

NETFLIX ANALYSIS

Analysis Team

Hanin Baher

Ahmed Talaat

Yasmin Mohamed

Mohamed Khedr

Ayat Saad

Ali Hassan

Supervisor
Mohamed AbdelMawla

Project Overview

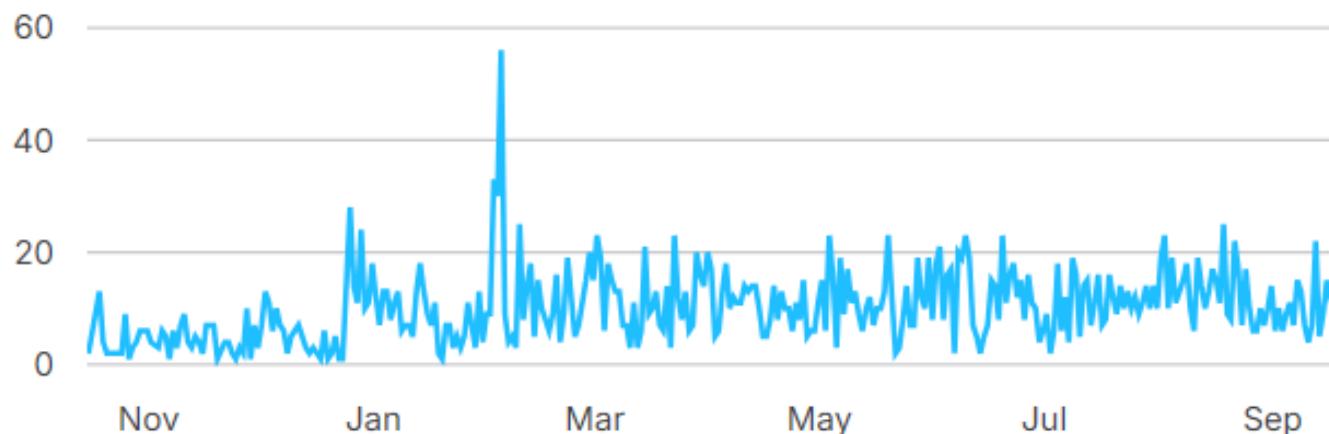
This project involved creating a well-structured relational database by extracting data, transforming it into a usable format, and importing it into a SQL Server. The objective was to ensure clean and accurate data to support further data analysis.





Views

All time ▾



Netflix Movie and TV Shows (June 2021)

Data of 7000 Netflix Movies and TV shows for Recommendation system

Data Card Code (3) Discussion (2) Suggestions (0)

About Dataset

Context

Dataset contains the list and metadata of all TV Shows and Movies available on Netflix currently about 7000 taken from the IMDB website.

Upvote if you liked it.

