



NETFLIX DATA ANALYSIS

USING SQL

INTRODUCTION TO NETFLIX

Netflix is a global streaming service that offers a wide variety of television shows, movies, documentaries, and more across a broad range of genres and languages. It was founded in 1997 by Reed Hastings and Marc Randolph in California, USA, originally as a DVD rental service. Over the years, Netflix has evolved into one of the largest online entertainment platforms, with a presence in nearly every country worldwide.

Data analysis plays a crucial role in shaping and optimizing content strategy for platforms like Netflix.

It helps in --

- 1) Understanding Viewer Preferences
- 2) Personalization and Recommendation Systems
- 3) Content Performance Evaluation
- 4) Content Acquisition and Licensing
- 5) Optimizing Content Releases
- 6) Audience Segmentation
- 7) Competitive Analysis
- 8) Predictive Analytics

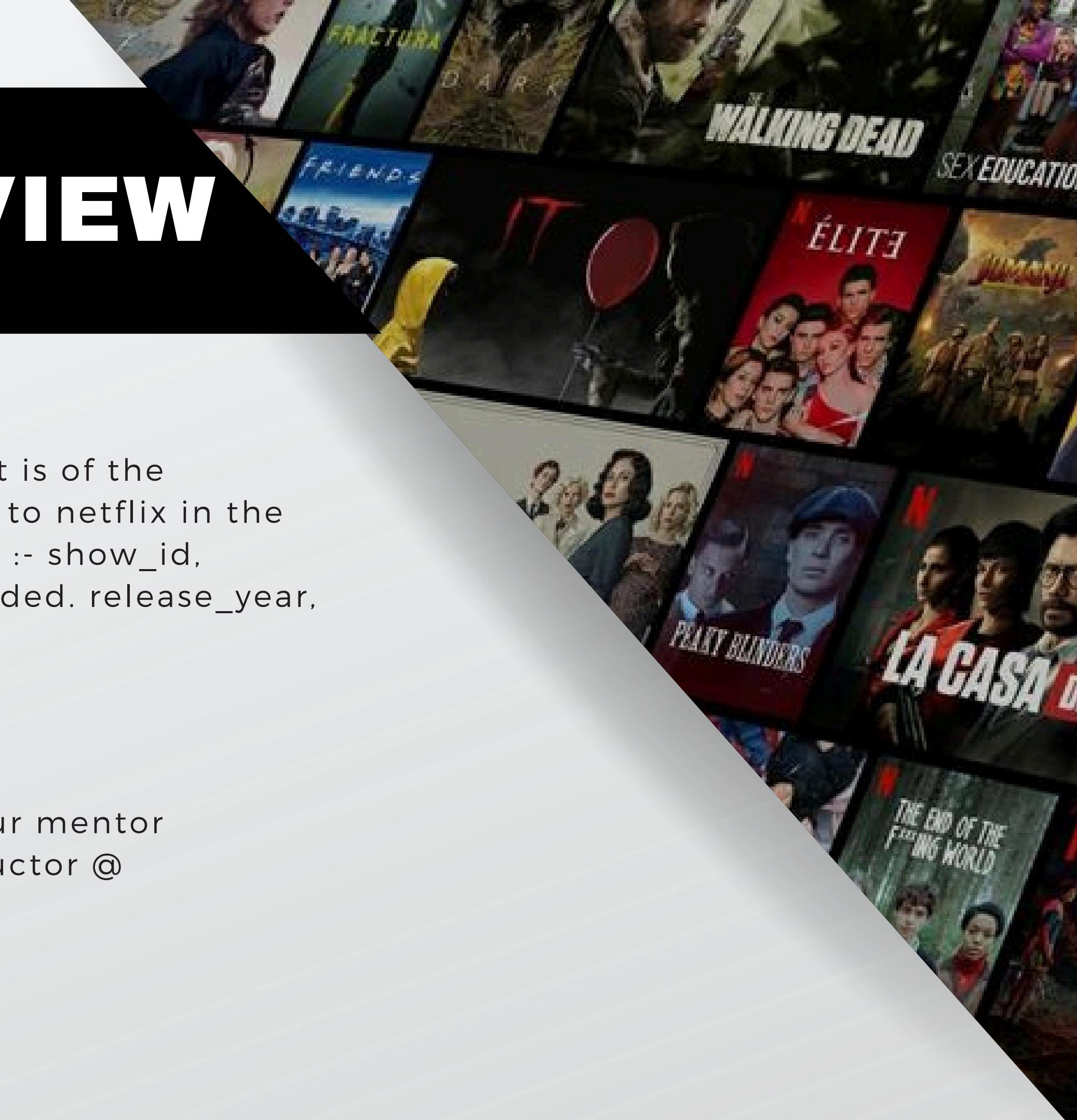
DATASET OVERVIEW



The dataset used in this project is of the shows/movies that were added to netflix in the year 2021. The columns include :- show_id, director, cast, country, date_added, release_year, rating, duration, listed_in and description

,

The dataset was provided by our mentor Krishna Madan Sir, SDE & Instructor @ Geekster.



