Netflix Data Analysis Using SQL

By Arjun Menon

1. Write a query to list all titles with their show_id, title, and type.

Query

```
-- Arjun= Write a query to list all titles with their show_id, title, and type.


select show_id,title,type from netflix;
```

Result

show_id	title	type
s1	Dick Johnson Is Dead	Movie
s2	Blood & Water	TV Show
s3	Ganglands	TV Show
s4	Jailbirds New Orleans	TV Show
s5	Kota Factory	TV Show
s6	Midnight Mass	TV Show
s7	My Little Pony: A New Generation	Movie
s8	Sankofa	Movie

2. Write a query to display all columns for titles that are Movies.

Query

```
1    -- Arjun=Write a query to display all columns for titles that are Movies.
2 •  select * from netflix where type="Movie";
```

Result

show_id	type	title	director	cast	country	date_added
s1352	Movie	Our Lady of San Juan, Four Centuries of Miracles	Noé González	Alejandro Peña Arenzana, Alejandra Yañez Rey...	Mexico	February 2, 2021
s1354	Movie	Beverly Hills Ninja	Dennis Dugan	Chris Farley, Nicollette Sheridan, Robin Shou, N...	United States	February 1, 2021
s1358	Movie	Rocks	Sarah Gavron	Bukky Bakray, Kosar Ali, D'angelou Osei Kissied...	United Kingdom	February 1, 2021
s1359	Movie	Shutter Island	Martin Scorsese	Leonardo DiCaprio, Mark Ruffalo, Ben Kingsley, ...	United States	February 1, 2021
s1360	Movie	The Bank Job	Roger Donaldson	Jason Statham, Saffron Burrows, Richard Linter...	United Kingdom, United States, Australia	February 1, 2021
s1365	Movie	Zathura	Jon Favreau	Jonah Bobo, Josh Hutcherson, Dax Shepard, Kr...	United States	February 1, 2021
s1367	Movie	Fatima	Marco Pontecorvo	Joaquim de Almeida, Goran Visnjic, Stephanie Gi...	Portugal, United States	January 31, 2021

3. Write a query to list TV shows that were released in the year 2021.

Query

```
1     -- Arjun=Write a query to list TV shows that were released in the year 2021.
2
3 •   select * from netflix where type="TV Show" and release_year=2021
```

Result

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
s2	TV Show	Blood & Water		Ama Qamata, Khosi Ngema, Gail Mabalane, Tha...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV
s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...		September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows
s4	TV Show	Jailbirds New Orleans				September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
s5	TV Show	Kota Factory		Mayur More, Jitendra Kumar, Ranjan Raj, Alam ...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Sh
s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, Hen...		September 24, 2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Holl...	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV
s11	TV Show	Vendetta: Truth, Lies and The Mafia				September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuseries, Internatio

4. Write a query to find all titles where the description contains the word family.
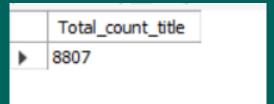
Query

```
1      -- Arjun= Write a query to find all titles where the description contains the word family.
2
3 •    select title from netflix where description like '%family%';
```

Result

title
▶ Ganglands
Je Suis Karl
Dark Skies
Omo Ghetto: the Saga
Shadow Parties
Gurgaon
Agatha Christie's Crooked House
Crocodile Dundee in Los Angeles

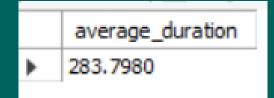# 5. Write a query to count the total number of titles in the dataset.

Query

```
1      -- Arjun=Write a query to count the total number of titles in the dataset.
2
3 ●    select count(title)as Total_count_title from netflix
```

Result

Total_count_title
8807

6. Write a query to find the average duration of all movies (in minutes, wherever the season is mentioned, consider 400 minutes per season).

