**Netflix Movies & TV Shows**

**1. Project Overview**

**Description:** This project involves building a dashboard to analyze Netflix's movies and shows. The project covers the end-to-end process, starting from data collection, cleaning, and transforming it into a structured format, followed by analysis using SQL and visualization in Power BI.

**Objective:** The main goal of this project is to create an insightful dashboard that helps users explore and analyze Netflix's content library.

**Key Features:**

* Data collection and cleaning
* Creating a Database
* Analysis using SQL
* Designing a star schema in Power BI
* Building powerful visualizations in Power BI

**2. Data Collection and Preparation**

**Data Sources:**

* Source data is acquired from Netflix’s public dataset ([Netflix Movie and TV Shows (June 2021)](https://www.kaggle.com/datasets/snehaanbhawal/netflix-tv-shows-and-movie-list)).
* The dataset includes attributes like Title, Type (Movie/TV Show), Release Year, Genres, Duration, Rating, Country, etc.

**Steps:**

**1. Downloading Data:**

* + Load the raw data from CSV.

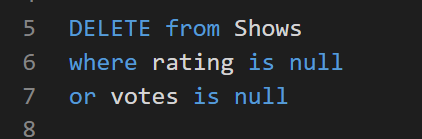
**2. Importing Data to SQL**

**Steps:**

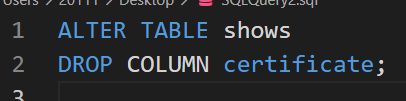
1. Load tables into SQL Server.
2. Data cleaning and transformation.

**Cleaning Steps:**

* **Handling Missing Data:** drop, or interpolate missing values.

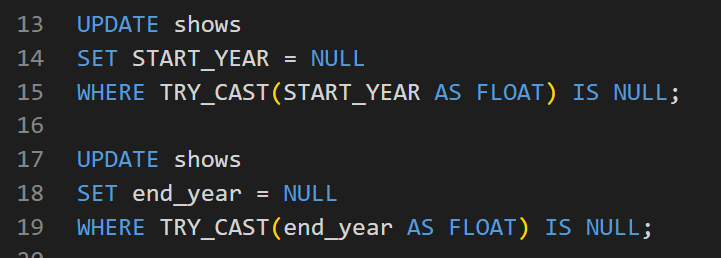


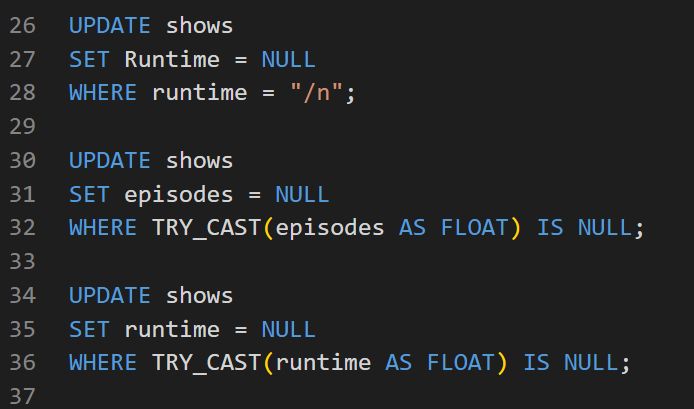
Deleting rows that has no rating or votes.



Dropping a column that was 66% null

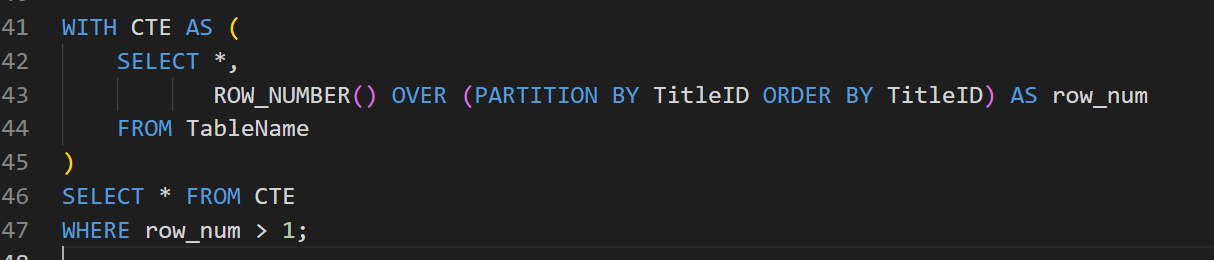
* **Cleaning random values:**

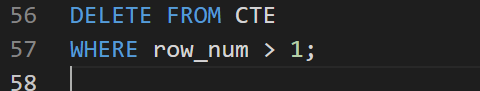




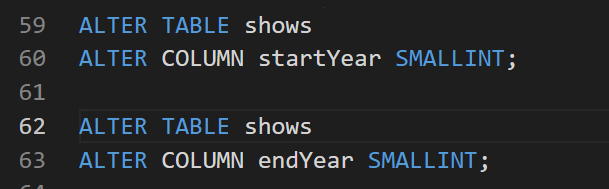
* **Duplicates:** Remove duplicate rows based on unique identifiers (like Title id).

