Table of Contents:

- Problem Statement
- Tool Used
- Table / Dataset Description
- Project Objectives
- Recommendations

NETFLIX

Problem Statement:

▶ This project aims to conduct a comprehensive analysis of **Netflix Movie** using **SQL**. The database contains 5 tables:

1. **netflix_movies**

2. **netflix_movies_directors**

3. **netflix_movies_countries**

4. **netflix_movies_cast**

5. **netflix_movies_categories**

▶ The goal is to utilize **SQL** to create insightful analyses and reports that provide a deeper understanding of **Netflix Movie** performance, rating, duration, and movies categories. The analysis aims to uncover trends, preferences, and patterns in the data to aid movies uploader and stakeholders in optimizing their **Netflix Movie** .

NETFLIX

Tool Used

▶ PostgreSQL

PostgreSQL is an open-source relational database management system emphasizing extensibility and SQL compliance.

PostgreSQL features transactions with atomicity, consistency, isolation, durability (ACID) properties, automatically updatable views, materialized views, triggers, foreign keys, and stored procedures.

Table Dataset Description:

1. show_id: Unique identifier for each Netflix Movie.
2. duration_minutes: Duration of the Netflix Movie.
3. duration_seasons: Seansons Duration of the Netflix Movie.
4. type: Netflix Movie type (e.g: TV Show, Movie)
5. title: Title of the Netflix Movie.
6. date_added: Date when the Netflix Movie was published.
7. release_year: the year when the Movie was released.
8. rating: Netflix Movie type (e.g: PG, TV-Y, PG-13, G, TV-MA, e.t.c)
9. description: Description provided for the Netflix Movie.
10. director: the director of each Netflix Movie
11. country: the originating country of the Netflix movie
12. movie_cast: the Netflix movie cast
13. movie_categories: the category of each Netflix movie (e.g: Comedies, Thrillers, Horror, e.t.c)

Project Objectives:

- ▶ Data Cleaning and Preparation

- ▶ Exploratory Data Analysis (EDA)

- ▶ Content, Ratings , Categories Analysis and JOINS

- ▶ Temporal Trends with 'CASE'

- ▶ Movie Cast Engagement Insight
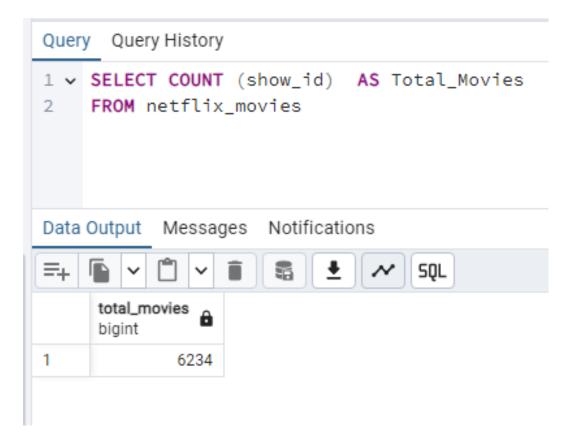
- ▶ Recommendation

NETFLIX

1. How many movies are in the datasets

According to the Netflix_movies table, analyses shows that there are 6,234 movies in the datasets

Query   Query History

```sql
1  SELECT COUNT (show_id)  AS Total_Movies
2  FROM netflix_movies
```

Data Output   Messages   Notifications

	total_movies
	bigint
1	6234

NETFLIX
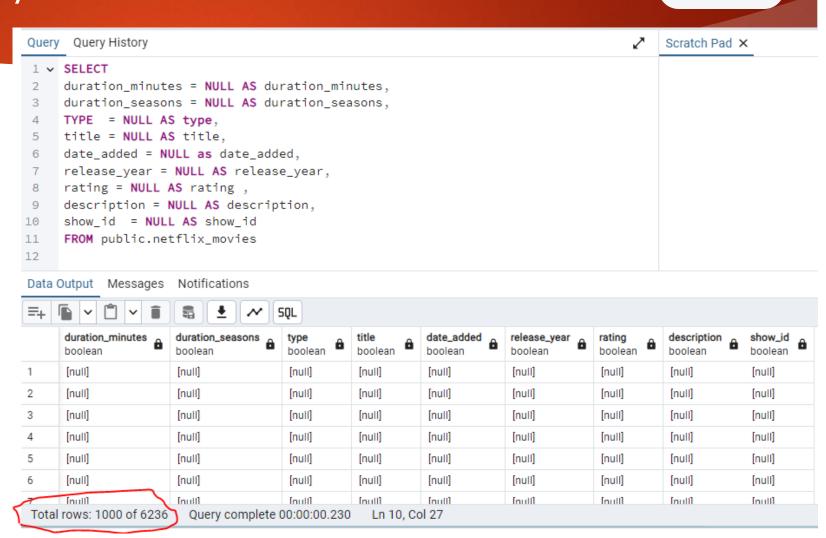
2. Write to check if there are(null values/figures) in the datasets)

By checking the for null values, result shows that there 1,000 missing information out 6,236 information in the database.

