

NETFLIX

Movies and TV Shows Data Analysis using SQL

By
Jagan Saravana

Table of Content

- Introduction
- Dataset Description
- Tool Used
- Questions Analyzed
- SQL Queries
- Conclusion



Introduction



Overview of the project

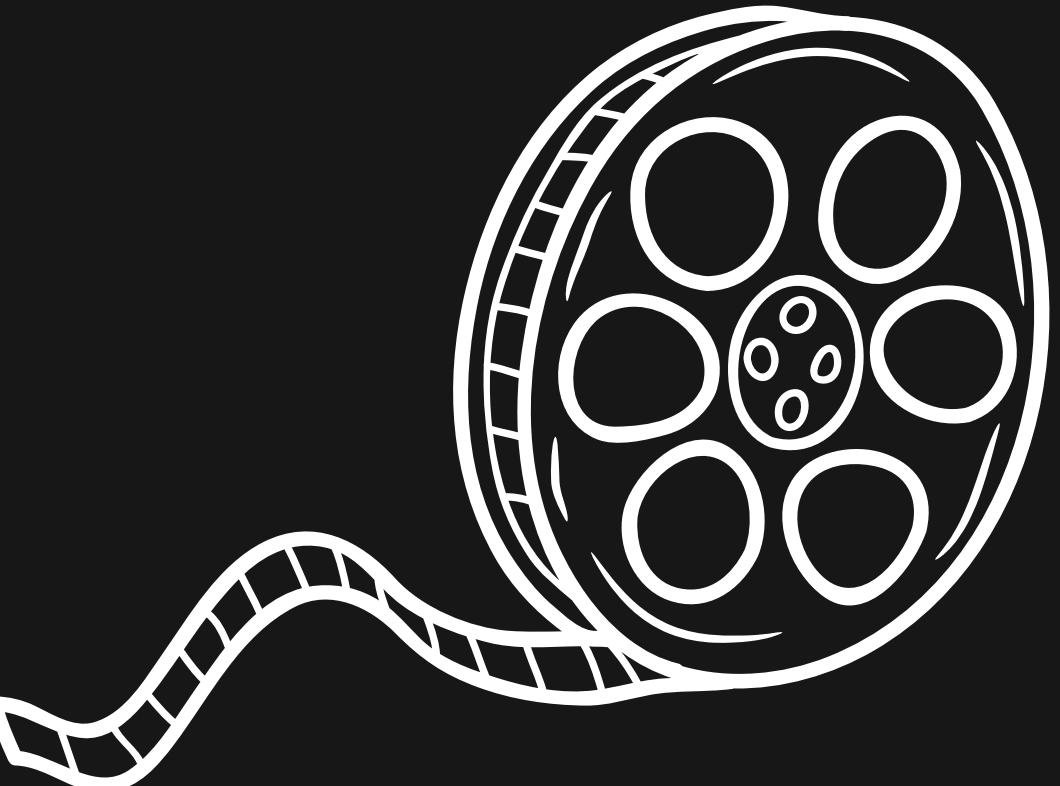
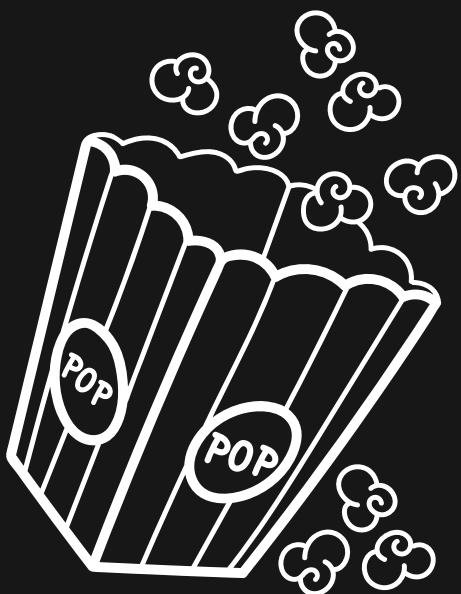
This project centers around analyzing Netflix's Movies and TV Shows dataset, leveraging the power of SQL for data exploration and insight generation. The dataset contains information about Netflix's extensive content library, including attributes such as titles, genres, release years, ratings, and more. By performing SQL queries, the project aims to uncover patterns and trends, offering a deeper understanding of Netflix's content distribution and evolution over time.

The primary focus of the analysis is to answer specific business and research-oriented questions. These include identifying the balance between movies and TV shows, the most common ratings, the longest movie on the platform, and trends in content additions over recent years. Such insights help shed light on content preferences, production trends, and market strategies.

