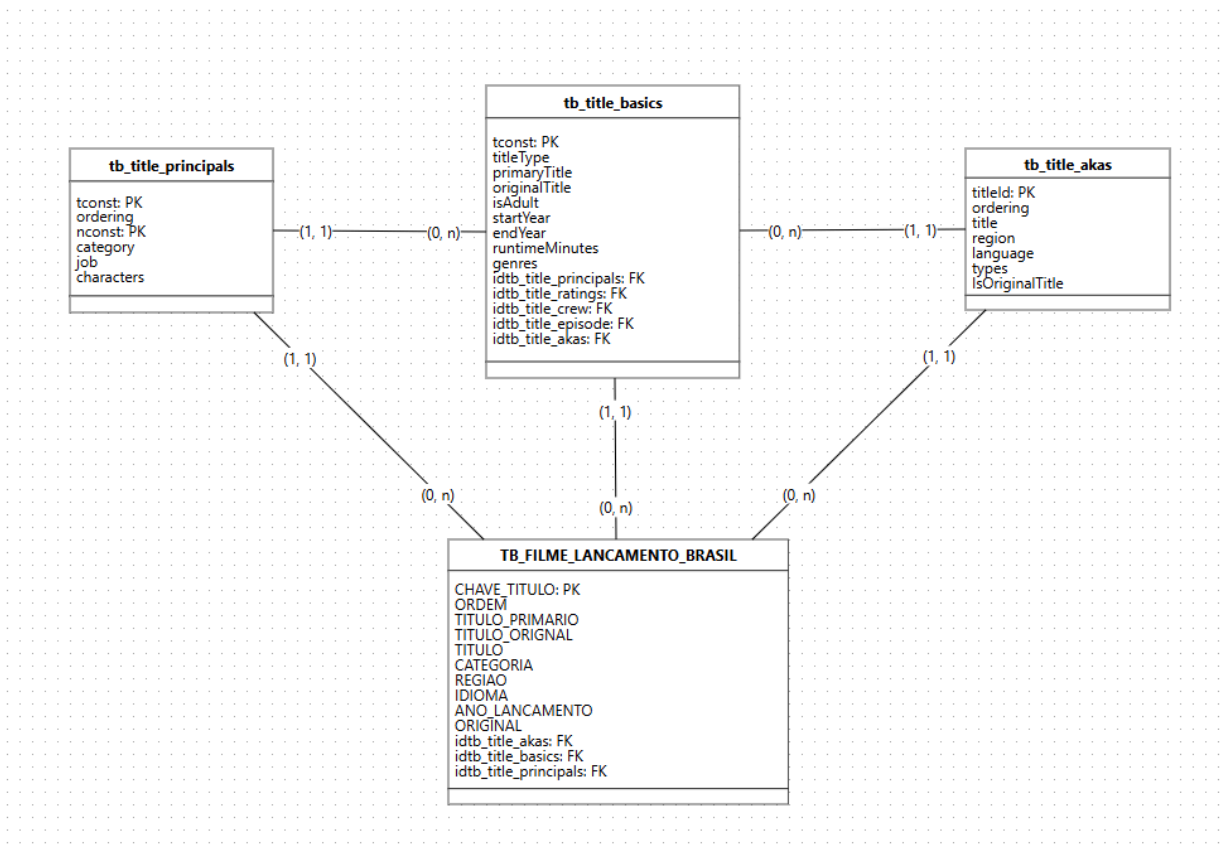
Dicionário de Dados (Camada Gold)

TB_TITULO_LANCAMENTO_BRASIL

Tabela com todos os título que sejam do tipo Movies, Short e Vídeo com lançamento no Brasil contendo as colunas:

- **CHAVE_TITULO** (coluna tconst da tabela imdb_db_silver.tb_title_basics): Identificador exclusivo do título.
- **ORDEM** (coluna ordering da tabela imdb_db_silver.tb_title_akas): Número para identificar exclusivamente as linhas para um determinado titleId na tabela imdb_db_silver.tb_title_akas.