Which is crucial for maintaining data integrity and making informed decisions.

NETFLIX

Query   Query History

```
1 v  SELECT
2    duration_minutes = NULL AS duration_minutes,
3    duration_seasons = NULL AS duration_seasons,
4    TYPE   = NULL AS type,
5    title = NULL AS title,
6    date_added = NULL as date_added,
7    release_year = NULL AS release_year,
8    rating = NULL AS rating ,
9    description = NULL AS description,
10   show_id  = NULL AS show_id
11   FROM public.netflix_movies
12
```

Scratch Pad ✕

Data Output   Messages   Notifications

	duration_minutes boolean	duration_seasons boolean	type boolean	title boolean	date_added boolean	release_year boolean	rating boolean	description boolean	show_id boolean
1	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
2	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
3	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
4	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
5	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
6	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
7	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]

Total rows: 1000 of 6236   Query complete 00:00:00.230   Ln 10, Col 27

3. Count the total number of movies and TV shows in the dataset

Query result shows that there are

4,264 Movies, **1,969** TV Show and **2**

media that was not stated to be

Movie or TV Show .

Query   Query History

```
1 v  SELECT DISTINCT "type" , COUNT (*) AS Total_Movies
2    FROM netflix_movies
3    GROUP BY "type"
4    ORDER by Total_Movies DESC
```

Data Output   Messages   Notifications

	type character varying	total_movies bigint
1	Movie	4264
2	TV Show	1969
3	Not Provided	2

4. Find the total and average duration of movies and TV shows

▶ Analyses shows that **MOVIES** takes lead of long duration over TV show with the :

Average Mins : 99mins 10sec

Total Mins: 422,571 mins

Which is equivalent to **TOTAL HOURS** of

7,042hrs

NETFLIX

Query   Query History

```
1 ∨  SELECT "type",
2    ROUND ( AVG (duration_mins), 2) AS Average_mins,
3    ROUND (SUM (duration_mins), 2) AS Total_mins,
4    ROUND (SUM (duration_mins)/ 60, 2) AS Total_Hours
5    FROM public.netflix_movies
6    GROUP BY "type"
7    ORDER BY Total_mins DESC
8
```

Data Output   Messages   Notifications

	type character varying	average_mins numeric	total_mins numeric	total_hours numeric
1	Movie	99.10	422571.00	7042.00
2	Not Provided	0.00	0.00	0.00
3	TV Show	0.00	0.00	0.00
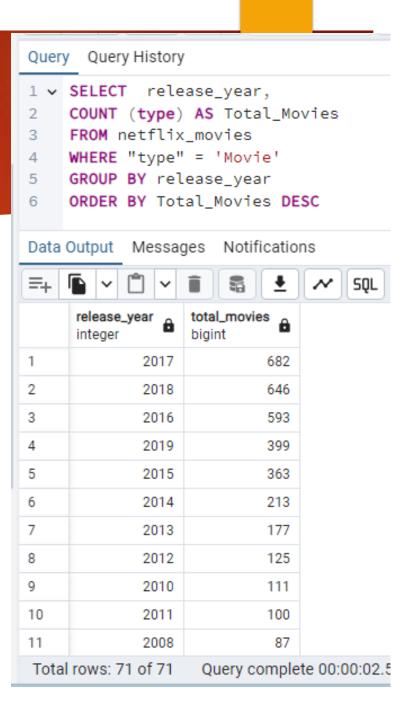
5. Find the number of movies released in each year.

```sql
SELECT  release_year,
COUNT (type) AS Total_Movies
FROM netflix_movies
WHERE "type" = 'Movie'
GROUP BY release_year
ORDER BY Total_Movies DESC
```

▶ By checking for the **number of movies released** each year.

Query result shows that year **2017** takes an **outstanding winning** with **682 movies** released while year **2018** and **2016** takes the 2nd and 3rd place with **646**, **593** respectively.