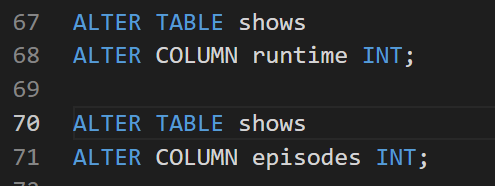
Checking duplicated rows

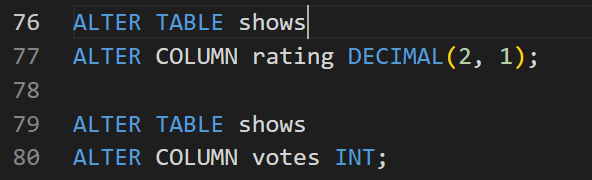




* **Data type:** setting the correct data type after cleaning

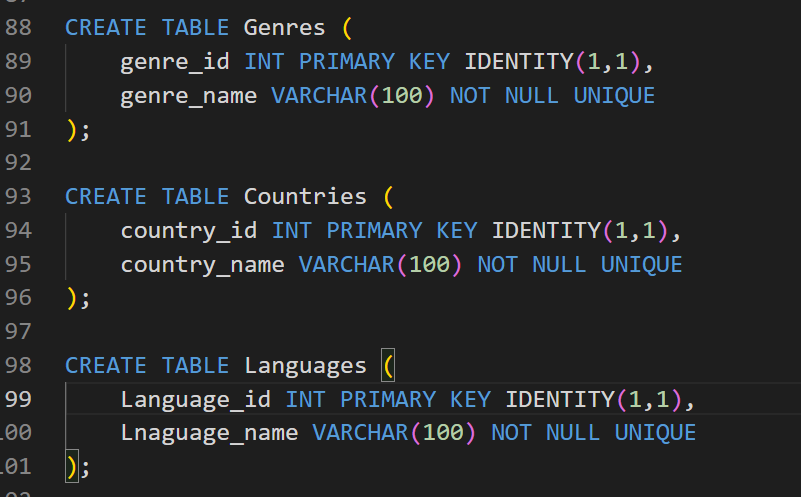


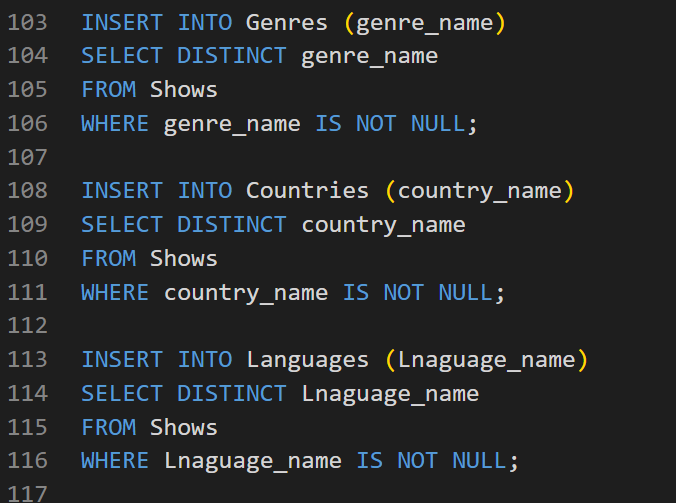




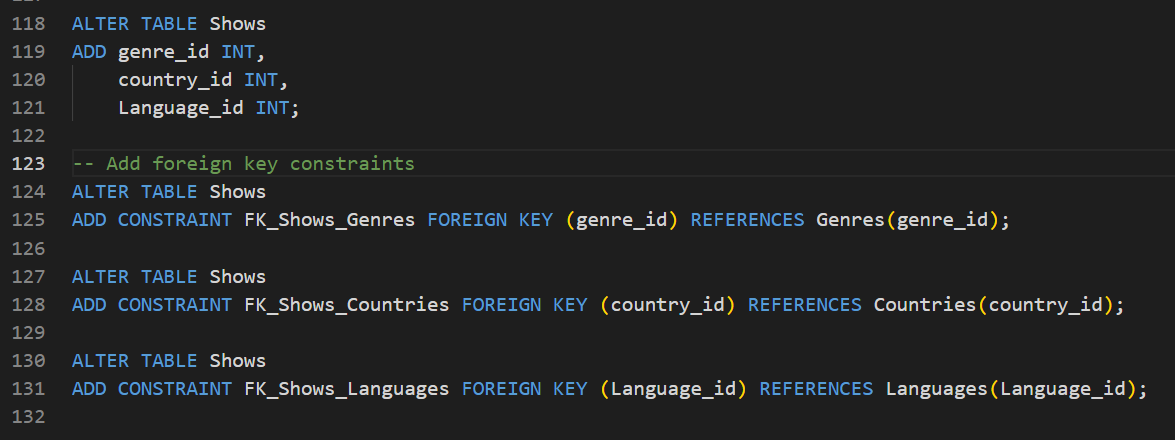
**Normalization:**

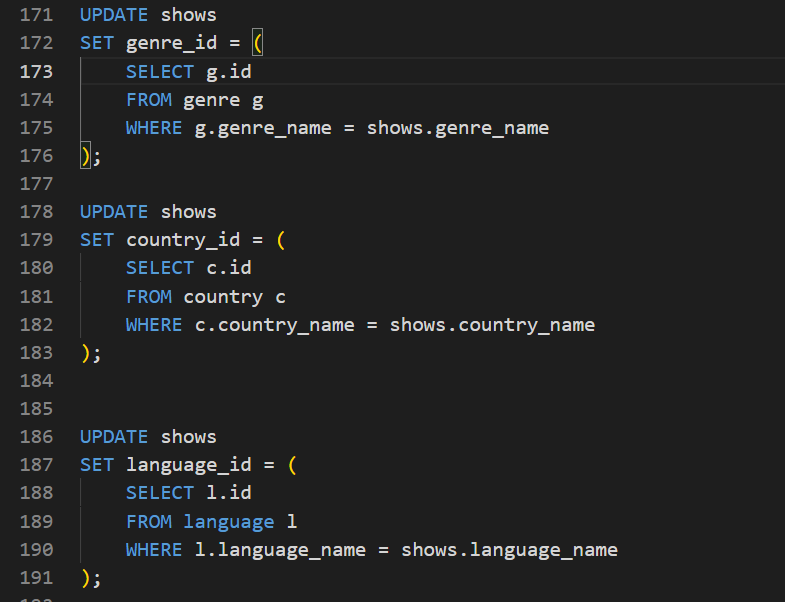
* Identify attributes with repetitive values, such as Genres, Country, and Language.