Data Collection & Analysis

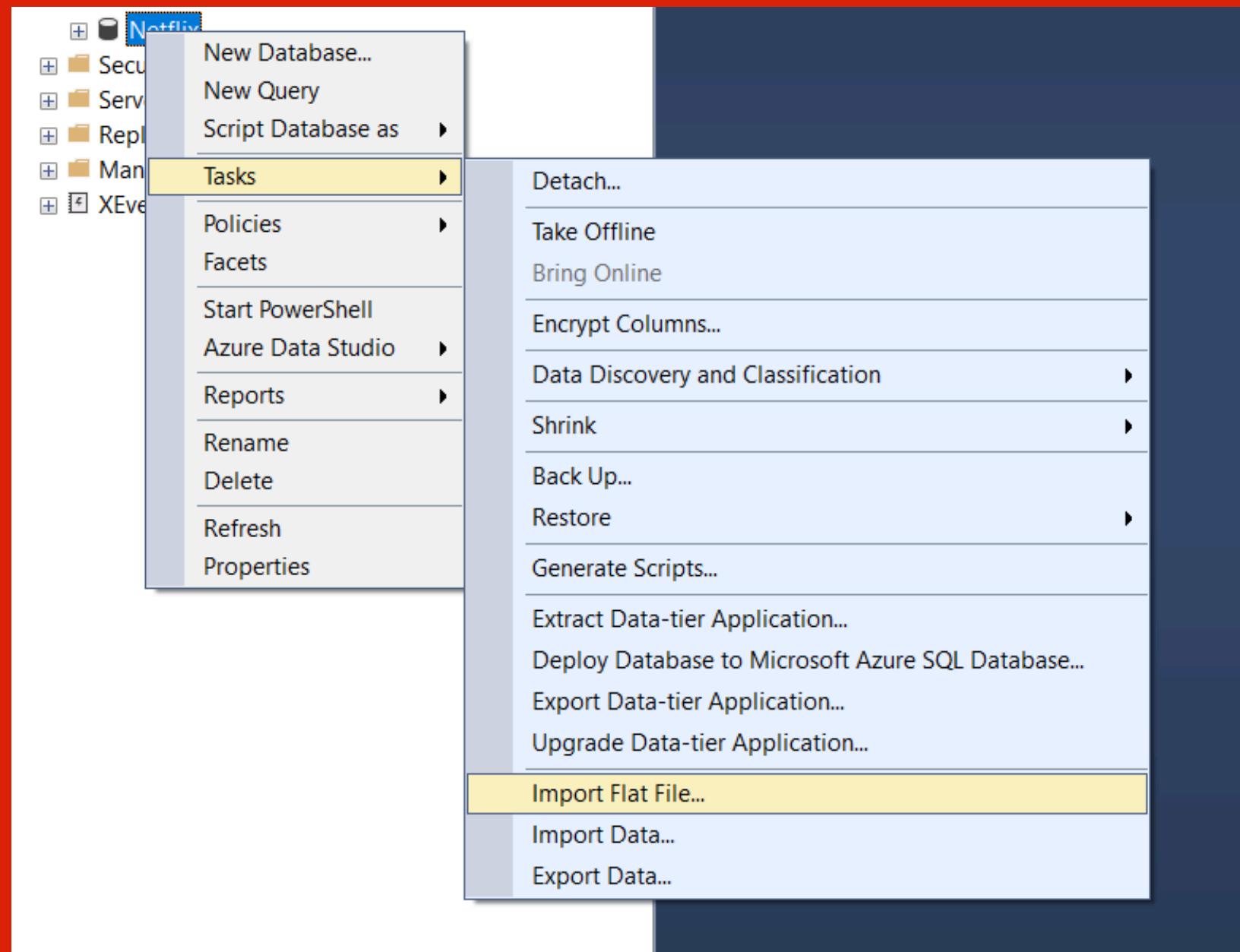
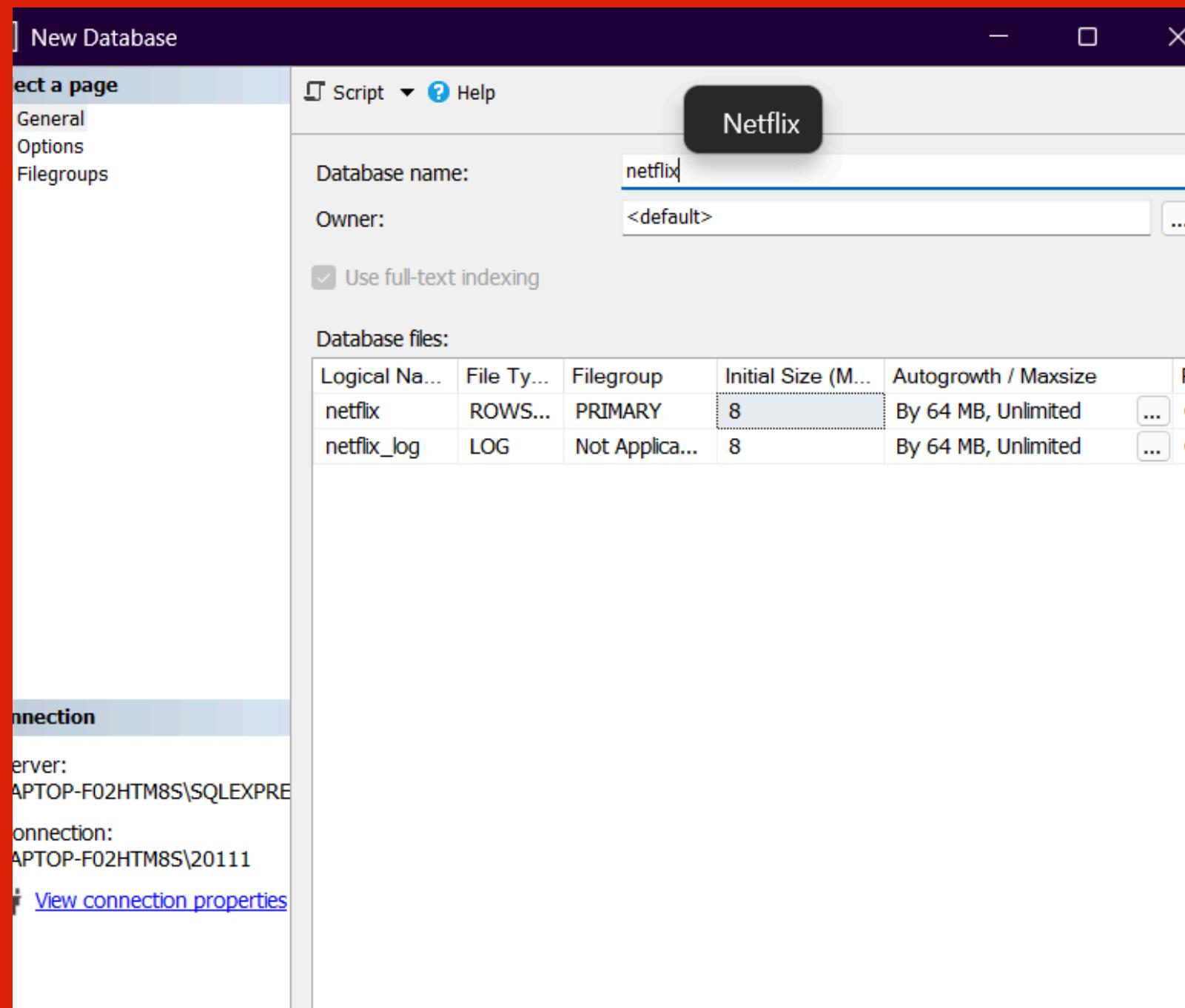
- 1- Extracting data into Excel sheets
- 2 Creating Database in SQL Server.
- 3 inserting data into SQL Server.
- 4- Cleaning numeric data and adjusting data types
- 5- Normalizing data by Creating dimension tables and constraints