NETFLIX

Query | Query History

Data Output | Messages | Notifications

	release_year integer	total_movies bigint
1	2017	682
2	2018	646
3	2016	593
4	2019	399
5	2015	363
6	2014	213
7	2013	177
8	2012	125
9	2010	111
10	2011	100
11	2008	87

Total rows: 71 of 71 | Query complete 00:00:02.5

6. List the top 10 longest movies based on duration

Query result shows the **TOP 10 longest** movies based on their respectful duration.

However, movie titled = **Black Mirror: Bandersnatch** is the longest movie with **312 mins equivalent to 5 hours** and the movie was **released** in the year **2018**.
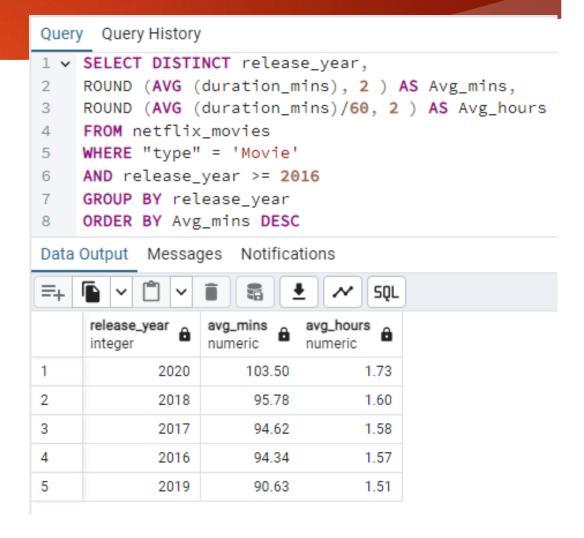
Query   Query History

```
1 ∨  SELECT "title", "duration_mins",
2    ROUND (SUM (duration_mins)/60, 2) AS Total_hours,
3    release_year
4    FROM netflix_movies
5    WHERE "type" = 'Movie'
6    GROUP BY title, duration_mins,release_year
7    ORDER by duration_mins DESC
8    LIMIT 10
```

Data Output   Messages   Notifications

	title character varying	duration_mins integer	total_hours numeric	release_year integer
1	Black Mirror: Bandersnatch	312	5.00	2018
2	Sangam	228	3.00	1964
3	Lagaan	224	3.00	2001
4	Jodhaa Akbar	214	3.00	2008
5	The Irishman	209	3.00	2019
6	The Gospel of Luke	205	3.00	2015
7	What's Your Raashee?	203	3.00	2009
8	The Lord of the Rings: The Return of the King	201	3.00	2003
9	Doctor Zhivago	200	3.00	1965
10	Elephants Dream 4 Hour	196	3.00	2006

7. Find the average duration of movies released in the last 5 years

▶ Analyses shows that movie released in the last **5 years** (assuming the present year is **2020**) has the maximum of average duration of **1 hour 73mins** and also minimum of average duration is **1 hour 51mins** of movies

```sql
1  ∨  SELECT DISTINCT release_year,
2      ROUND (AVG (duration_mins), 2 ) AS Avg_mins,
3      ROUND (AVG (duration_mins)/60, 2 ) AS Avg_hours
4      FROM netflix_movies
5      WHERE "type" = 'Movie'
6      AND release_year >= 2016
7      GROUP BY release_year
8      ORDER BY Avg_mins DESC
```

Data Output   Messages   Notifications

release_year integer	avg_mins numeric	avg_hours numeric
2020	103.50	1.73
2018	95.78	1.60
2017	94.62	1.58
2016	94.34	1.57
2019	90.63	1.51

NETFLIX

8. Determine the most common rating for TV shows

```
Query   Query History

1  ∨  SELECT DISTINCT rating,
2      COUNT (*) AS Total_rating
3      FROM netflix_movies
4      WHERE "type" = 'TV Show'
5      GROUP BY "rating"
6      ORDER BY Total rating DESC
```

Data Output   Messages   Notifications

	rating character varying	total_rating bigint
1	TV-MA	679
2	TV-14	660
3	TV-PG	269
4	TV-Y	102
5	TV-Y7	100
6	TV-G	69
7	TV-Y7-FV	68
8	NR	16
9	Not Provided	2
10	R	2
11	G	1
12	PG	1

Total rows: 12 of 12   Query complete 00:0

▶ From the query analyses result, there are **12 different type** of **ratings** for TV shows.

    However, result shows that **TV-MA rating** is the top and most common rating for TV Shows with **total** of **679** all time counts of rating while **TV-14 and TV-PG** takes the **2nd and 3rd** position with **660 , 269** respectively.