* Create separate dimension tables for these attributes to maintain unique values.
  + **Dimension Tables**: Tables like Genres, Countries, language.

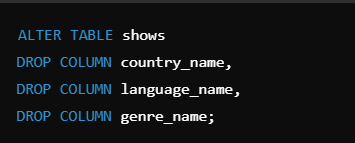




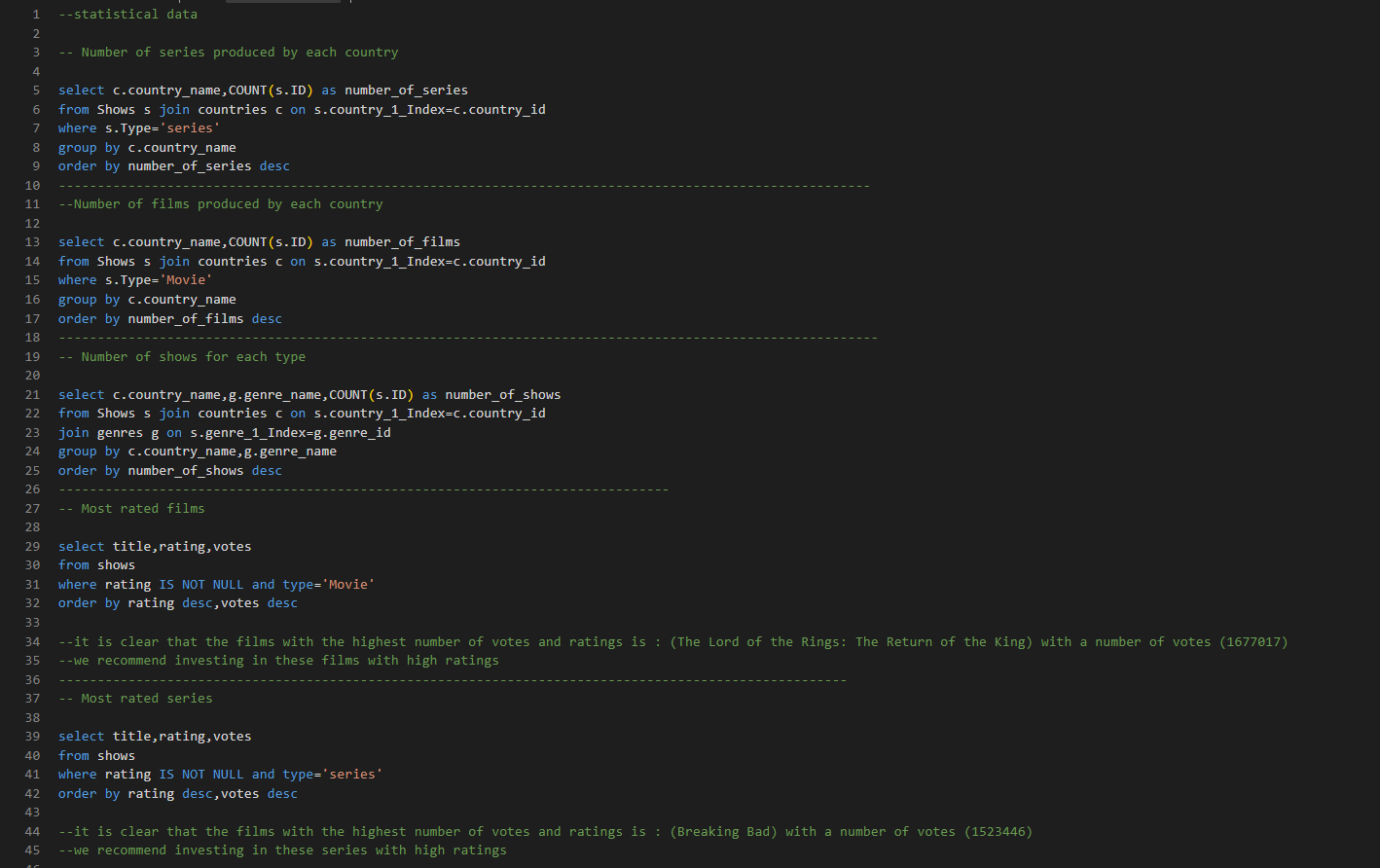
1. **Fact Table**: A central fact table that contains Title, Release Year, Duration, and foreign keys referencing the dimension tables