Introduction

Objective

- To explore Netflix's content catalog and extract meaningful insights.
- To answer specific questions related to the content types, ratings, trends, and more using SQL queries.
- To demonstrate proficiency in SQL by formulating and executing queries to derive actionable information from the dataset.



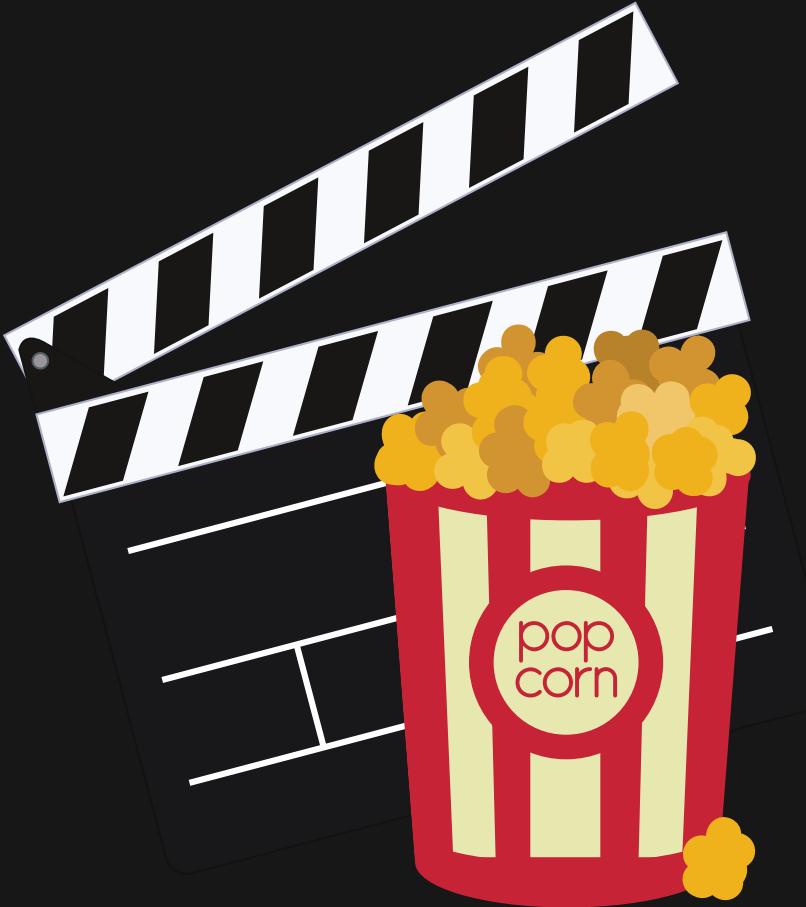
Dataset Description

Dataset Source and Format

- Source: The dataset used is a publicly available Netflix Movies and TV Shows dataset.
- Format: The data is in CSV format, containing structured information about titles, directors, genres, release years, countries, durations, and more.

Key Attributes in the Dataset

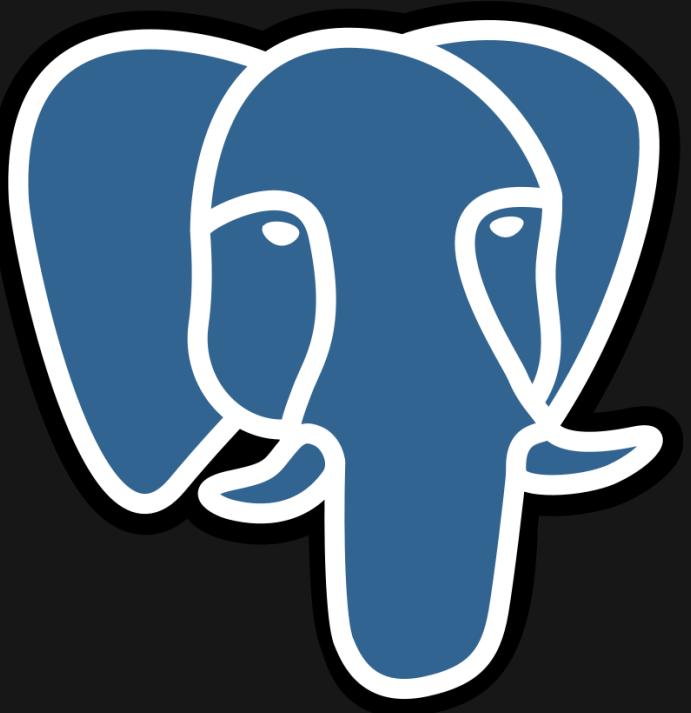
- Title: Name of the movie or TV show.
- Director: Name of the director.
- Genre: Type of content (e.g., Comedy, Drama).
- Country: Country where the content was produced.
- Date Added: Date when the content was added to Netflix.
- Release Year: Year of release.
- Rating: Content rating (e.g., PG-13, R).
- Duration: Length of the movie or number of seasons for TV shows.