9. List the titles of movies with a duration greater than 180 minutes and a rating of 'R'

► Query result shows that there are only 3

movie with a duration greater than 180

minutes and a rating of 'R'

Query | Query History

```sql
1  SELECT title, duration_mins, rating
2  from netflix_movies
3  WHERE "type" = 'Movie'
4  AND duration_mins >= 180
5  AND rating ='R'
```

Data Output | Messages | Notifications

	title character varying	duration_mins integer	rating character varying
1	The Irishman	209	R
2	Magnolia	189	R
3	Schindler's List	195	R
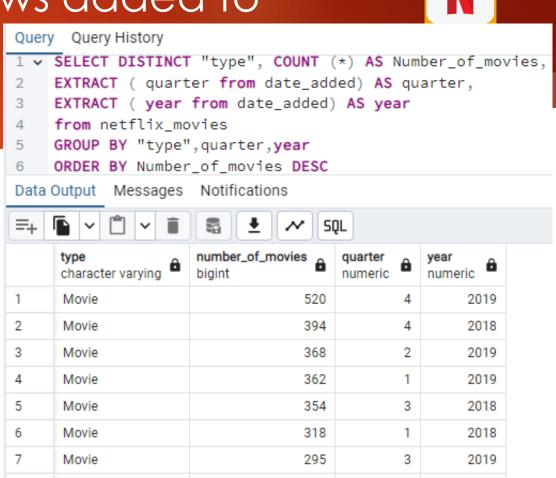
10. Find the number of TV shows added to Netflix in each quarter

```sql
1  SELECT DISTINCT "type", COUNT (*) AS Number_of_movies,
2  EXTRACT ( quarter from date_added) AS quarter,
3  EXTRACT ( year from date_added) AS year
4  from netflix_movies
5  GROUP BY "type",quarter,year
6  ORDER BY Number_of_movies DESC
```

Query   Query History

Data Output   Messages   Notifications

	type character varying	number_of_movies bigint	quarter numeric	year numeric
1	Movie	520	4	2019
2	Movie	394	4	2018
3	Movie	368	2	2019
4	Movie	362	1	2019
5	Movie	354	3	2018
6	Movie	318	1	2018
7	Movie	295	3	2019
8	Movie	245	4	2017
9	Movie	233	1	2017
10	Movie	224	2	2018
11	TV Show	224	4	2019
12	Movie	222	2	2017

▶ Analyses shows certain number of movies published on Netflix in each quarter.

   Further queries result shows that year 2019 Q4 has the highest number movies published which is 520 movies follow by year 2018 Q4 and 2019 Q2 with total of 394, 368 movies respectively.

NETFLIX

11. Calculate the average rating for movies released before 2000.
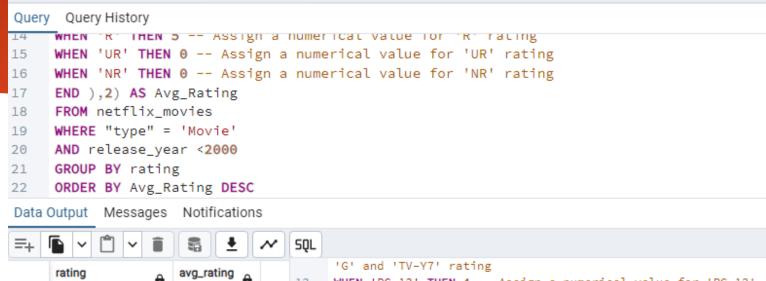
- Query result display all types of *Rating* for **Movies** released before *2000*.

  However, '*R'* ratings takes the lead with the **highest average rating** before year 2000 with **5.00** while '*TV-G'* ratings has the *least average rating* of *1.00*.

  Moreso, '*UR and NR*' ratings has none with *0.00* average rating before *2000*.

```
14    WHEN 'R' THEN 5 -- Assign a numerical value for 'R' rating
15    WHEN 'UR' THEN 0 -- Assign a numerical value for 'UR' rating
16    WHEN 'NR' THEN 0 -- Assign a numerical value for 'NR' rating
17    END ),2) AS Avg_Rating
18    FROM netflix_movies
19    WHERE "type" = 'Movie'
20    AND release_year <2000
21    GROUP BY rating
22    ORDER BY Avg_Rating DESC
```