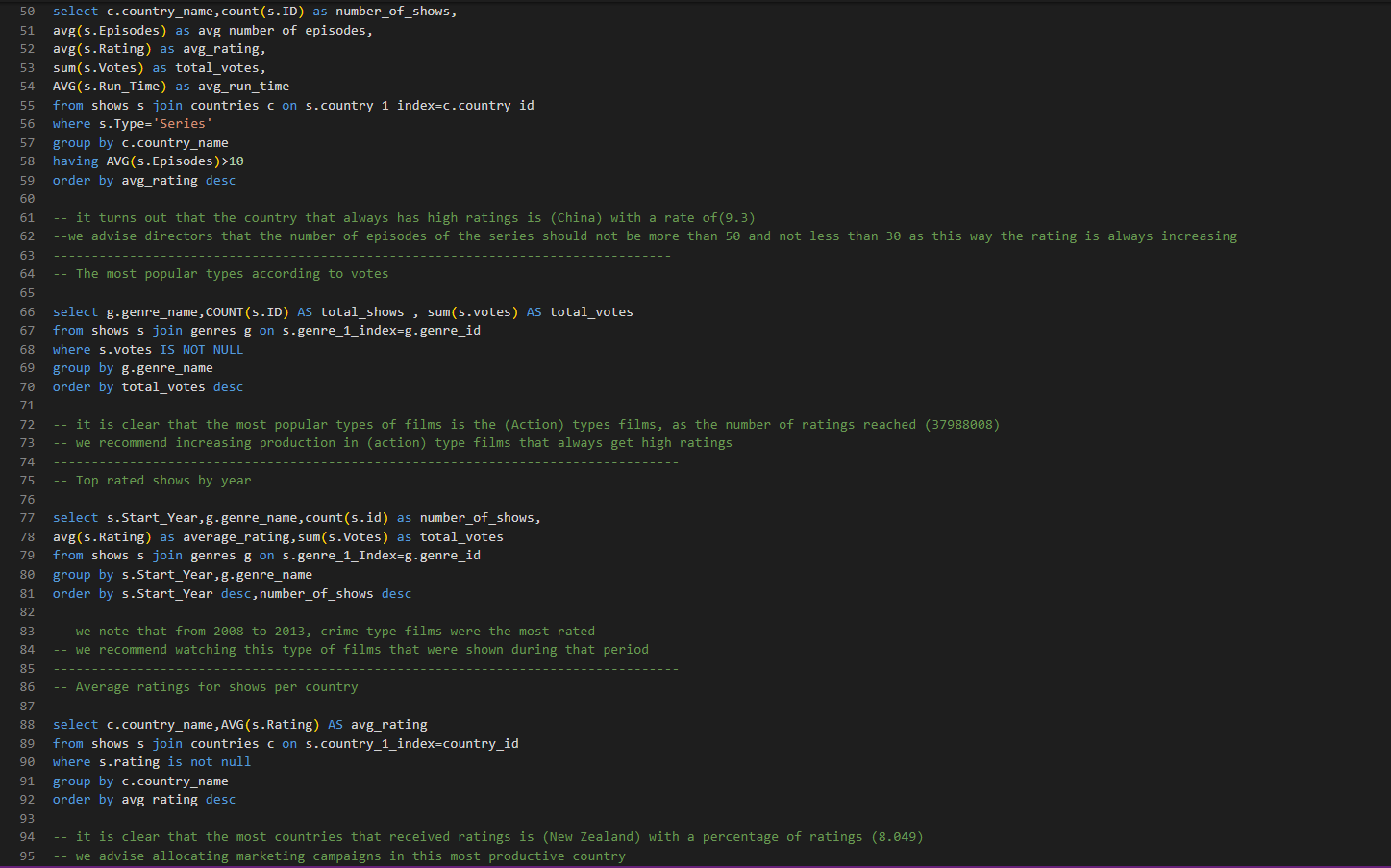


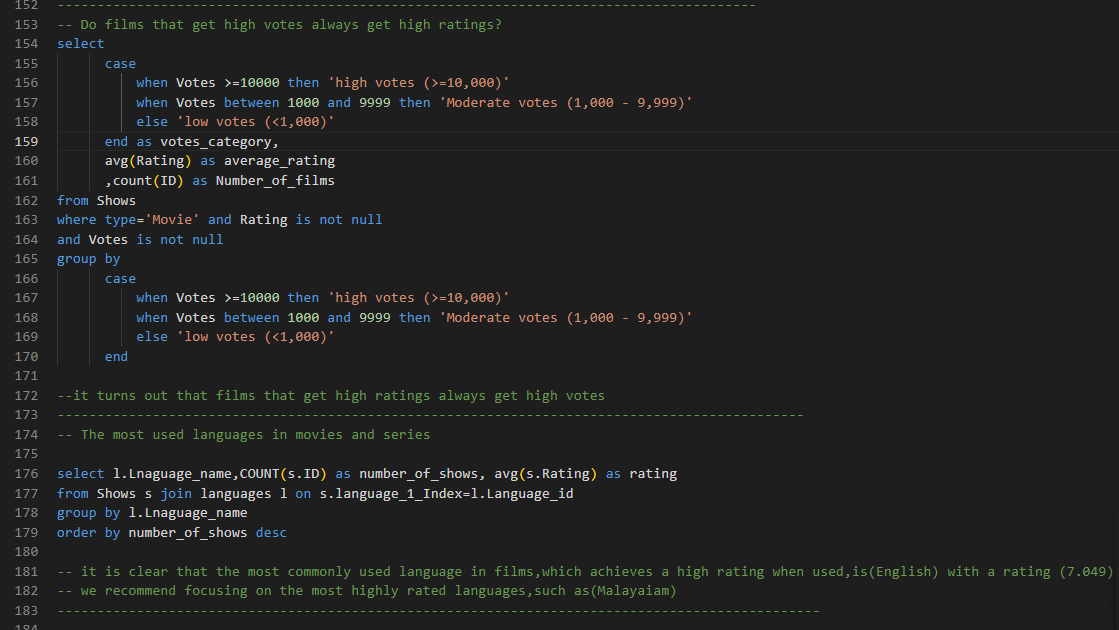