Tool Used

PostgreSQL

PostgreSQL is a powerful open-source relational database management system (RDBMS). It supports advanced data types, complex queries, and data integrity, making it ideal for analyzing large datasets. Its compatibility with structured data and powerful querying capabilities ensured a seamless data analysis experience for this project.



Questions Analyzed

1. Count the Number of Movies vs TV Shows
2. Find the Most Common Rating for Movies and TV Shows
3. List All Movies Released in a Specific Year (e.g., 2020)
4. Find the Top 10 Countries with the Most Content on Netflix
5. Identify the Longest Movie
6. Find Content Added in the Last 5 Years
7. Find All Movies/TV Shows by Director 'K.S. Ravikumar'
8. List All TV Shows with More Than 8 Seasons
9. Count the Number of Content Items in Each Genre
10. Find each year and the average numbers of content release
in India on netflix



SQL Queries

1. Count the Number of Movies vs TV Shows

```
SELECT type, COUNT(*) total_content FROM netflix  
GROUP BY type;
```

Data Output Messages Notifications

	type character varying (10)	total_content bigint
1	Movie	6131
2	TV Show	2676

SQL Queries

2. Find the Most Common Rating for Movies and TV Shows

```
SELECT
    type,
    rating
FROM
(
    SELECT
        type,
        rating,
        COUNT(*),
        RANK() OVER(PARTITION BY type ORDER BY COUNT(*) DESC) as ranking
    FROM netflix
    GROUP BY 1,2
) as t1
WHERE
    ranking = 1
```

Data Output Messages Notifications

	type character varying (10)	rating character varying (10)
1	Movie	TV-MA
2	TV Show	TV-MA

SQL Queries

3. List All Movies Released in a Specific Year (e.g., 2020)

```
SELECT * FROM netflix  
WHERE type = 'Movie' AND release_year=2021;
```

	show_id character varying (6) 	type character varying (10) 	title character varying (150) 	director character varying (210)
1	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha
2	s10	Movie	The Starling	Theodore Melfi
3	s13	Movie	Je Suis Karl	Christian Schwochow
4	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti
5	s19	Movie	Intrusion	Adam Salky
6	s24	Movie	Go! Go! Cory Carson: Chrissy Takes the Wheel	Alex Woo, Stanley Moore
7	s31	Movie	Ankahi Kahaniya	Ashwiny Iyer Tiwari, Abhishek Chaubey, Saket Chaudhary
8	s36	Movie	The Father Who Moves Mountains	Daniel Sandu
9	s37	Movie	The Stronghold	Cédric Jimenez
10	s46	Movie	My Heroes Were Cowboys	Tyler Greco
11	s65	Movie	Nightbooks	David Yarovesky
12	s69	Movie	Schumacher	Hanns-Bruno Kammertöns, Vanessa Nöcker, Michael Wech

SQL Queries

4. Find the Top 10 Countries with the Most Content on Netflix

```
SELECT *
FROM
(
    SELECT
        UNNEST(STRING_TO_ARRAY(country, ',')) AS country,
        COUNT(*) AS total_content
    FROM netflix
    GROUP BY 1
) AS t1
WHERE country IS NOT NULL
ORDER BY total_content DESC
LIMIT 10;
```

	country	total_content
1	United States	3211
2	India	1008
3	United Kingd...	628
4	United States	479
5	Canada	271
6	Japan	259
7	France	212
8	South Korea	211
9	France	181
10	Spain	181

SQL Queries

5. Identify the Longest Movie

```
SELECT
    *
FROM netflix
WHERE type = 'Movie'
ORDER BY SPLIT_PART(duration, ' ', 1)::INT DESC;
```