Data Output   Messages   Notifications

	rating character varying	avg_rating numeric
1	R	5.00
2	TV-MA	4.00
3	PG-13	4.00
4	PG	3.00
5	TV-14	3.00
6	TV-PG	2.50
7	G	2.00
8	TV-G	1.00
9	UR	0.00
10	NR	0.00

```
      'G' and 'TV-Y7' rating
13    WHEN 'PG-13' THEN 4 -- Assign a numerical value for 'PG-13'
      rating
14    WHEN 'R' THEN 5 -- Assign a numerical value for 'R' rating
15    WHEN 'UR' THEN 0 -- Assign a numerical value for 'UR'
      rating
16    WHEN 'NR' THEN 0 -- Assign a numerical value for 'NR' rating
17    END ),2) AS Avg_Rating
18    FROM netflix_movies
19    WHERE "type" = 'Movie'
20    AND release_year <2000
21    GROUP BY rating
22    ORDER BY Avg_Rating DESC
```
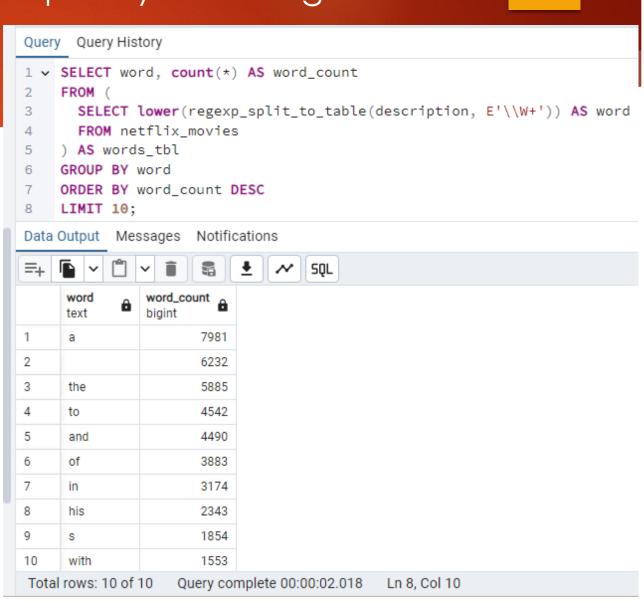
12. Determine the top 5 most frequently occurring words in the description column.

▶ **SQL data mining techniques** was used to ascertain the most common word used in the description column.
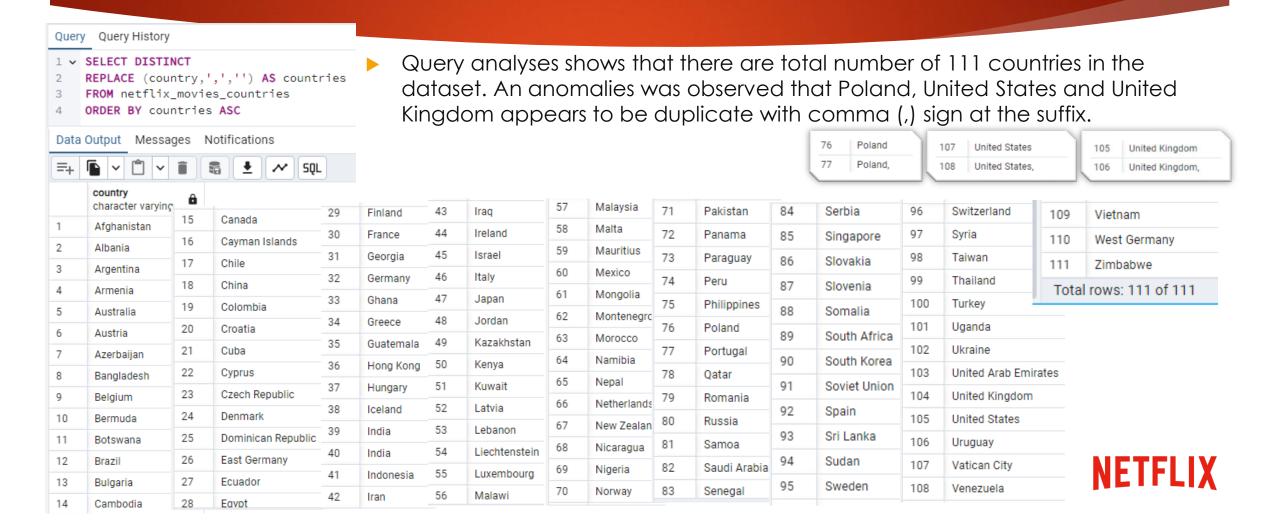
Analyses show that **'a'** is the most common character word used with **total occurrence of 7,981** followed by **space ' '** character and **'the'** with **total occurrence of 6,232 , 5885 respectively**.