QUERY 1

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

```
7 #1. Write a query to list all titles with their show_id, title, and type.  
8 -- Jhalak Bora  
9• select show_id,title,type from shows_2021;  
a
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

| 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 |
| | | |

QUERY 2

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

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

-- Jhalak Bora

```
select * from shows_2021 where type = 'Movie';
```

| w_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in |
|------|-------|----------------------------------|-------------------------------|---|--|------------|--------------|--------|----------|---------------------|
| | Movie | Dick Johnson Is Dead | Kirsten Johnson | | United States | 2021-09-25 | 2020 | PG-13 | 90 min | Documentaries |
| | Movie | My Little Pony: A New Generation | Robert Cullen, José Luis Ucha | Vanessa Hudgens, Kimiko Glenn, James Marsden... | | 2021-09-24 | 2021 | PG | 91 min | Children & Family |
| | Movie | Sankofa | Haile Gerima | Kofi Ghanaba, Oyafunmike Ogunlano, Alexander... | United States, Ghana, Burkina Faso, United Kingdom | 2021-09-24 | 1993 | TV-MA | 125 min | Dramas, Independent |
| | Movie | The Starling | Theodore Melfi | Melissa McCarthy, Chris O'Dowd, Kevin Kline, Tim... | United States | 2021-09-24 | 2021 | PG-13 | 104 min | Comedies, Drama |

QUERY 3

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

```
l9 #3. Write a query to list TV shows that were released in the year 2021.  
20 -- Jhalak Bora  
21 • select * from shows_2021  
22      where type = 'TV Show' AND release_year = 2021;
```

| show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---------|---------|-----------------------|----------------|---|--------------|------------|--------------|--------|-----------|--|---|
| s2 | TV Show | Blood & Water | | Ama Qamata, Khosi Ngema, Gail Mabalane, Tha... | South Africa | 2021-09-24 | 2021 | TV-MA | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Tow |
| s3 | TV Show | Ganglands | Julien Ledercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... | | 2021-09-24 | 2021 | TV-MA | 1 Season | Crime TV Shows, International TV Shows, TV Ac... | To protect his family from a powerful drug l |
| s4 | TV Show | Jailbirds New Orleans | | | | 2021-09-24 | 2021 | TV-MA | 1 Season | Docuseries, Reality TV | Feuds, flirtations and toilet talk go down am |
| s5 | TV Show | Kota Factory | | Mayur More, Jitendra Kumar, Ranjan Raj, Alam ... | India | 2021-09-24 | 2021 | TV-MA | 2 Seasons | International TV Shows, Romantic TV Shows, TV... | In a city of coaching centers known to train |

QUERY 4

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

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

27 -- Jhalak Bora

```
28 • select title from shows_2021  
29     where description like '%family%';
```

30

| Result Grid | |
|----------------------|--------------------|
| | Filter Rows: |
| | Export: |
| | Wrap Cell Content: |
| title | |
| ▶ Ganglands | |
| Je Suis Karl | |
| Dark Skies | |
| Omo Ghetto: the Saga | |
| Shadow Parties | |

QUERY 5

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

```
2 #5. Write a query to count the total number of titles in the dataset.  
3 -- Jhalak Bora  
4 • select count(title) from shows_2021;  
5
```

| Result Grid | Filter Rows: | Export: | Wrap Cell Content: |
|--------------|--------------|---------|--------------------|
| count(title) | | | |
| 8788 | | | |

QUERY 6

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