- **TITULO_PRIMARIO** (coluna primaryTitle da tabela imdb_db_silver.tb_title_basics): Descrição do título popularmente conhecido ou usado pelos cineastas em materiais promocionais no seu lançamento.
- **TITULO_ORIGINAL** (coluna primaryTitle da tabela imdb_db_silver.tb_title_basics): Descrição do título original, no idioma original.
- **TITULO** (coluna title na tabela imdb_db_silver.tb_title_akas): Descrição do título no Brasil.
- **CATEGORIA** (coluna titleType na tabela imdb_db_silver.tb_title_basics): Formato ou tipo do título, se é um filme (movie), curta (short) ou vídeo (video).
- **REGIAO** (coluna region na tabela imdb_db_silver.tb_title_akas): Contendo sigla com 2 caracteres da região de lançamento do título
- **IDIOMA** (coluna language na tabela imdb_db_silver.tb_title_akas): Contendo sigla com dois caracteres do idioma.
- **ANO_LANCAMENTO** (Coluna startYear na tabela imdb_db_silver.tb_title_basics)
- **ORIGINAL** (coluna isOriginalTitle na tabela imdb_db_silver.tb_title_akas): 1 indica que o título é original e 0 indica que o título não é original.



Relacionamento da tabela IMDB_DB_GOLD.TB_LANCAMENTO_BRASIL

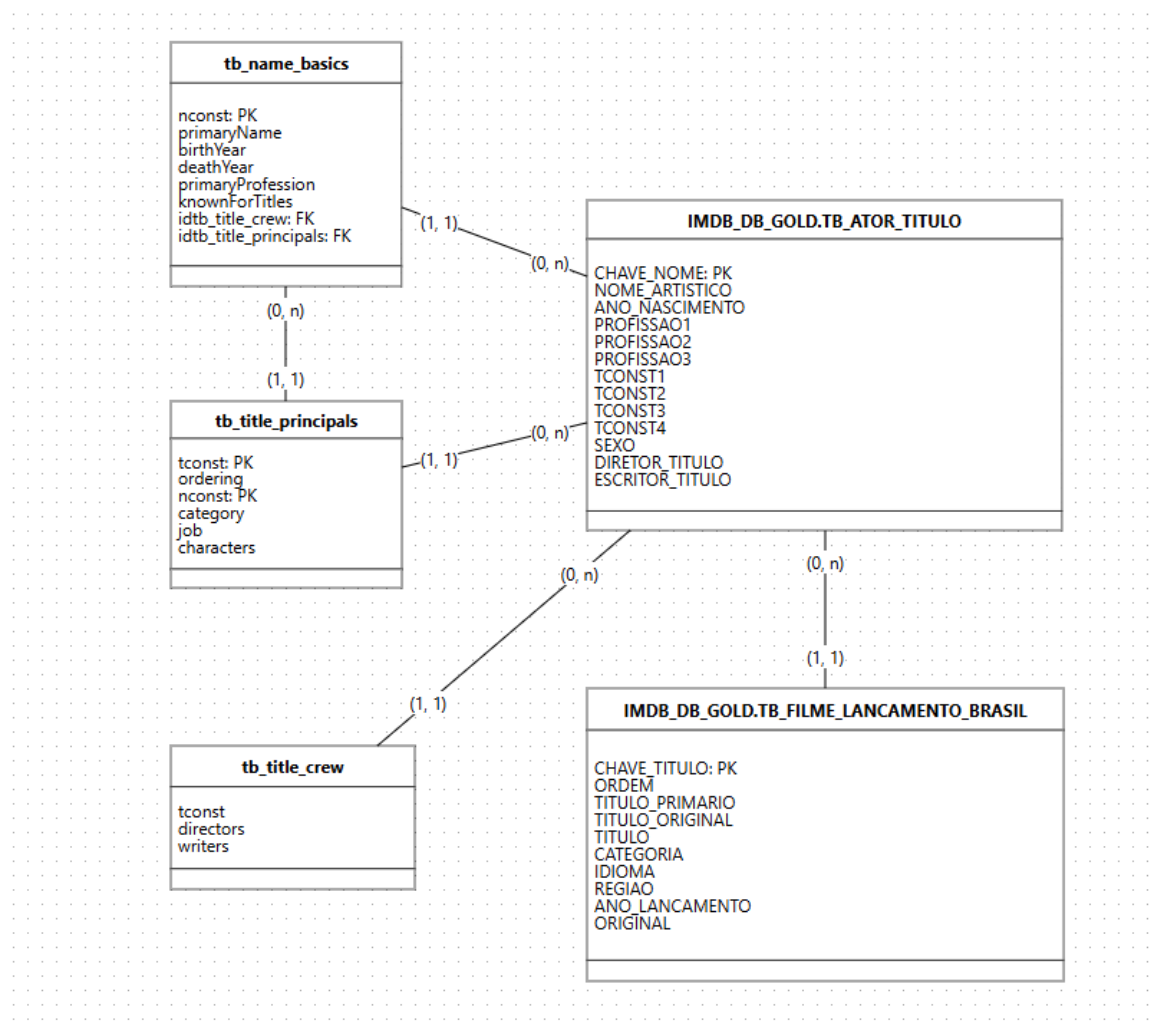
TB_ATOM_TITULO

OBSERVAÇÃO

- Tabela criada com a finalidade de diminuir a carga para listagem de títulos ou profissões vinculadas aos atores, pois as colunas primaryProfession e knowForTitles é um array em formato string separado por vírgula.