2- inserting data into SQL Server.

1-Extracting data into Excel sheets

2-Creating DB



Cleaning Steps

Deleting rows that has no rating or votes.

```
5  DELETE from Shows  
6  where rating is null  
7  or votes is null
```

Handling Missing Data: drop, or interpolate missing values.

```
1  ALTER TABLE shows  
2  DROP COLUMN certificate;
```

Cleaning random values.

```
13 UPDATE shows  
14 SET START_YEAR = NULL  
15 WHERE TRY_CAST(START_YEAR AS FLOAT) IS NULL;  
16  
17 UPDATE shows  
18 SET end_year = NULL  
19 WHERE TRY_CAST(end_year AS FLOAT) IS NULL;  
20  
26 UPDATE shows  
27 SET Runtime = NULL  
28 WHERE runtime = "/n";  
29  
30 UPDATE shows  
31 SET episodes = NULL  
32 WHERE TRY_CAST(episodes AS FLOAT) IS NULL;  
33  
34 UPDATE shows  
35 SET runtime = NULL  
36 WHERE TRY_CAST(runtime AS FLOAT) IS NULL;
```



Cleaning Steps

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

setting the correct data type after cleaning

```
WITH CTE AS (
    SELECT *,
    ROW_NUMBER() OVER (PARTITION BY TitleID ORDER BY TitleID) AS row_num
    FROM TableName
)
SELECT * FROM CTE
WHERE row_num > 1;
```

```
56  DELETE FROM CTE
57 WHERE row_num > 1;
58 |
```

```
ALTER TABLE shows
ALTER COLUMN rating DECIMAL(2, 1);

ALTER TABLE shows
ALTER COLUMN votes INT;
```



Dimension Tables

Tables like Genres, Countries, language

```
88 CREATE TABLE Genres (
89     genre_id INT PRIMARY KEY IDENTITY(1,1),
90     genre_name VARCHAR(100) NOT NULL UNIQUE
91 );
92
93 CREATE TABLE Countries (
94     country_id INT PRIMARY KEY IDENTITY(1,1),
95     country_name VARCHAR(100) NOT NULL UNIQUE
96 );
97
98 CREATE TABLE Languages (
99     Language_id INT PRIMARY KEY IDENTITY(1,1),
100    Lnaguage_name VARCHAR(100) NOT NULL UNIQUE
101 );
```

```
103    INSERT INTO Genres (genre_name)
104    SELECT DISTINCT genre_name
105    FROM Shows
106    WHERE genre_name IS NOT NULL;
107
108    INSERT INTO Countries (country_name)
109    SELECT DISTINCT country_name
110    FROM Shows
111    WHERE country_name IS NOT NULL;
112
113    INSERT INTO Languages (Lnaguage_name)
114    SELECT DISTINCT Lnaguage_name
115    FROM Shows
116    WHERE Lnaguage_name IS NOT NULL;
117
```