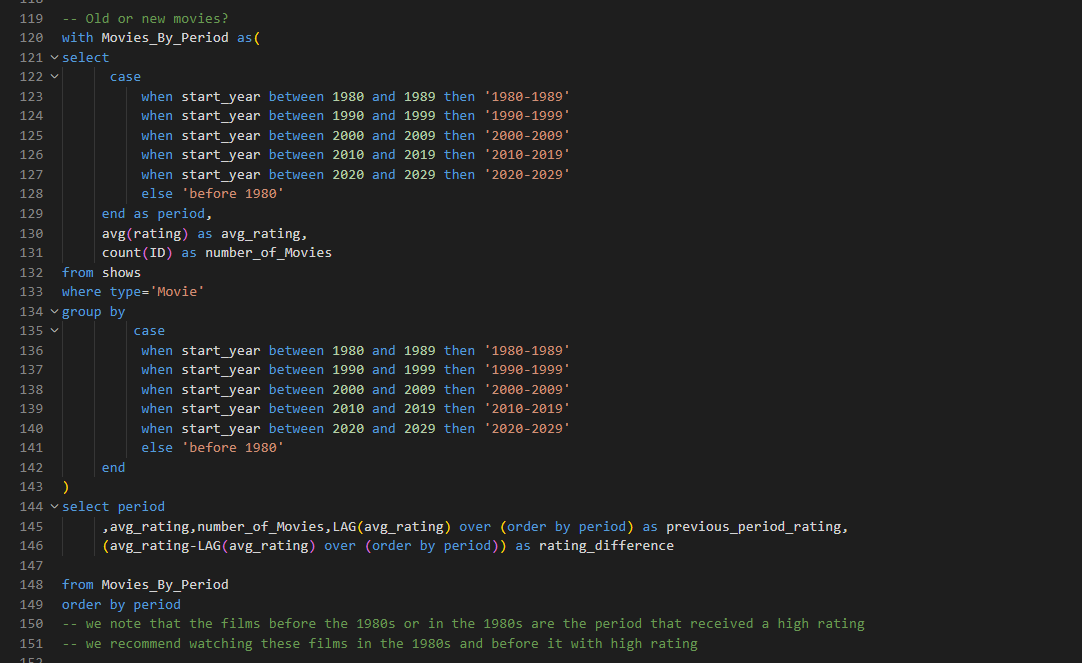
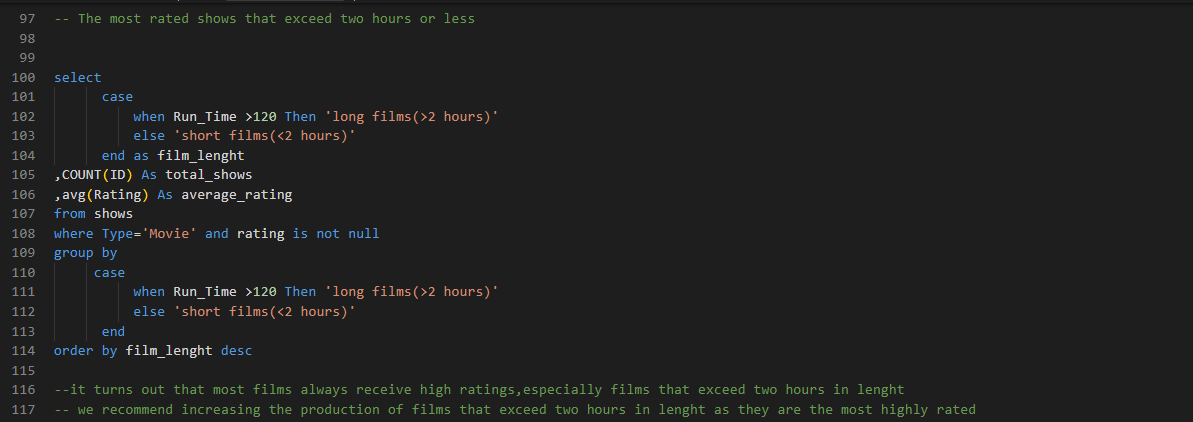