- Desta forma a busca direta sem "separar" as informações dos arrays sobrecarrega a consulta sem necessidade.

Tabela com todas informações de pessoas que tenham participado de algum título, que sejam do tipo Movies, Short e Vídeo contendo as colunas:

- **CHAVE_NOME** (coluna tconst da tabela imdb_db_silver.): Identificador exclusivo do título.
- **NOME_ARTISTICO** (coluna primaryName da tabela imdb_db_silver.tb_name_basics): Descrição do título popularmente conhecido ou usado pelos cineastas em materiais promocionais no seu lançamento.
- **ANO_NASCIMENTO** (coluna birthYear da tabela imdb_db_silver.tb_name_basics): Ano de nascimento do Ator/Atriz.
- **PROFISSAO[NÚMERO]** (coluna primaryProfession da tabela imdb_db_silver.tb_name_basics): Convertida em 3 colunas, onde cada coluna contém a descrição de uma profissão que o ator exerce.
- **TCONST[NÚMERO]** (coluna knowForTitles da tabela imdb_db_silver.tb_name_basics): Convertida em 4 colunas, onde cada coluna possui a tconst (CHAVE_TITULO) de um título que o ator colaborou.
- **SEXO** (coluna): Informa com valor "M" se for Masculino, "F" se for Feminino ou "-" se não houver informação..
- **DIRETOR_TITULO** (coluna directors da tabela imdb_db_silver.tb_title_crew): Informa com "S" se a pessoa foi um dos diretores no Título.
- **ESCRITOR_TITULO** (coluna writers da tabela imdb_db_silver.tb_title_crew): Informa com "S" se a pessoa foi um dos escritores do Título.