Fact Table

A central fact table that contains Title, Release Year, Duration, and foreign keys referencing the dimension tables

```
118 ALTER TABLE Shows
119 ADD genre_id INT,
120     country_id INT,
121     Language_id INT;
122
123 -- Add foreign key constraints
124 ALTER TABLE Shows
125 ADD CONSTRAINT FK_Shows_Genres FOREIGN KEY (genre_id) REFERENCES Genres(genre_id);
126
127 ALTER TABLE Shows
128 ADD CONSTRAINT FK_Shows_Countries FOREIGN KEY (country_id) REFERENCES Countries(country_id);
129
130 ALTER TABLE Shows
131 ADD CONSTRAINT FK_Shows_Languages FOREIGN KEY (Language_id) REFERENCES Languages(Language_id);
132
```





Exploratory Data Analysis

Analysis using
SQL

Data cleaning

Designing a star schema
in Power BI

Building powerful
visualizations
in Power BI



Some Analyzing Queries

```
-- Number of series produced by each country

select c.country_name,COUNT(s.ID) as number_of_series
from Shows s join countries c on s.country_1_Index=c.country_id
where s.Type='series'
group by c.country_name
order by number_of_series desc

----- 

--Number of films produced by each country

select c.country_name,COUNT(s.ID) as number_of_films
from Shows s join countries c on s.country_1_Index=c.country_id
where s.Type='Movie'
group by c.country_name
order by number_of_films desc

----- 

-- Number of shows for each type

select c.country_name,g.genre_name,COUNT(s.ID) as number_of_shows
from Shows s join countries c on s.country_1_Index=c.country_id
join genres g on s.genre_1_Index=g.genre_id
group by c.country_name,g.genre_name
order by number_of_shows desc

----- 

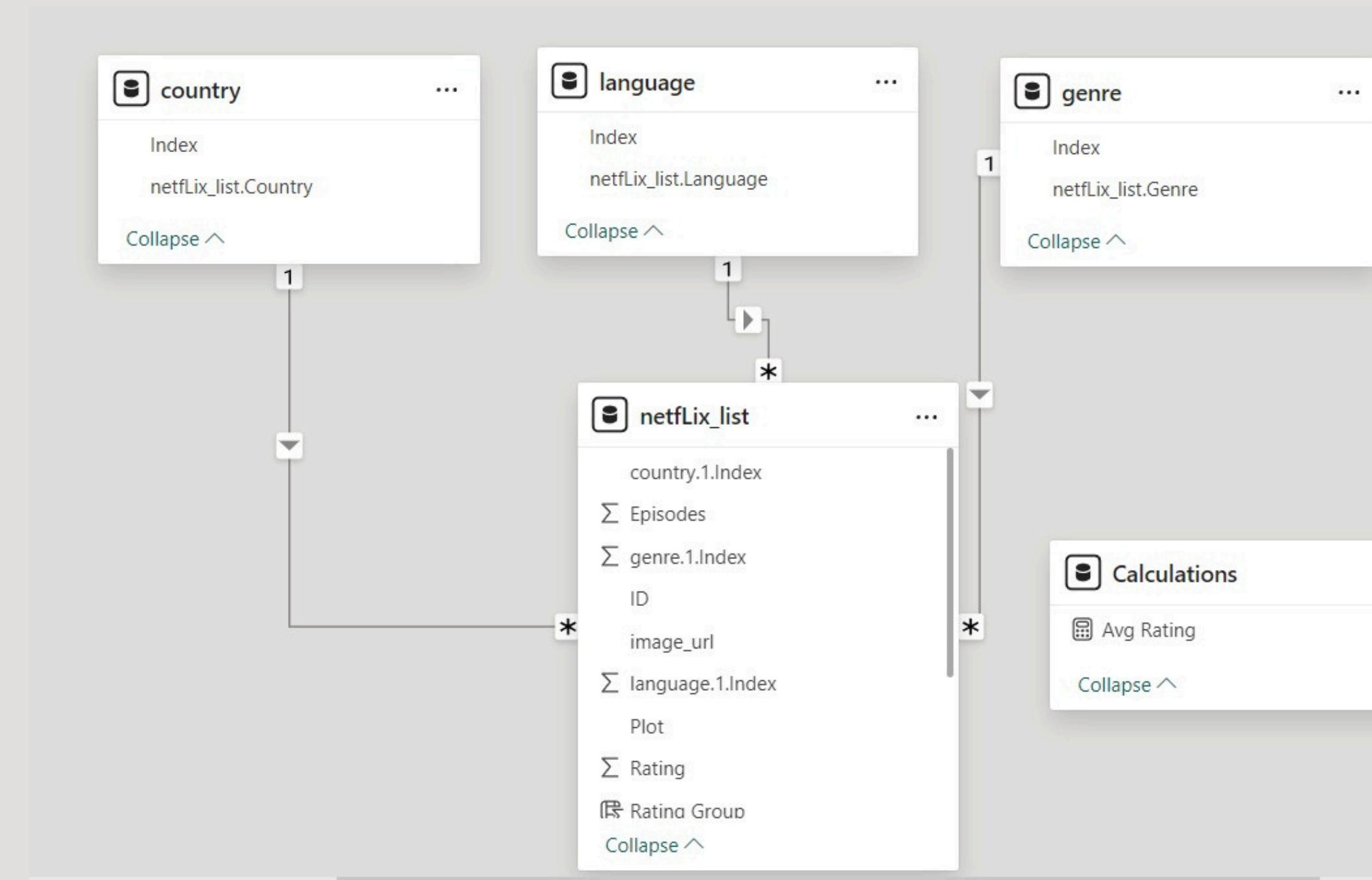
-- Most rated films

select title,rating,votes
from shows
where rating IS NOT NULL and type='Movie'
order by rating desc,votes desc
```