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (210)
1	s5814	Movie	Louis C.K.: Live at the Comedy Store	Louis C.K.
2	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.
3	s5542	Movie	Louis C.K. 2017	Louis C.K.
4	s4254	Movie	Black Mirror: Bandersnatch	[null]
5	s718	Movie	Headspace: Unwind Your Mind	[null]
6	s2492	Movie	The School of Mischief	Houssam El-Din Mustafa
7	s2488	Movie	No Longer kids	Samir Al Asfory
8	s2485	Movie	Lock Your Girls In	Fouad El-Mohandes
9	s2489	Movie	Raya and Sakina	Hussein Kamal
10	s167	Movie	Once Upon a Time in America	Sergio Leone
11	s7933	Movie	Sangam	Raj Kapoor
12	s1020	Movie	Lagaan	Ashutosh Gowariker

SQL Queries

6. Find Content Added in the Last 5 Years

```
SELECT *
FROM netflix
WHERE TO_DATE(date_added, 'Month DD, YYYY') >= CURRENT_DATE - INTERVAL '5 years';
```

	show_id character varying (6) 	type character varying (10) 	title character varying (150) 	director character varying (210)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
2	s2	TV Show	Blood & Water	[null]
3	s3	TV Show	Ganglands	Julien Leclercq
4	s4	TV Show	Jailbirds New Orleans	[null]
5	s5	TV Show	Kota Factory	[null]
6	s6	TV Show	Midnight Mass	Mike Flanagan
7	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha
8	s8	Movie	Sankofa	Haile Gerima
9	s9	TV Show	The Great British Baking Show	Andy Devonshire
10	s10	Movie	The Starling	Theodore Melfi
11	s11	TV Show	Vendetta: Truth, Lies and The Mafia	[null]
12	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri

SQL Queries

7. Find All Movies/TV Shows by Director 'K.S. Ravikumar'

```
SELECT *
FROM (
  SELECT
    *,
    UNNEST(STRING_TO_ARRAY(director, ',')) AS director_name
  FROM netflix
) AS t
WHERE director_name = 'K.S. Ravikumar';
```

Data Output Messages Notifications

Showing rows: 1 to 1 | | Page No: 1 of 1 |

	show_id character varying (6) 	type character varying (10) 	title character varying (150) 	director character varying (210) 	casts character varying (1000) 
1	s23	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Rajgopal, Nassar, S.P. Balasubrahmany...

SQL Queries

8.List All TV Shows with More Than 8 Seasons

```
SELECT *
FROM netflix
WHERE type = 'TV Show'
    AND SPLIT_PART(duration, ' ', 1)::INT > 8;
```

SQL Queries

9. Count the Number of Content Items in Each Genre

```
SELECT
    UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre,
    COUNT(*) AS total_content
FROM netflix
GROUP BY 1;
```

	genre text	total_content bigint
1	International TV Shows	774
2	Stand-Up Comedy	9
3	Spanish-Language TV Shows	172
4	Romantic Movies	613
5	Anime Features	21
6	TV Horror	11
7	Stand-Up Comedy & Talk Shows	34
8	Kids' TV	63
9	Docuseries	221
10	Children & Family Movies	605
11	Movies	57
12	Horror Movies	82

Total rows: 73 | Query complete 00:00:00.170

SQL Queries

10. Find each year and the average numbers of content release in India on netflix

```
SELECT
    country,
    release_year,
    COUNT(show_id) AS total_release,
    ROUND(
        COUNT(show_id)::numeric /
        (SELECT COUNT(show_id) FROM netflix WHERE country = 'India')::numeric * 100, 2
    ) AS avg_release
FROM netflix
WHERE country = 'India'
GROUP BY country, release_year
ORDER BY avg_release DESC
LIMIT 5;
```

	country character varying (150) 	release_year integer 	total_release bigint 	avg_release numeric 
1	India	2017	101	10.39
2	India	2018	94	9.67
3	India	2019	87	8.95
4	India	2020	75	7.72
5	India	2016	73	7.51

Conclusion

This project successfully analyzed Netflix's Movies and TV Shows dataset using SQL, providing valuable insights into the platform's content library. By answering ten carefully crafted questions, we uncovered trends and patterns, such as the distribution of movies and TV shows, the most common content ratings, and the geographical diversity of content production. The analysis also highlighted specific details, like the longest movie, frequently featured genres, and content additions in recent years.

The use of PostgreSQL proved instrumental in managing and querying the dataset efficiently, enabling accurate and meaningful results. These insights can be applied to understand viewer preferences, guide content acquisition strategies, and track market trends. Overall, this project demonstrated the power of SQL in extracting actionable intelligence from large datasets, reinforcing its value in data analytics.

THANKS