Relacionamento da tabela IMDB_DB_GOLD.TB_ATOM_TITULO

TB_ATOM_TITULO_BRASIL

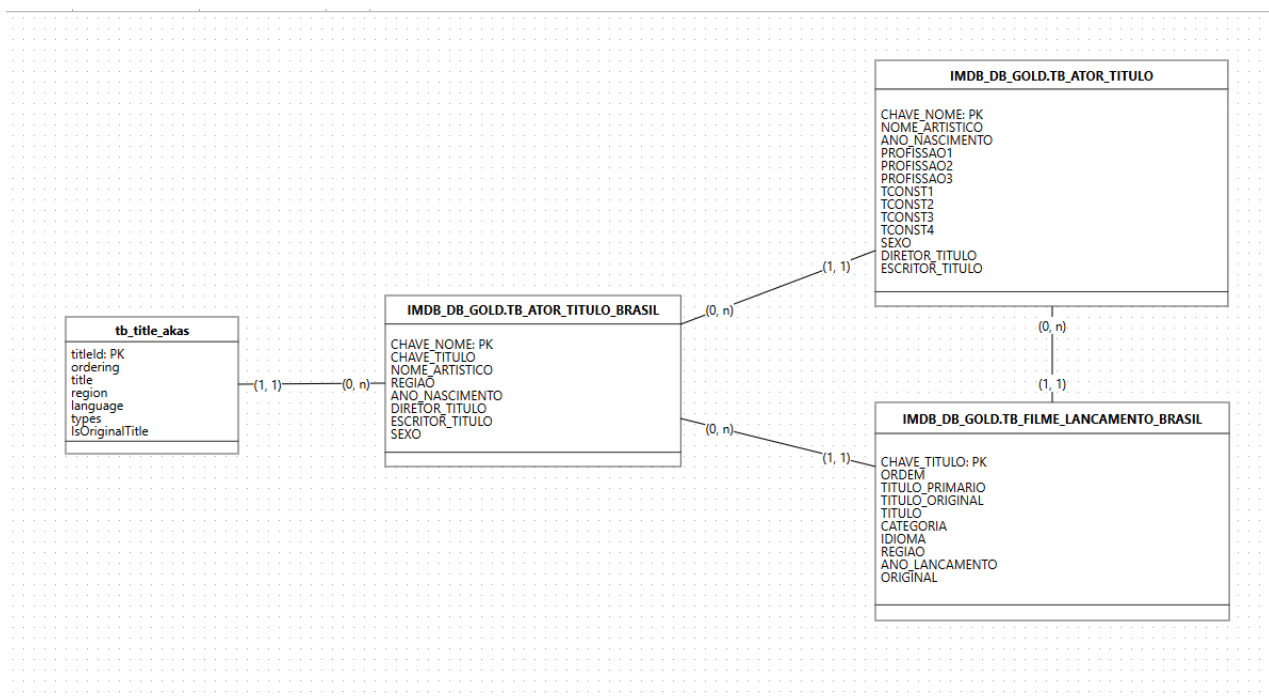
Lista dos atores que atuaram em um título lançado no Brasil do tipo Movie, Short ou Vídeo, contendo as colunas:

- **CHAVE_NOME** (Coluna nconst): Identificador exclusivo do ator/atriz.
- **CHAVE_TITULO** (Coluna nconst): Identificador exclusivo do título que o ator/atriz participou.
- **NOME_ARTISTICO** (coluna NOME_ARTISTICO da tabela IMDB_DB_GOLD.TB_ATOM_TITULO): Nome artístico do ator/atriz.
- **REGIAO** (coluna region da tabela imdb_db_silver.tb_title_akas): Região (país) com dois caracteres em que o título foi lançado.

- **ANO_NASCIMENTO** (coluna ANO_NASCIMENTO da tabela IMDB_DB_GOLD.TB_ATOM_TITULO):
Ano de nascimento do ator/atriz.
- **DIRETOR_TITULO** (coluna DIRETOR_TITULO da tabela IMDB_DB_GOLD.TB_ATOM_TITULO):
Informa com "S" se a pessoa foi um dos diretores do Título.
- **ESCRITOR_TITULO** (coluna ESCRITOR_TITULO da tabela IMDB_DB_GOLD.TB_ATOM_TITULO):
Informa com "S" se a pessoa foi um dos escritores do Título.
- **SEXO** (coluna SEXO da tabela IMDB_DB_GOLD.TB_ATOM_TITULO):
Informa "F" para Feminino, "M" para Masculino ou "-" quando não há informação.

OBSERVAÇÃO

- As colunas primaryProfession consistem em apenas 3 elementos.
- Foi realizada a consulta com a imdb_db_silver.tb_name_basics verificando a quantidade máxima de títulos por Ator/Atriz.
- Desta forma a lista foi separada em colunas com a finalidade de facilitar consultas com valor do campo "cheio" (completo).