Query

```
1      -- Arjun=Write a query to find the average duration of all movies (in minutes, wherever the season is mentioned, consider 400 minutes per season).
2
3  •   SELECT
4          AVG(duration_in_minutes) AS average_duration
5      FROM (
6          SELECT *,
7              CASE
8                  WHEN duration LIKE '%seasons%' OR duration LIKE '%season%'
9                      THEN CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED) * 400
10                 WHEN duration LIKE '%min%'
11                     THEN CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED)
12                 ELSE 0
13             END AS duration_in_minutes
14         FROM netflix
15     ) AS durations;
16
```

Result

average_duration
283.7980

7. Write a query to list the top 5 latest titles based on the date_added, sorted in descending order.

Query

```
1       -- Arjun=Write a query to list the top 5 latest titles based on the date_added, sorted in descending order.
2
3   •   SELECT title,DATE_FORMAT(STR_TO_DATE(date_added, '%M %d, %Y'), '%Y-%m-%d') AS formatted_date from netflix order by formatted_date desc limit 5
4
```

Result

title	formatted_date
► Dick Johnson Is Dead	2021-09-25
Blood & Water	2021-09-24
Ganglands	2021-09-24
Jailbirds New Orleans	2021-09-24
Kota Factory	2021-09-24

8. Write a query to list all titles along with the number of other titles by the same director. Include columns for show_id, title, director, and number_of_titles_by_director.

Query

```
1      -- Arjun=8. Write a query to list all titles along with the number of other titles by the sa
2
3 •    SELECT
4          show_id,
5          title,
6          director,
7          (SELECT COUNT(*)
8           FROM netflix n2
9           WHERE n1.director = n2.director
10             AND n1.show_id != n2.show_id) AS number_of_titles_by_director
11     FROM
12         netflix n1;
13
```

Result

show_id	title	director	number_of_titles_by_director
s1	Dick Johnson Is Dead	Kirsten Johnson	0
s2	Blood & Water		2633
s3	Ganglands	Julien Leclercq	2
s4	Jailbirds New Orleans		2633
s5	Kota Factory		2633
s6	Midnight Mass	Mike Flanagan	3
s7	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	0
s8	Sankofa	Haile Gerima	0

9. Write a query to find the total number of titles for each country. Display country and the count of titles.

Query

```
1    -- Arjun=Write a query to find the total number of titles for each country. Display country and the count of titles.
2
3 •  select country,count(title) as count_number from netflix group by country order by count_number desc;
```

Result

country	count_number
United States	2818
India	972
	831
United Kingdom	419
Japan	245
South Korea	199
Canada	181
Spain	145

10. Write a query using a CASE statement to categorize titles into three categories based on their rating: Family for ratings G, PG, PG-13, Kids for TV-Y, TV-Y7, TV-G, and Adult for all other ratings.

Query

```
1   -- Arjun= Write a query using a CASE statement to categorize titles into three categories based on their rating: F
2
3
4   select title,rating,
5   case when rating in('G', 'PG', 'PG-13') then 'family'
6   when rating in('TV-Y', 'TV-Y7', 'TV-G') then 'kids'
7   else 'Adult' end rating_category
8
9
10  from netflix
11  |
12
```

Result

title	rating	rating_categor
Dive Club	TV-G	kids
La casa de papel	TV-MA	Adult
Money Heist: From Tokyo to Berlin	TV-MA	Adult
Sharkdog	TV-Y	kids
Worth	PG-13	family
Afterlife of the Party	TV-PG	Adult
Anjaam	TV-14	Adult
Bright Star	PG	family

11. Write a query to add a new column title_length to the titles table that calculates the length of each title.

Query

```
1    -- Arjun=Write a query to add a new column title_length to the titles table that calculates the length of each title.
2
3  ● alter table netflix add column title_length int;
4
5
6  ● update netflix
7
8    set title_length=length(title);
9
10 ● select title,title_length from netflix;
```
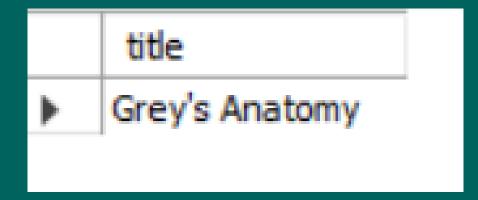
Result

title	title_length
Dick Johnson Is Dead	20
Blood & Water	13
Ganglands	9
Jailbirds New Orleans	21
Kota Factory	12
Midnight Mass	13
My Little Pony: A New Generation	32
Sankofa	7