**SQL Queries:** Write SQL queries to analyze the data. Examples:







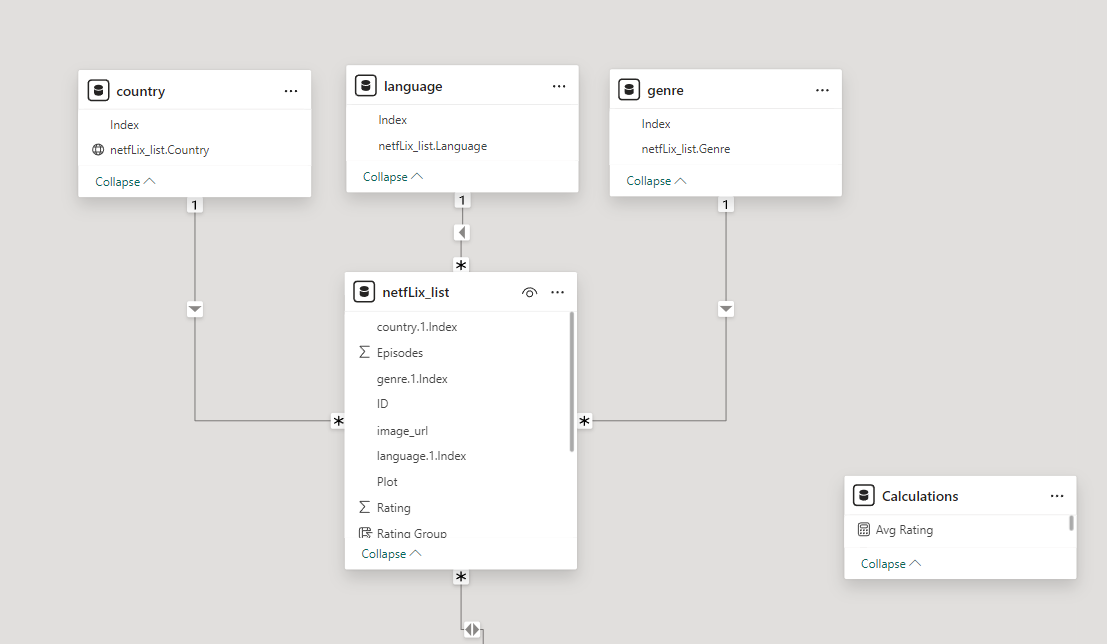
**5. Star Schema Design in Power BI**

**Schema Design:**

* Import the tables (fact and dimension tables) into Power BI.
* Design a **star schema** by linking the fact table to the dimension tables through foreign keys.
* Ensure proper relationships are established in the Power BI model for efficient querying.

**Example Schema:**

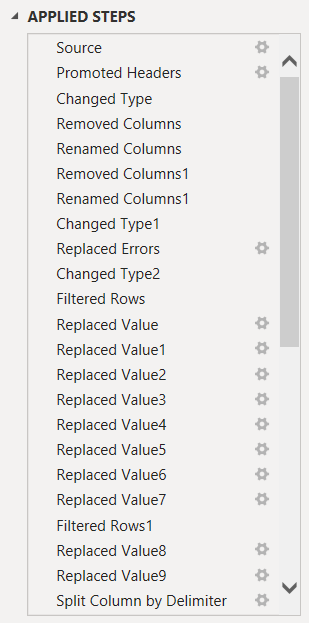
* **Shows** (Fact Table):
  + MovieID, Title, Type, ReleaseYear, Duration, GenreID, languageID, CountryID, etc.
* **Genres** (Dimension Table):
  + GenreID, GenreName
* **languages** (Dimension Table):
  + languadeID, RatingValue
* **Countries** (Dimension Table):
  + CountryID, CountryName



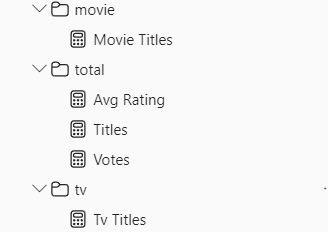
**6. Data Cleaning in Power BI**

**Data Modeling and Cleaning:**

* Use Power BI’s Query Editor to handle any additional cleaning tasks like:
  + Handling missing or null values.
  + Changing data types (e.g., date fields, numeric fields).
  + Replacing unwanted values
  + Splitting columns



* + Creating measures using DAX



Tv Titles = CALCULATE([Titles], FILTER(netfLix\_list, netfLix\_list[Type]= "Series"))

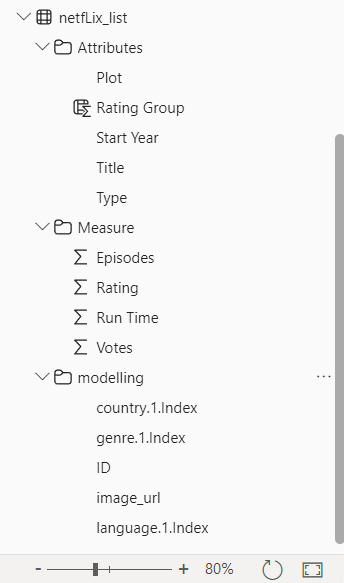
Votes = SUM(netfLix\_list[Votes])

Titles = DISTINCTCOUNT(netfLix\_list[ID])

Avg Rating = AVERAGE(netfLix\_list[Rating])

Movie Titles = CALCULATE([Titles], FILTER(netfLix\_list, netfLix\_list[Type]= "Movie")

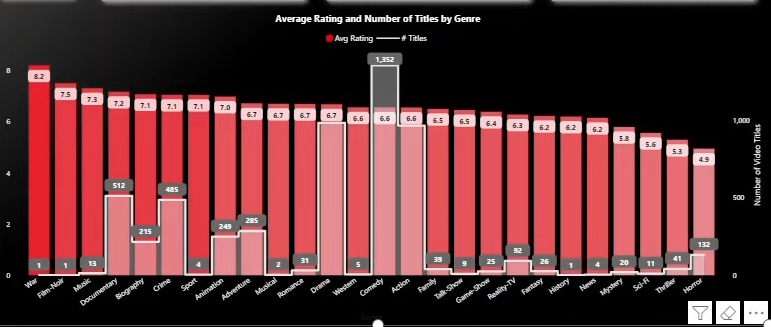
* + Organizing the fact table to folders (attributes, measures, modelling)



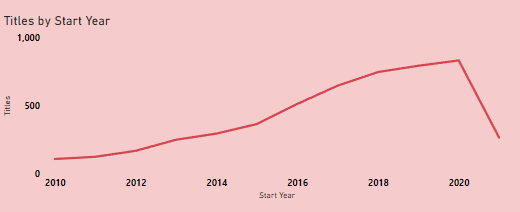
**7. Building Visuals in Power BI**

**Key Visualizations:**

* **Movie/Show Count and rating by Genre:** A bar chart showing the distribution of genres.