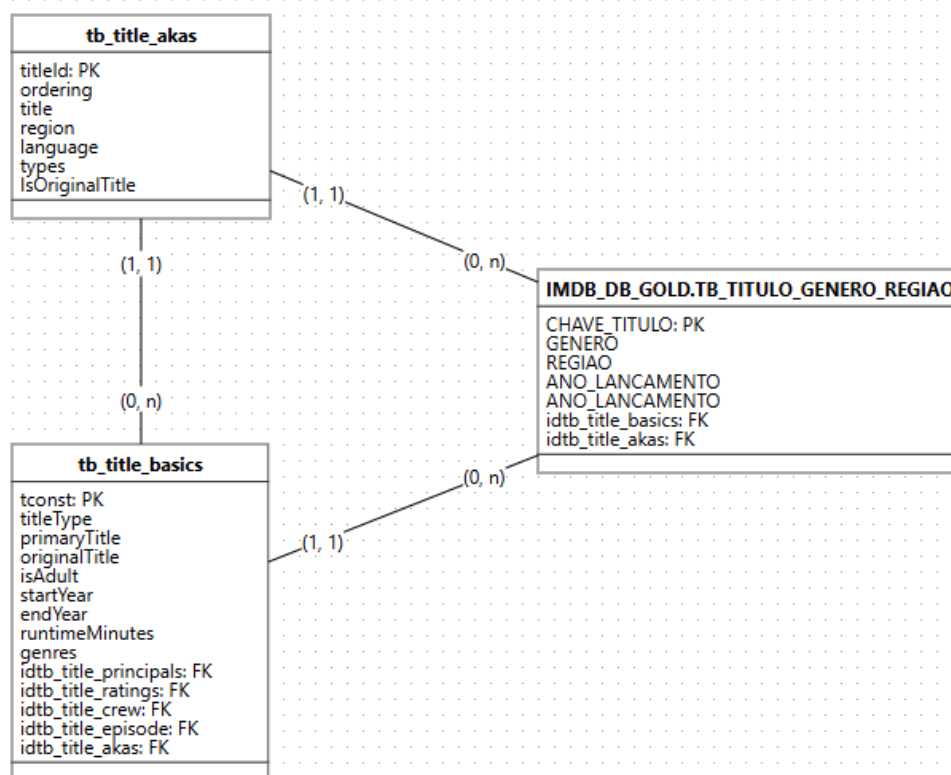


Relacionamento da tabela IMDB_DB_GOLD.TB_ATOM_TITULO_BRASIL

TB_TITULO_GENERO_REGIAO

Lista de todos os Gêneros de Títulos, disponíveis na base fornecida pelo IMDB por título e região de lançamento, contendo as seguintes colunas:

- **CHAVE_TITULO** (coluna tconst da tabela imdb_db_silver.tb_title_basics):
Identificador exclusivo do título.
- **GENERO** (coluna genres da tabela imdb_db_silver.tb_title_basics):
Informa o nome do gênero do título.
- **REGIAO** (coluna region da tabela imdb_db_silver.tb_title_akas):
Informa a região que o título foi lançado.
- **ANO_LANCAMENTO** (coluna startYear da tabela imdb_db_silver.tb_title_basics):
Informa ano de lançamento do título.



Relacionamento da tabela IMDB_DB_GOLD.TB_TITULO_GENERO_REGIAO

Análises e Resultados

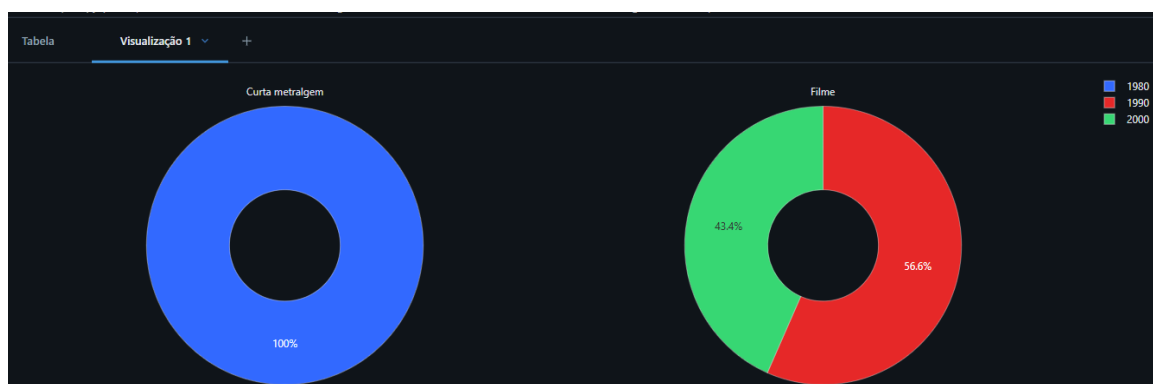
Banco de Dados: IMDB_RELATORIO

Resposta às perguntas mencionadas no Objetivo

Conforme mencionado no tópico objeto, com o tratamento dos dados na Camada Silver e na Camada Gold foi possível responder às perguntas.

A fonte de dados de todas as respostas são visões (views) materializadas em IMDB_RELATORIO com as tabelas da Camada Gold.

