

NETFLIX MOVIE DATA ANALYSIS

NAME :- MAHAK BANSAL

REG NO :- PCE21CS096

SEC :- B

-- Question:-

1. Write a query to list all titles with their show_id, title, and type.
2. Write a query to display all columns for titles that are Movies.
3. Write a query to list TV shows that were released in the year 2021.
4. Write a query to find all titles where the description contains the word family.
5. Write a query to count the total number of titles in the dataset.
6. Write a query to find the average duration of all movies (in minutes, wherever the season is mentioned, consider 400 minutes per season).
7. Write a query to list the top 5 latest titles based on the date_added, sorted in descending order.
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.
9. Write a query to find the total number of titles for each country. Display country and the count of titles.
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.
11. Write a query to add a new column title_length to the titles table that calculates the length of each title.
12. Write a query using an advanced function to find the title with the longest duration in minutes.
13. Create a view named RecentTitles that includes titles added in the last 30 days.
14. Write a query using a window function to rank titles based on their release_year within each country.
15. Write a query to calculate the cumulative count of titles added each month sorted by date_added.
16. Write a stored procedure to update the rating of a title given its show_id and new rating.
17. Write a query to find the country with the highest average rating for titles. Use subqueries and aggregate functions to achieve this.
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.

-- create database netflix for table NETFLIX

CREATE DATABASE netflix;

-- using database netflix for table data USE netflix; -- import table netflix from netflix.csv using 'Table Data Import Wizard' - refresh the SCHEMAS ---- SCHEMAS - -- > netflix -> Tables -> Table Data Import Wizard

-- viewing table Netflix

```
SELECT * FROM netflix;
```

#Mahak Bansal:- 1. Write a query to list all titles with their show_id, title, and type.

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

#Mahak Bansal:- 2. Write a query to display all columns for titles that are Movies.

```
SELECT * FROM netflix WHERE type = 'Movie';
```

#Mahak Bansal:- 3. Write a query to list TV shows that were released in the year 2021.

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

#Mahak Bansal:- 4. Write a query to find all titles where the description contains the word family.

```
select title from netflix where description LIKE '%family%';
```

#Mahak Bansal:- 5. Write a query to count the total number of titles in the dataset.

```
SELECT count(*) AS total_titles From netflix;
```

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

```
SELECT AVG( CASE
  WHEN duration LIKE '%seasons' THEN SUBSTRING(duration, 1, LENGTH(duration) -
    LENGTH('seasons')) * 400
  WHEN duration LIKE '%season' THEN 400
  ELSE duration
END) AS average_duration
FROM netflix where type = 'movie';
```

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

```
SELECT title FROM netflix order by date_added DESC limit 5;
```

#Mahak Bansal:- 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.

```
SELECT show_id, title, director, count(*) OVER (partition by director) as No_of_title_by_director  
FROM netflix;
```

#Mahak Bansal:- 9. Write a query to find the total number of titles for each country. Display country and the count of titles.

```
select country, count(*) as count_of_titles from netflix group by country;
```

#Mahak Bansal:- 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.

```
select title,rating, case  
when rating in ('G','PG','PG-13') THEN 'Family'  
when rating in ('TV-Y','TV-Y7','TV-G') THEN 'Kids'  
else 'Adult'  
end as category  
from netflix;
```

#Mahak Bansal:- 11. Write a query to add a new column title_length to the titles table that calculates the length of each title.

```
ALTER TABLE netflix  
ADD COLUMN title_length INT;
```

```
-- To update the new column in the table
```

```
SET SQL_SAFE_UPDATES = 0;
```

```
UPDATE netflix
```

```
SET title_length = LENGTH(title);  
SET SQL_SAFE_UPDATES = 1;  
  
-- checking the new column 'title_length' in table netflix  
SELECT title, title_length  
FROM netflix;
```

#Mahak Bansal:- 12. Write a query using an advanced function to find the title with the longest duration in minutes.

```
select title,case  
when duration like'%season%' then cast(substring_index(duration,' ',1) as unsigned) * 400  
else cast(substring_index(duration,' ',1)as unsigned)  
END AS duration_in_min  
from netflix  
order by duration_in_min desc limit 1;
```

#Mahak Bansal:- 13. Create a view named RecentTitles that includes titles added in the last 30 days.

```
CREATE VIEW RecentTitles AS  
SELECT * FROM netflix  
WHERE date_added >= DATE_SUB(CURDATE(), INTERVAL 30 DAY);  
  
-- to show view RecentTitles  
SELECT * FROM RecentTitles;
```

#Mahak Bansal:- 14. Write a query using a window function to rank titles based on their release_year within each country.

```
SELECT show_id,title, country,release_year,  
RANK() OVER (PARTITION BY country ORDER BY release_year DESC) AS ranking  
FROM netflix;
```

#Mahak Bansal:- 15. Write a query to calculate the cumulative count of titles added each month sorted by date_added.

```
SELECT
    DATE_FORMAT(date_added, '%Y-%m') AS month_added, COUNT(*) AS titles_added
FROM netflix
GROUP BY month_added
ORDER BY month_added;
```

Mahak Bansal:- 16. Write a stored procedure to update the rating of a title given its show_id and new rating.

```
DELIMITER $$
CREATE PROCEDURE UpdateRating(IN p_show_id VARCHAR(10), IN p_new_rating VARCHAR(10))
BEGIN UPDATE netflix SET rating = p_new_rating
WHERE show_id = p_show_id;
END $$
DELIMITER ;
```

rating before calling procedure

```
SELECT show_id, rating
FROM netflix
WHERE show_id = 's20';
```

-- to change the rating

```
SET SQL_SAFE_UPDATES = 0;
```

-- call UpdateRating

```
CALL UpdateRating('s20', 'PG-13');
```

-- rating after calling procedure

```
SELECT show_id, rating
```

```
FROM netflix

WHERE show_id = 's20';

SET SQL_SAFE_UPDATES = 1;
```

Mahak Bansal:- 17. Write a query to find the country with the highest average rating for titles. Use subqueries and aggregate functions to achieve this.

```
SELECT country, AVG(rating_value) AS average_rating
FROM (SELECT country,CASE rating
      WHEN 'G' THEN 1
      WHEN 'PG' THEN 2
      WHEN 'PG-13' THEN 3
      WHEN 'TV-Y' THEN 4
      WHEN 'TV-Y7' THEN 5
      WHEN 'TV-G' THEN 6
      ELSE 7
      END AS rating_value
FROM netflix
) AS rating_values
GROUP BY country
ORDER BY average_rating DESC LIMIT 1;
```

Mahak Bansal:- 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.

```
SELECT t1.show_id AS show_id_1,
      t1.title AS title_1,
      t1.rating AS rating_1,
      t2.show_id AS show_id_2,
      t2.title AS title_2,
      t2.rating AS rating_2
```

```
FROM netflix t1  
JOIN netflix t2 ON t1.country = t2.country  
WHERE t1.show_id < t2.show_id AND t1.rating > t2.rating;
```