NETFLIX

```
Query   Query History

1 ∨  SELECT word, count(*) AS word_count
2    FROM (
3      SELECT lower(regexp_split_to_table(description, E'\\W+')) AS word
4      FROM netflix_movies
5    ) AS words_tbl
6    GROUP BY word
7    ORDER BY word_count DESC
8    LIMIT 10;
```

Data Output   Messages   Notifications

	word text	word_count bigint
1	a	7981
2		6232
3	the	5885
4	to	4542
5	and	4490
6	of	3883
7	in	3174
8	his	2343
9	s	1854
10	with	1553

Total rows: 10 of 10   Query complete 00:00:02.018   Ln 8, Col 10

13. Find and display all Countries in the dataset

```
Query   Query History

1  v  SELECT DISTINCT
2     REPLACE (country,',','') AS countries
3     FROM netflix_movies_countries
4     ORDER BY countries ASC
```

Data Output   Messages   Notifications

▶ Query analyses shows that there are total number of 111 countries in the dataset. An anomalies was observed that Poland, United States and United Kingdom appears to be duplicate with comma (,) sign at the suffix.

76	Poland		107	United States		105	United Kingdom
77	Poland,		108	United States,		106	United Kingdom,

#	country character varying		#			#			#			#			#			#			#			#	
1	Afghanistan	15	Canada	29	Finland	43	Iraq	57	Malaysia	71	Pakistan	84	Serbia	96	Switzerland	109	Vietnam
2	Albania	16	Cayman Islands	30	France	44	Ireland	58	Malta	72	Panama	85	Singapore	97	Syria	110	West Germany
3	Argentina	17	Chile	31	Georgia	45	Israel	59	Mauritius	73	Paraguay	86	Slovakia	98	Taiwan	111	Zimbabwe
4	Armenia	18	China	32	Germany	46	Italy	60	Mexico	74	Peru	87	Slovenia	99	Thailand		
5	Australia	19	Colombia	33	Ghana	47	Japan	61	Mongolia	75	Philippines	88	Somalia	100	Turkey		Total rows: 111 of 111
6	Austria	20	Croatia	34	Greece	48	Jordan	62	Montenegro	76	Poland	89	South Africa	101	Uganda		
7	Azerbaijan	21	Cuba	35	Guatemala	49	Kazakhstan	63	Morocco	77	Portugal	90	South Korea	102	Ukraine		
8	Bangladesh	22	Cyprus	36	Hong Kong	50	Kenya	64	Namibia	78	Qatar	91	Soviet Union	103	United Arab Emirates		
9	Belgium	23	Czech Republic	37	Hungary	51	Kuwait	65	Nepal	79	Romania	92	Spain	104	United Kingdom		
10	Bermuda	24	Denmark	38	Iceland	52	Latvia	66	Netherlands	80	Russia	93	Sri Lanka	105	United States		
11	Botswana	25	Dominican Republic	39	India	53	Lebanon	67	New Zealand	81	Samoa	94	Sudan	106	Uruguay		
12	Brazil	26	East Germany	40	India	54	Liechtenstein	68	Nicaragua	82	Saudi Arabia	95	Sweden	107	Vatican City		
13	Bulgaria	27	Ecuador	41	Indonesia	55	Luxembourg	69	Nigeria	83	Senegal					108	Venezuela
14	Cambodia	28	Egypt	42	Iran	56	Malawi	70	Norway								

NETFLIX

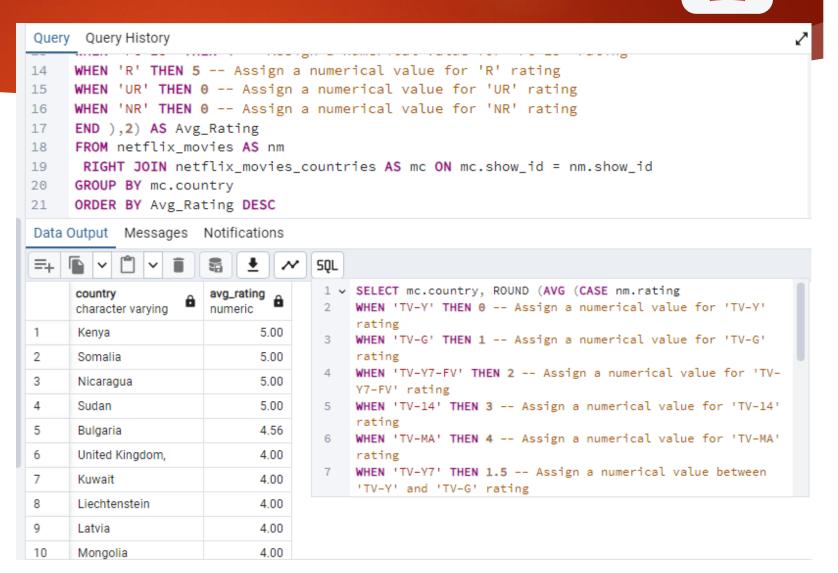# 14. Find the Country with the highest Netflix Movies