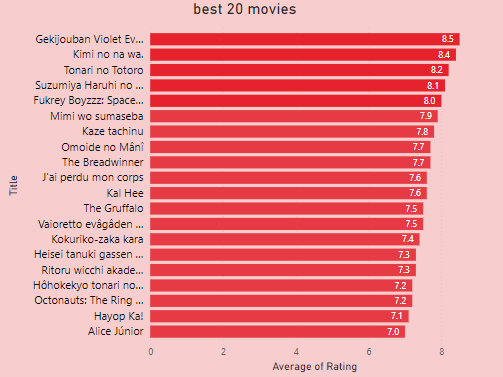


* **Production over Time:** A line chart visualizing production of Netflix content over time.

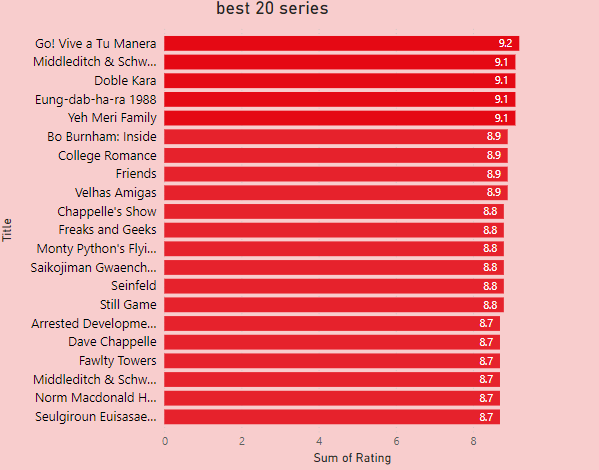


* **Top 20 Movies:** a bar chart showing top 20 content by average rating

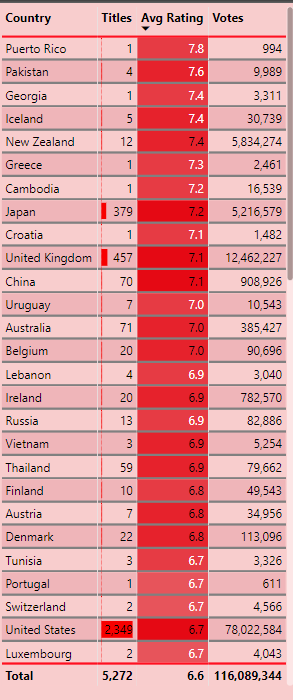
For movies:

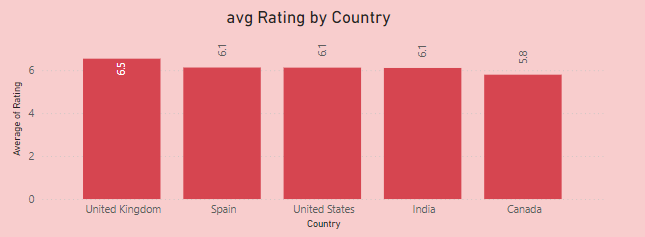


For series:

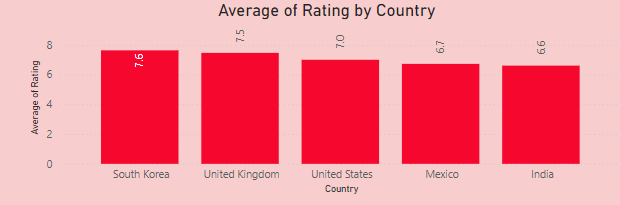


* **Top Countries by Content:** A table chart visualizing which countries produce the most content and has most rating.

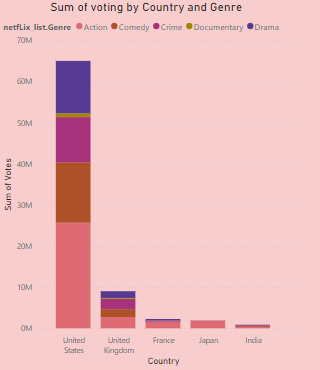


For Movies: 

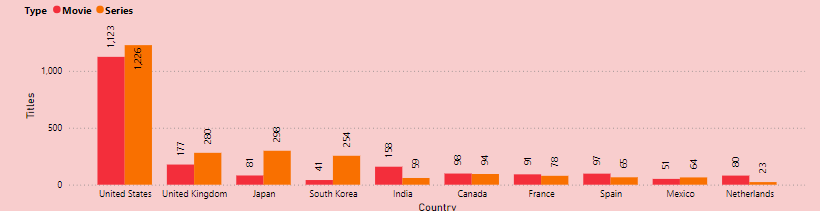
For Series:



By genre:

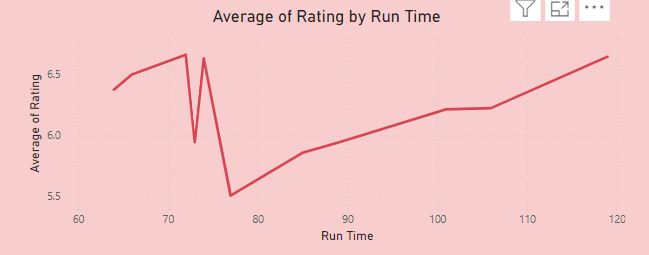


* **Content Type Comparison (Movies vs. TV Shows):** A stacked bar chart comparing movies vs. TV shows by country.