12. Write a query using an advanced function to find the title with the longest duration in minutes.

Query

Result

```
1    -- Arjun= Write a query using an advanced function to find the title with the longest duration in minutes.
2  • ⊖ with cte as(
3      SELECT *,
4    ⊖    CASE
5             WHEN duration LIKE '%seasons%' OR duration LIKE '%season%'
6                 THEN CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED) * 400
7             WHEN duration LIKE '%min%'
8                 THEN CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED)
9             ELSE 0
10         END AS duration_in_minutes
11     FROM netflix),
12   ⊖ cte2 as(
13
14     select title,duration_in_minutes,rank() over (order by duration_in_minutes desc) as rnk from cte)
15
16     select title from cte2 where rnk=1
17
```

title
► Grey's Anatomy

13. Create a view named RecentTitles that includes titles added in the last 30 days.

Query

```
1    -- Arjun=Create a view named RecentTitles that includes titles added in the last 30 days.
2
3
4  •  CREATE VIEW RecentTitles AS
5    select title,formatted_date from(SELECT title,DATE_FORMAT(STR_TO_DATE(date_added, '%M %d, %Y'), '%Y-%m-%d') AS formatted_date from netflix )x
6    WHERE
7        formatted_date >= CURDATE() - INTERVAL 30 DAY;
8
9  •    select * from recenttitles
10
```
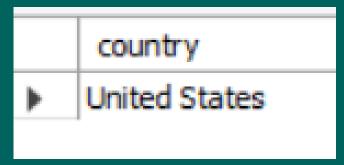
Result

	title	formatted_date

14. Write a query to find the country with the highest average rating for titles. Use subqueries and aggregate functions to achieve this.

Query

```sql
-- Arjun=Write a query to find the country with the highest average rating for titles. Use subqueries and aggregate functions to achieve this.
SELECT
    country
FROM
    netflix
WHERE title IN (SELECT title FROM
        (SELECT
            title, AVG(rating) AS avg_rating
        FROM
            netflix
        GROUP BY title) AS title_avg_ratings
    WHERE
        avg_rating = (SELECT
                MAX(avg_rating)
            FROM
                (SELECT
                    title, AVG(rating) AS avg_rating
                FROM
                    netflix
                GROUP BY title) AS highest_title_avg)) GROUP BY country;
```

Result

country
▶ United States

15. Write a query to find pairs of titles from the same country where one title has a higher rating than the other. Display columns for show_id_1, title_1, rating_1, show_id_2, title_2, and rating_2.

Query

```
1    -- Arjun=Write a query to find pairs of titles from the same country where one title has a higher rating than the other. Display columns for show_id_1, title_1, r
2
3
4 •  select n1.show_id show_id_1,n1.title title_1,n1.rating rating_1,n2.show_id show_id_2,n2.title title_2,n2.rating rating_2 from netflix n1 join netflix n2
5    on n1.country=n2.country and
6    n1.rating>n2.rating and
7    n1.show_id <> n2.show_id
8
9    |
```

Result

show_id_1	title_1		rating_1	show_id_2	title_2	rating_2
s4961	Happy Anniversary		TV-MA	s1	Dick Johnson Is Dead	PG-13
s4960	First Match		TV-MA	s1	Dick Johnson Is Dead	PG-13
s4951	John Mulaney: New in Town		TV-MA	s1	Dick Johnson Is Dead	PG-13
s4947	Star Trek: The Next Generation		TV-PG	s1	Dick Johnson Is Dead	PG-13
s4939	Seth Rogen's Hilarity for Charity		TV-MA	s1	Dick Johnson Is Dead	PG-13
s4938	Ram Dass, Going Home		TV-14	s1	Dick Johnson Is Dead	PG-13
s4936	Amateur		TV-MA	s1	Dick Johnson Is Dead	PG-13
s4935	6 Balloons		TV-MA	s1	Dick Johnson Is Dead	PG-13