1 - Quantidade de títulos lançados no Brasil por categoria e década (80, 90, etc.)?



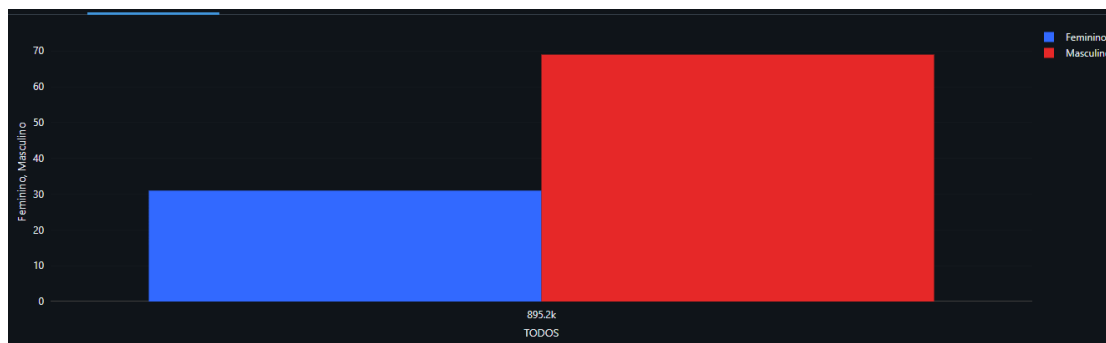
2 - Qual categoria teve mais lançamentos no Brasil por década?



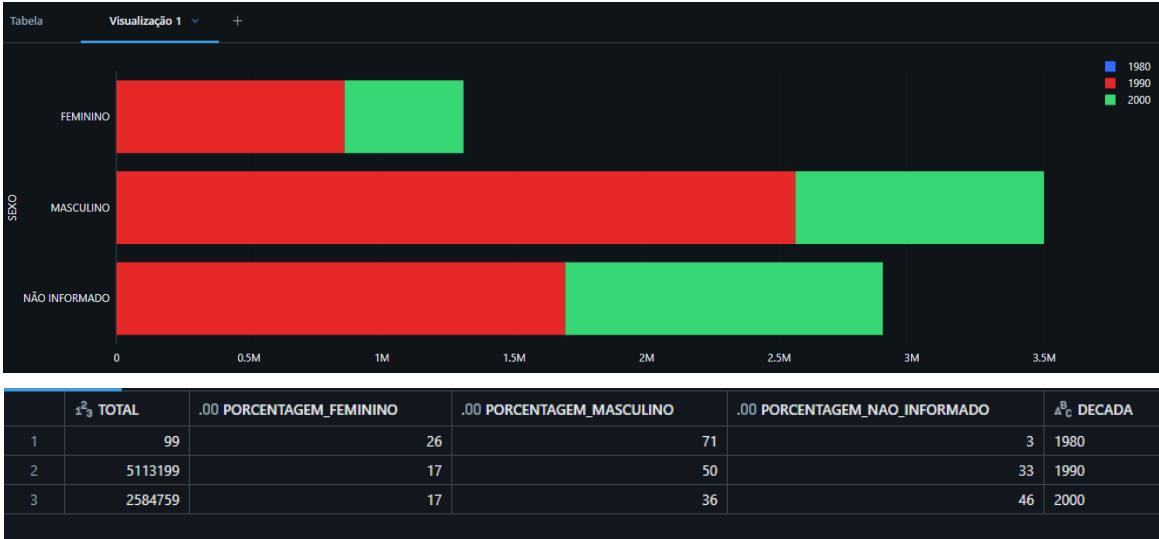
3 - Todos os títulos que possuem nome exclusivo no Brasil (Com nome diferente do nome original ou comercial).