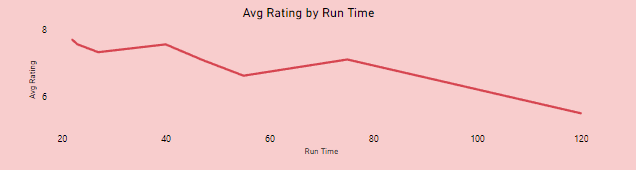


* **Duration Distribution:** A line chart describing the distribution of durations.

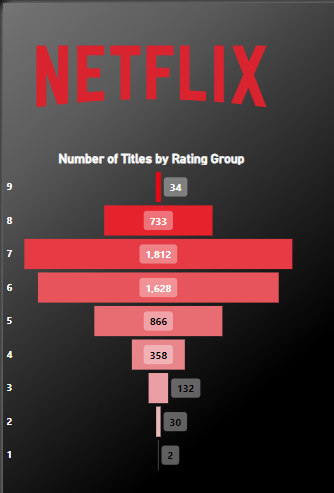
for movies:



For series:



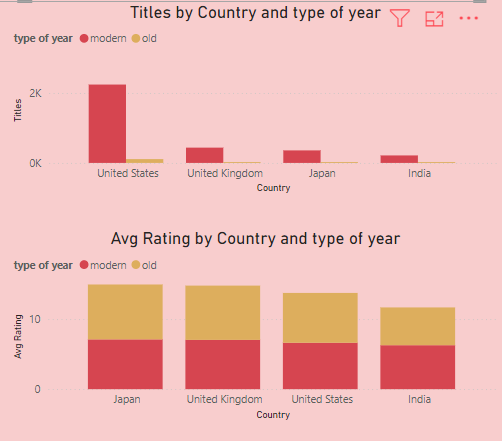
* **Number of titles by rating group:** A funnel showing number of titles for each rating group

****

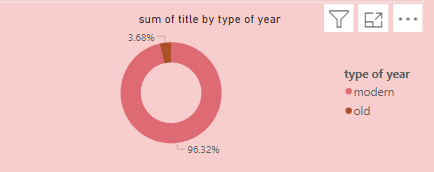
* **Cards for KPI’s:** showing number of titles for movies, TV shows and both

****

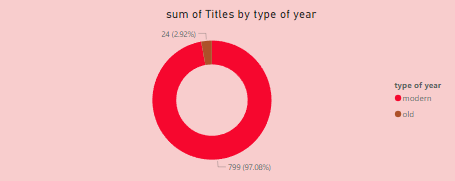
* **Old VS New content:** column charts comparing modern and old content by genre and country



**Number of titles for movies**

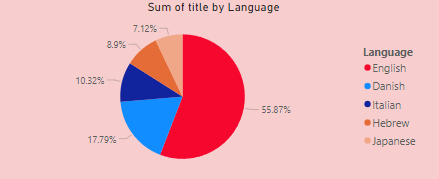


**Number of titles for series**

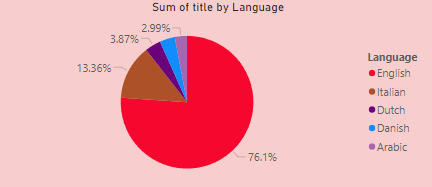


* **Number of titles by languages:** a pie chart describes the proportion of each language in the total titles.

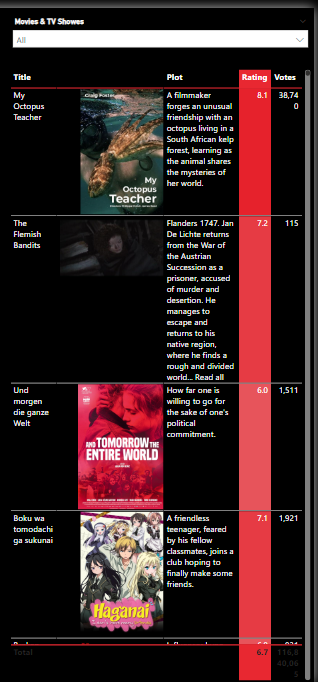
For movies



For series



* **Table showing content details:**

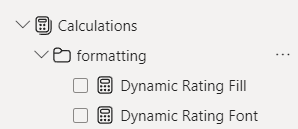
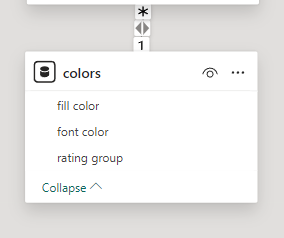
****

**Additional Features:**

* **Filters and Slicers:** Allow users to filter data by movies and tv shows



* **Interactive Dashboard:** the dashboard is dynamic, so users can click on a genre or country and see relevant data for that filter.
* **Creating table and measures to use dynamic coloring**

**8. Conclusion**

This dashboard allows users to analyze Netflix’s movies and shows by different dimensions such as genre, country, and language. By splitting the data into fact and dimension tables and using a star schema, the performance of the analysis is improved, and Power BI offers robust visualizations for deeper insights.