Data Model

Star Scheme

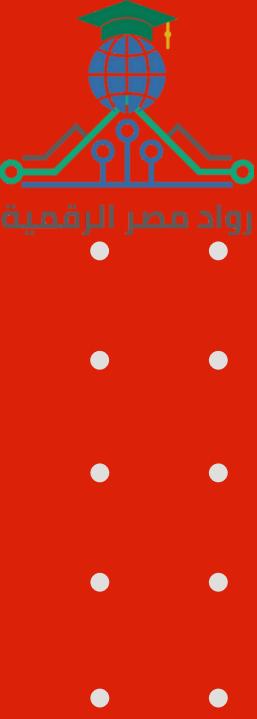


Data Cleaning In Power Bi

- **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**

◀ APPLIED STEPS

Applied Step	Action
1	Source
2	Promoted Headers
3	Changed Type
4	Removed Columns
5	Renamed Columns
6	Removed Columns1
7	Renamed Columns1
8	Changed Type1
9	Replaced Errors
10	Changed Type2
11	Filtered Rows
12	Replaced Value
13	Replaced Value1
14	Replaced Value2
15	Replaced Value3
16	Replaced Value4
17	Replaced Value5
18	Replaced Value6
19	Replaced Value7
20	Filtered Rows1
21	Replaced Value8
22	Replaced Value9
23	Split Column by Delimiter



Organizing Fact Table

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

The screenshot shows a hierarchical view of a fact table structure:

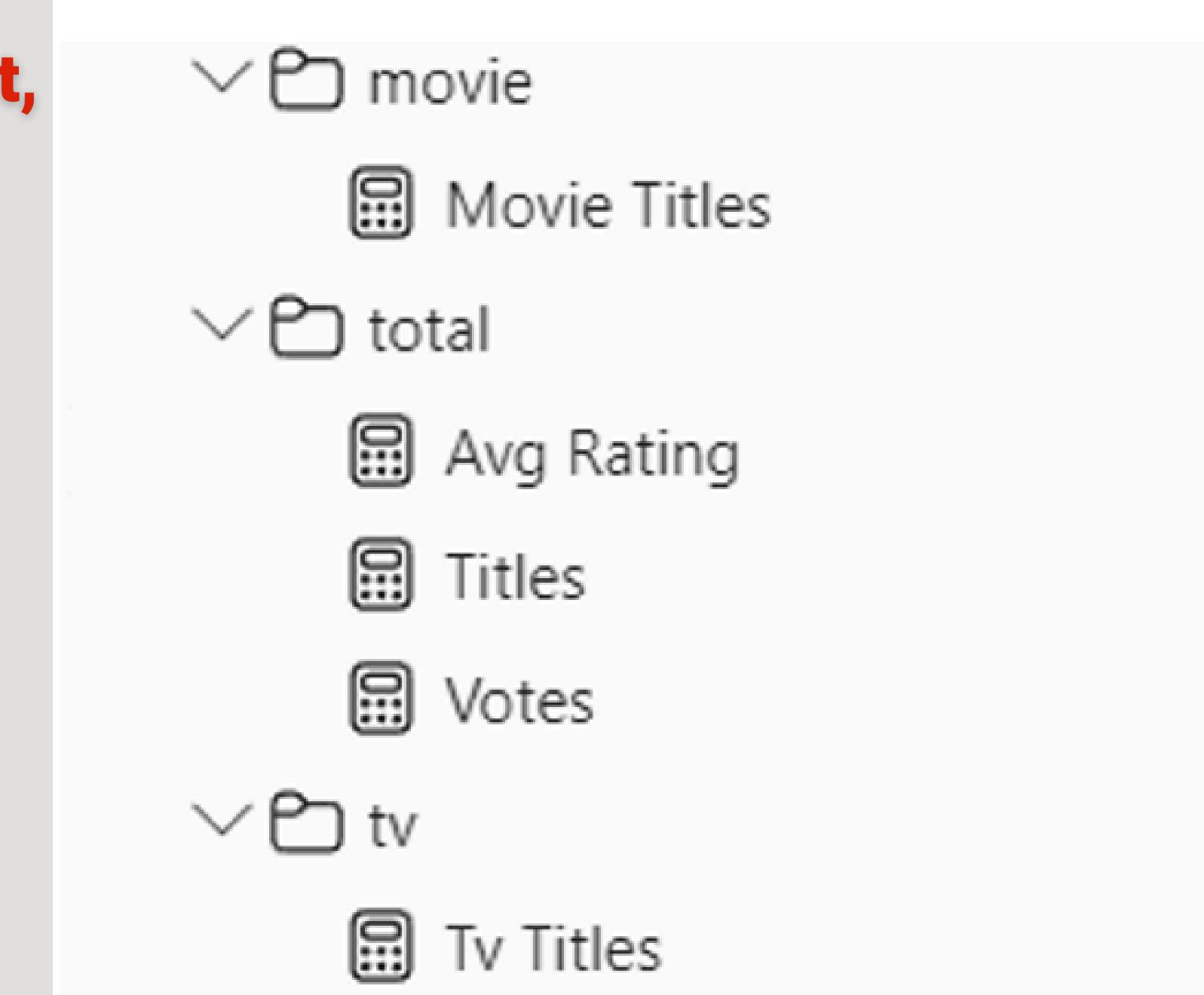
- Attributes**:
 - Plot
 - Rating Group
 - Start Year
 - Title
 - Type
- Measure**:
 - Σ Episodes
 - Σ Rating
 - Σ Run Time
 - Σ Votes
- modelling**:
 - country.1.Index
 - genre.1.Index
 - ID
 - image_url
 - language.1.Index

At the bottom of the interface, there are standard navigation controls: a zoom slider set to 80%, a refresh button, and a full-screen button.



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"))**





Dashboards

Categorization

NETFLIX

Movie
Series

num of shows
5595

movie
2645

TV series
2950

Number of Titles by Rating Group

Rating Group	Number of Titles
9+	34
8+	733
7+	1,812
6+	1,628
5+	866
4+	358
3+	132
2+	90
1+	2
0	