	TÍTULO_BRASIL	TÍTULO_ORIGINAL	TÍTULO_COMERCIAL	ANO_LANC
1	#ALIVE	#SARAITDA	#ALIVE	
2	#ERADOSGIGANTES	#ERADOSGIGANTES	#ONTHESHOULDERSOFGIANTS	
3	#GAROTAS: O FILME	#GAROTAS: O FILME	#THOSEGIRLS	
4	#MANHOLE - DESVIO FATAL	#MANHOLE	#MANHOLE	
5	#OITAVO	OITAVO	OITAVO	
6	#PARTIUFAMA: CANCELADO NO AMOR	#PARTIUFAMA: CANCELADO NO AMOR	#FAMETIME	
7	#RESISTÊNCIA	#RESISTÊNCIA	#RESISTANCE	
8	#SEMFILTRO	#NOFILTER	#NOFILTER	
9	#SEMSAÍDA	#FOLLOWME	#FOLLOWME	
10	#SEMSAÍDA	FOLLOW ME	FOLLOW ME	
11	#SEM_FILTRO - NOITE SANGRENTA	#NO_FILTER	#NO_FILTER	
12	#SOBREVIDA	#SARAITDA	#ALIVE	
13	#UBERXTAXIS	#UBERXTAXIS	UBER VS. TAXI	
14	#BENNY: MEU FILHOTE FAVORITO	UGLY BENNY	UGLY BENNY	

4 - Qual a porcentagem de pessoas que atuaram como Ator ou Atriz em Títulos do tipo Filmes, Curtas ou Vídeo, lançados no Brasil?



8 - Porcentagem de atores do sexo feminino e masculino que atuaram nos títulos do tipo Filme em cada década existente na base.



6 - Filmes lançados no Brasil, onde o Diretor(a) atuou como Ator/Atriz.

	NOME ARTISTICO	TITULO_ORIGINAL	ANO_LANCAMENTO	REGIAO	PROFISSAO1	PROFISSAO2	PROFISSAO3
1	Philthy Phil Phillips	CITY DRAGON	1995	BR	director	actor	writer
2	A. Edward Sutherland	ABIE'S IRISH ROSE	1946	BR	director	producer	actor
3	A. Edward Sutherland	SKY DEVILS	1932	BR	director	producer	actor
4	A. Edward Sutherland	STEEL AGAINST THE SKY	1941	BR	director	producer	actor
5	A. Edward Sutherland	THE SAP FROM SYRACUSE	1930	BR	director	producer	actor
6	A.R. Murugadoss	GHAJINI	2008	BR	writer	director	actor
7	AJ Jones	THE CURSE OF LA PATASOLA	2022	BR	actor	writer	director
8	Aage Rais-Nordentoft	2 RYK OG EN AFLEVERING	2003	BR	director	writer	actor
9	Aakash Bhatia	LOOOOP LAPETA	2022	BR	director	writer	actor
10	Aarne Tarkas	NINA JA ERIK	1960	BR	director	writer	actor
11	Aaron Bertrand-Ruiz	ERÓTICO	2024	BR	director	actor	writer
12	Aaron Fisher	INSIDE THE RAIN	2019	BR	actor	director	writer
13	Aaron Harnick	30 DAYS	1999	BR	actor	writer	producer
14	Aaron Harnick	JUDY BERLIN	1999	BR	actor	writer	producer
15	Aaron Sorkin	A FEW GOOD MEN	1992	BR	writer	producer	actor

7 - Ranking dos diretores com títulos lançados no Brasil.

	^A _C NOME_ARTISTICO	¹ ₃ TOTAL
1	Pedro Murad	9
2	John Waters	8
3	Luciano Mello	8
4	Bruno de Oliveira	8
5	Bruno Costa	8
6	Marcelo Leme	8
7	Leonardo Martinelli	8
8	Anna Azevedo	7
9	George Miller	7
10	Joaquim Haickel	7
11	Maria Ribeiro	7
12	Paulo Miranda	7
13	Robert Gordon	7
14	Christian Caselli	7
15	Carlos Pronzato	6

Conclusão**Autoavaliação**

Por ter bastante familiaridade com SQL e python, foquei muito em testes de novas consultas, refazendo os dados e deveria também ter focado no resultado final para ter uma margem maior no desenvolvimento do MVP. E somente com a V0 (versão zero) finalizada poderia testar com calma.

Inicialmente achei as disciplinas um pouco morosa mas com o desenvolvimento do MVP, possibilitou conhecer um ambiente (Databricks) que permitiu colocar mais a “mão na massa” e gostar bastante no aprendizado que proporcionou.

Somente nas aulas de dúvidas que tomei conhecimento que ainda faltava muita coisa para confeccionar este MVP. Ficou o aprendizado para tenha mais atenção.