Analyses shows that **United States** has the highest

number of movie with **2,609 movies** while **India** and

United Kingdom has **837, 601** movies respectively.

NETFLIX

Query   Query History

```
1 ∨  SELECT
2        DISTINCT (country),
3            COUNT (show_id)   AS Total_movies
4    FROM public.netflix_movies_countries
5        GROUP BY country
6            ORDER BY Total_movies DESC
```

Data Output   Messages   Notifications

	country character varying	total_movies bigint
1	United States	2609
2	India	837
3	United Kingdom	601
4	Canada	318
5	France	271
6	Japan	231
7	Spain	178
8	South Korea	162
9	Germany	151
10	Mexico	129
11	Australia	126

15. Country with the highest Average Rating

▶ Sequel to the query result which shows that *Kenya, Somalia, Nicaragua, Sudan* takes top place with 5.00 avg. rating


```
14   WHEN 'R' THEN 5 -- Assign a numerical value for 'R' rating
15   WHEN 'UR' THEN 0 -- Assign a numerical value for 'UR' rating
16   WHEN 'NR' THEN 0 -- Assign a numerical value for 'NR' rating
17   END ),2) AS Avg_Rating
18   FROM netflix_movies AS nm
19    RIGHT JOIN netflix_movies_countries AS mc ON mc.show_id = nm.show_id
20   GROUP BY mc.country
21   ORDER BY Avg_Rating DESC
```

Data Output   Messages   Notifications

	country character varying	avg_rating numeric
1	Kenya	5.00
2	Somalia	5.00
3	Nicaragua	5.00
4	Sudan	5.00
5	Bulgaria	4.56
6	United Kingdom,	4.00
7	Kuwait	4.00
8	Liechtenstein	4.00
9	Latvia	4.00
10	Mongolia	4.00

```
1 ∨  SELECT mc.country, ROUND (AVG (CASE nm.rating
2    WHEN 'TV-Y' THEN 0 -- Assign a numerical value for 'TV-Y'
     rating
3    WHEN 'TV-G' THEN 1 -- Assign a numerical value for 'TV-G'
     rating
4    WHEN 'TV-Y7-FV' THEN 2 -- Assign a numerical value for 'TV-
     Y7-FV' rating
5    WHEN 'TV-14' THEN 3 -- Assign a numerical value for 'TV-14'
     rating
6    WHEN 'TV-MA' THEN 4 -- Assign a numerical value for 'TV-MA'
     rating
7    WHEN 'TV-Y7' THEN 1.5 -- Assign a numerical value between
     'TV-Y' and 'TV-G' rating
```
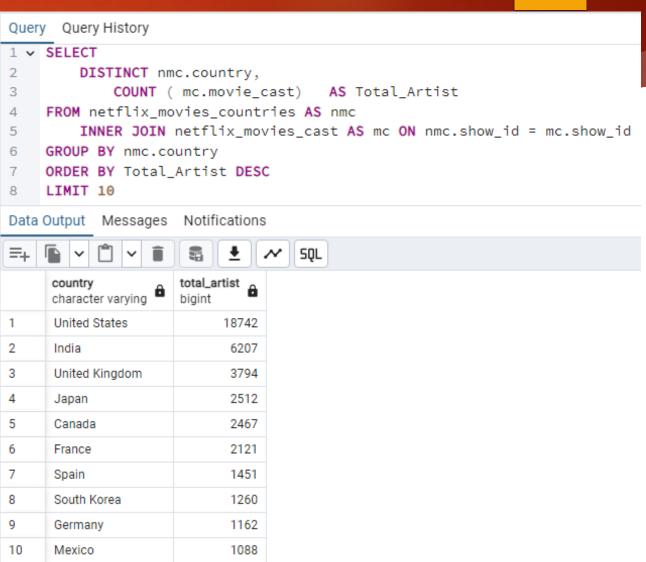
NETFLIX