```
37 #6. Write a query to find the average duration of all movies (in minutes, wherever the season is mentioned, consider 400 minutes per season).
38 -- Jhalak Bora
39 • ALTER TABLE shows_2021 ADD COLUMN duration_in_minutes INT;
40 • UPDATE shows_2021
41 SET duration_in_minutes =
42 CASE
43     WHEN duration LIKE '% min' THEN CAST(REPLACE(duration, ' min', '') AS UNSIGNED)
44     WHEN duration LIKE '% Seasons' THEN CAST(REPLACE(duration, ' Seasons', '') AS UNSIGNED) * 400
45     WHEN duration LIKE '% Season' THEN CAST(REPLACE(duration, ' Season', '') AS UNSIGNED) * 400
46 END;
47 • SELECT type,title,duration_in_minutes AS average_duration FROM shows_2021
48 WHERE type = 'movie';
49
```

| Result Grid | | |
|-------------|----------------------------------|------------------|
| type | title | average_duration |
| Movie | Dick Johnson Is Dead | 90 |
| Movie | My Little Pony: A New Generation | 91 |
| Movie | Sankofa | 125 |
| Movie | The Starling | 104 |
| Movie | Je Suis Karl | 127 |

QUERY 7

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

```
51  #7. Write a query to list the top 5 latest titles based on the date_added, sorted in descending order
52  -- Jhalak Bora
53 • select title,date_added from shows_2021
54      order by date_added DESC LIMIT 5;
```

Result Grid | Filter Rows: _____ | Export: | Wrap Cell Content: | Fetch rows:

| title | date_added |
|-----------------------|------------|
| 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 |

QUERY 8

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.

```
#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
-- Jhalak Bora
SELECT n1.show_id, n1.title, n1.director, n2.number_of_titles_by_director
FROM shows_2021 n1
JOIN (SELECT director, COUNT(*) AS number_of_titles_by_director FROM shows_2021 GROUP BY director) n2
ON n1.director = n2.director
ORDER BY n2.number_of_titles_by_director;
```

| show_id | title | director | number_of_titles_by_director |
|---------|-------------------------|--------------------------------|------------------------------|
| 209 | King's Ransom | Jeffrey W. Byrd | 1 |
| 210 | Kingdoms of the Sky | Steve Greenwood | 1 |
| 211 | Kingpin | Peter Farrelly, Bobby Farrelly | 1 |
| 212 | Kings | Deniz Gamze Ergüven | 1 |
| 217 | KL Zombi | Woo Ming Jin | 1 |
| 220 | Know Your Enemy - Japan | Frank Capra, Joris Ivens | 1 |

QUERY 9

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

```
66  #9. Write a query to find the total number of titles for each country. Display country and the count of titles.
67  -- Jhalak Bora
68 • select country,
69  count(*) as 'count_of_titles'
70  from shows_2021
71  group by country
72  order by count_of_titles desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

| country | count_of_titles |
|----------------|-----------------|
| United States | 2806 |
| India | 972 |
| | 830 |
| United Kingdom | 419 |
| Japan | 245 |
| South Korea | 199 |
| Canada | 181 |

QUERY 10

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.

```
#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.  
-- Jhalak Bora  
• select show_id,type,title,  
CASE  
when rating = 'G' then 'Family'  
when rating = 'PG' then 'Family'  
when rating = 'PG-13' then 'Family'  
when rating = 'TV-Y' then 'Kids'  
when rating = 'TV-Y7' then 'Kids'  
when rating = 'TV-G' then 'Kids'  
else  
'Adult'  
END AS CATEGORY  
from shows_2021;
```

| show_id | type | title | CATEGORY |
|---------|---------|----------------------|----------|
| 1 | Movie | Dick Johnson Is Dead | Family |
| 2 | TV Show | Blood & Water | Adult |
| 3 | TV Show | Ganglands | Adult |

QUERY 11

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

```
93  #11. Write a query to add a new column title_length to the titles table that calculates the length of each title.  
94  -- Jhalak Bora  
95 • alter table shows_2021 add column title_length INT;  
96 • SET SQL_SAFE_UPDATES = 0;  
97 • update shows_2021 set title_length = LENGTH(title) WHERE show_id IS NOT NULL;  
98 • select title,title_length from shows_2021;  
qq
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

| title | title_length |
|-----------------------|--------------|
| Dick Johnson Is Dead | 20 |
| Blood & Water | 13 |
| Ganglands | 9 |
| Jailbirds New Orleans | 21 |
| Kota Factory | 12 |

QUERY 12

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

```
101 #12. Write a query using an advanced function to find the title with the longest duration in minutes
102 -- Jhalak Bora
103 • SELECT title, duration_in_minutes FROM shows_2021 ORDER BY duration_in_minutes DESC LIMIT 1;
104
```

| Result Grid | |
|----------------|---------------------|
| title | duration_in_minutes |
| Grey's Anatomy | 6800 |

QUERY 13

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

```
#13. Create a view named RecentTitles that includes titles added in the last 30 days.  
-- Jhalak Bora  
CREATE VIEW RecentTitles AS select * from shows_2021  
    WHERE date_added >= CURDATE() - INTERVAL 30 DAY;  
SELECT * from RecentTitles;
```

QUERY 14

14. Write a query using a window function to rank titles based on their release_year within each country.