Titles by Start Year

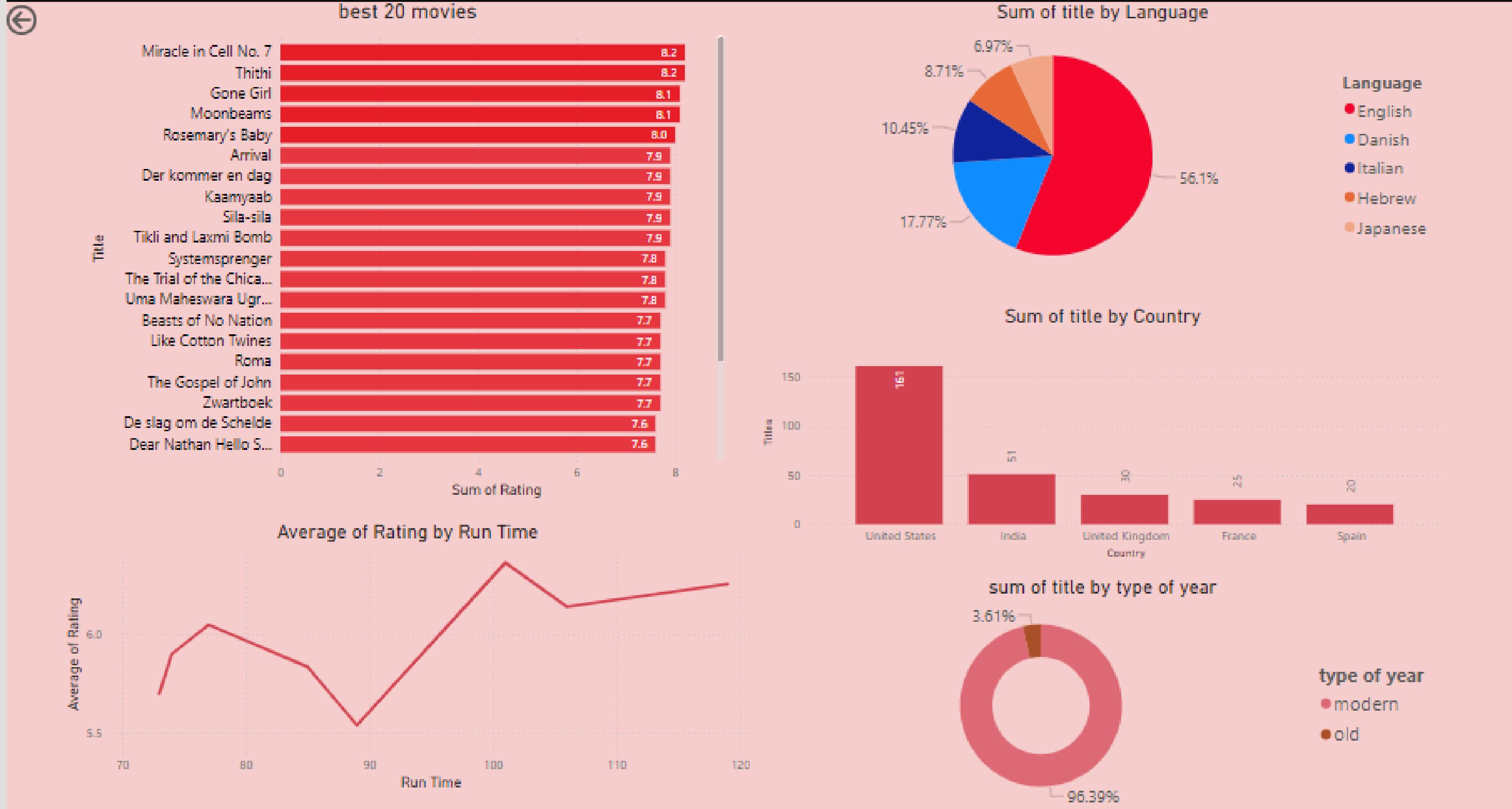
Start Year	Titles
2010	~300
2012	~400
2014	~500
2016	~600
2018	~700
2020	~1,000