16. Find the Top 10 Countries with the highest number of artist

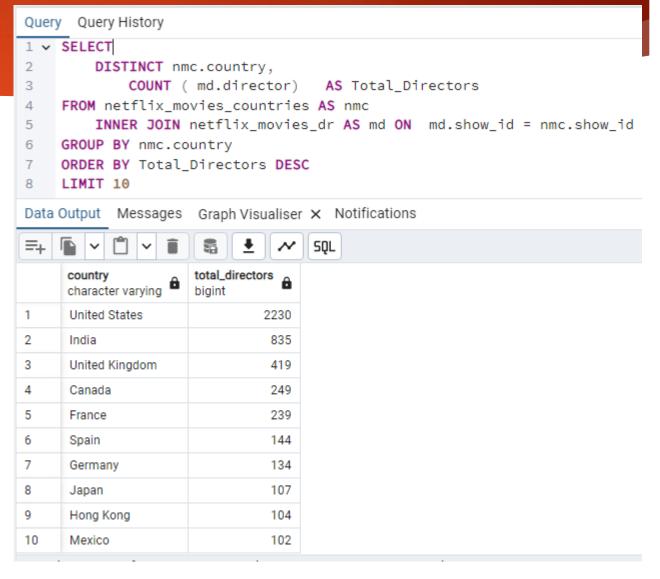▶ Query result shows the TOP 10 Countries with the highest number of artist in the dataset.

However, as seen that **United States** takes the lead which have the highest number of artist with **18,742** artist in the dataset.

N NETFLIX

Query   Query History

```sql
1 ∨  SELECT
2        DISTINCT nmc.country,
3            COUNT ( mc.movie_cast)    AS Total_Artist
4    FROM netflix_movies_countries AS nmc
5        INNER JOIN netflix_movies_cast AS mc ON nmc.show_id = mc.show_id
6    GROUP BY nmc.country
7    ORDER BY Total_Artist DESC
8    LIMIT 10
```

Data Output   Messages   Notifications

	country character varying	total_artist bigint
1	United States	18742
2	India	6207
3	United Kingdom	3794
4	Japan	2512
5	Canada	2467
6	France	2121
7	Spain	1451
8	South Korea	1260
9	Germany	1162
10	Mexico	1088

17. Find the Top 10 Countries with the highest number of directors

▶ Sequel to the query result showing the Top 10 countries with highest number of directors, it shows that United State produced the highest number of director of movies on Netflix which takes the top position with 2,230 total number of directors.

NETFLIX

```
Query   Query History

1 ∨   SELECT
2          DISTINCT nmc.country,
3              COUNT ( md.director)    AS Total_Directors
4      FROM netflix_movies_countries AS nmc
5          INNER JOIN netflix_movies_dr AS md ON  md.show_id = nmc.show_id
6      GROUP BY nmc.country
7      ORDER BY Total_Directors DESC
8      LIMIT 10
```

Data Output   Messages   Graph Visualiser ✕   Notifications

	country character varying 🔒	total_directors bigint 🔒
1	United States	2230
2	India	835
3	United Kingdom	419
4	Canada	249
5	France	239
6	Spain	144
7	Germany	134
8	Japan	107
9	Hong Kong	104
10	Mexico	102

Recommendations

❖ Movies publisher are advice to use popular tags, attractive titles, clear descriptions, sharp graphics cover design and high quality thumbnails to improve discoverability and attract viewers.

❖ Producers are encourage to produce and publish **movie** within the **Average Mins : 99mins 10sec** to attracting more viewers and build a strong community.

❖ Movies publisher should published Movie with the **TV-MA rating** to maximize views and engagement.

❖ Producers are encourage to produce **movies & TV shows** mostly at the last quarter of the year, since insight shows that users engagement are high during the last quarter **(oct,nov,dec).**

❖ Movie crew should ensure good audio and Movie quality, create valuable contents, knowledgeable contents engaging the audience (users) .

❖ Respectively, countries should invest more in the movie industry as insight shows that **US-United States** has the highest number of artists and directors of movies on **NETFLIX**. Thus, other countries should invest and develop their movie industry.

❖ Movies publisher should ensure data accuracy that is data inputted are free from error at the time of uploading. For instance, that Poland, United States and United Kingdom appears to be duplicate with comma (,) sign at the suffix. *(see page 19)*

❖ Regularly review Netflix Analytics to understand what works and refine your content strategy accordingly.

NETFLIX

THANK YOU

NETFLIX MOVIES ANALYSES

WITH

SQL

By

ADEDOKUN ABDULMALIK ADEYEMI