```
.13 #14. Write a query using a window function to rank titles based on their release_year within each country.  
.14 -- Jhalak Bora  
.15 • SELECT title, country, release_year, RANK() OVER (PARTITION BY country ORDER BY release_year)  
.16 AS release_year_rank FROM shows_2021;  
17
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

| title | country | release_year | release_year_rank |
|---|---------|--------------|-------------------|
| Pioneers: First Women Filmmakers* | | 1925 | 1 |
| Jis Desh Men Ganga Behti Hai | | 1960 | 2 |
| Mobile Suit Gundam I | | 1981 | 3 |
| Mobile Suit Gundam II: Soldiers of Sorrow | | 1981 | 3 |
| Mobile Suit Zeta | | 1992 | 5 |

QUERY 15

15. Write a query to calculate the cumulative count of titles added each month sorted by date_added.

```
L19 #15. Write a query to calculate the cumulative count of titles added each month sorted by date_added.  
L20 -- Jhalak Bora  
L21 • SELECT  
L22     date_format(date_added, '%Y/%m') AS month_added,  
L23     Count(*) AS titles_added  
L24 FROM shows_2021  
L25 GROUP BY month_added  
L26 ORDER BY month_added;  
L27
```

Result Grid | Filter Rows: _____ | Export: _____ | Wrap Cell Content:

| month_added | titles_added |
|-------------|--------------|
| HULL | 10 |
| 2008/01 | 1 |
| 2008/02 | 1 |
| 2009/05 | 1 |
| 2009/06 | 1 |

QUERY 16

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

```
L29 #17. Write a query to find the country with the highest average rating for titles. Use subqueries and aggregate functions to achieve this.  
L30 -- Jhalak Bora  
L31 • SELECT s.country, s.title AS title_with_highest_rating, s.rating AS highest_rating  
L32 FROM shows_2021 s  
L33 ⚡ INNER JOIN (  
L34     SELECT country,  
L35         MAX(rating) AS max_rating FROM shows_2021  
L36         GROUP BY country) AS max_ratings  
L37     ON s.country = max_ratings.country AND s.rating = max_ratings.max_rating;  
L38
```

| Result Grid | | |
|---|---------------------------|----------------|
| country | title_with_highest_rating | highest_rating |
| South Africa | French Toast | TV-PG |
| South Africa | Liefling | TV-PG |
| South Africa | Pretville | TV-PG |
| South Africa | The Letter Reader | TV-PG |
| United States, Ghana, Burkina Faso, United Kin... | Sankofa | TV-MA |

QUERY 17

17. 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.

```
0  #18. 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.
1  -- Jhalak Bora
2 • SELECT s1.show_id AS show_id_1, s1.title AS title_1, s1.rating AS rating_1, s2.show_id AS show_id_2, s2.title AS title_2, s2.rating AS rating_2
3   FROM shows_2021 s1
4     INNER JOIN shows_2021 s2 ON s1.country = s2.country
5       WHERE
6         s1.show_id < s2.show_id AND s1.rating > s2.rating;
7
```

| show_id_1 | title_1 | rating_1 | show_id_2 | title_2 | rating_2 |
|-----------|--|----------|-----------|-----------|----------|
| s2910 | Monster High: Electrified | TV-Y | s3 | Ganglands | TV-MA |
| s2859 | Calico Critters: Everyone's Big Dream Flying in t... | TV-Y | s3 | Ganglands | TV-MA |
| s2858 | Calico Critters Mini Episodes Clover | TV-Y | s3 | Ganglands | TV-MA |
| s2815 | Steam Team to the Rescue | TV-Y | s3 | Ganglands | TV-MA |
| -2912 | Dino & Dinosaur: Mine of Mystery | TV-Y | -2 | Ganglands | TV-MA |

THANK YOU
THANK YOU
THANK YOU
THANK YOU
THANK YOU