Avg Rating and num of Titles by Genre

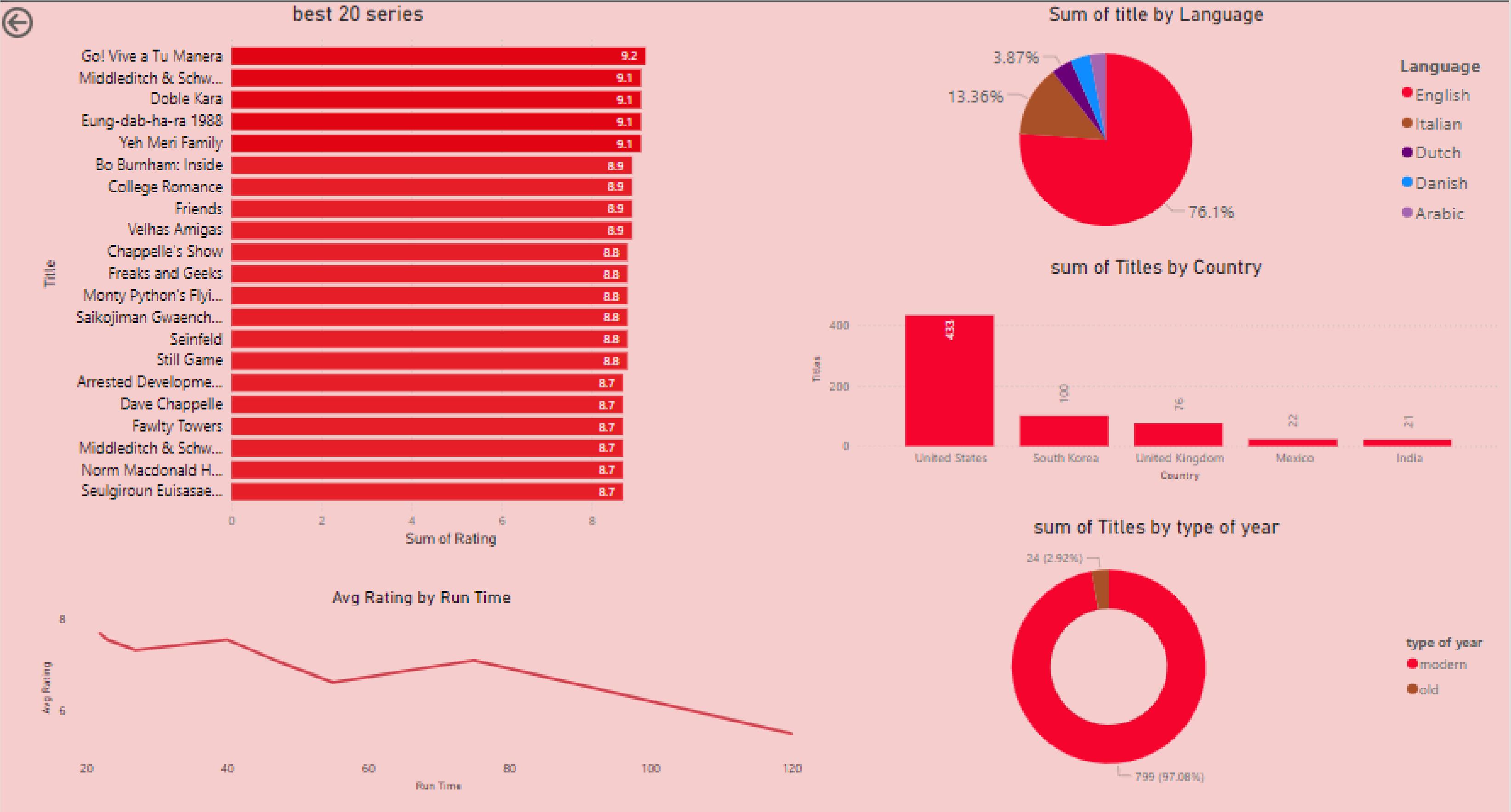
Genre	Avg Rating	Titles	Video Titles
War	8.2	1	1,384
Film-Noir	7.5	1	
Music	7.4	14	
Documentary	7.2	522	
Biography	7.1	215	
Crime	7.1	498	
Sport	7.1	4	
Animation	7.0	251	
Adventure	6.7	287	
Musical	6.7	2	
Drama	6.7	66	
Romance	6.7	31	
Action	6.6	66	
Comedy	6.6	65	
Western	6.6	65	
Family	6.5	39	
Talk-Show	6.5	10	
Game-Sho...	6.4	25	
Reality-TV	6.3	26	
Fantasy	6.2	1	
History	6.2	4	
News	6.2	20	
Mystery	5.8	41	
Thriller	5.3		

Movies & TV Shows				
All				
Title	Plot	Rating	Votes	
Super Songs	award-winning educational apps, "StoryBots Super Songs" centers on the StoryBots, the curious little creatures who live in the world beneath our screens and help answer humans' biggest questions.			
Tae-yang-eui hoo-ye	This drama tells of the love story that develops between a surgeon and a special forces officer.	8.3	10,544	
Scams	Kusano has a bright future with his fancy university degree and new job, yet, it is not to be. The blood sucking financial giant which he works at goes broke and leaves him on the pavement. When it rains it pours and his... Read all	6.3	265	
Sobat Ambyar	Coping with heartbreak, the shy owner of floundering cafe	6.4	144	

Ratings of Movies



Ratings of Tv-Shows



Rating Titles By Countries and Type





Insights

- **the country that always has high ratings is (China) with a rate of(9.3)**
- **the most popular types of shows is the (Action) shows**
- **we note that from 2008 to 2013, crime-type films were the most rated**
- **the most countries that received ratings is (New Zealand) with a percentage of ratings (8.049)**
- **most films always receive high ratings,especially films that exceed two hours in length**
- **films that get high ratings always get high votes**
- **the most commonly used language in films,which achieves a high rating when used,is(English)**



Recommendations



- we advise allocating marketing campaign in the most high rated and productive countries like United States 7.03

- United Kingdom 7.25
- Japan 8.00

- we recommend investing in movies with high ratings like action movies
- we advise allocating marketing campaigns and increase production in the most high rated countries like New Zealand
- we recommend increase the production of films that exceed two hours as they are highly rated
- we recommend focusing on the most highly rated languages as Malayaiam as it has low production
- we recommend increase the production of films that uses English language

Any Questions



THANK YOU