Netflix Data Analysis with SQLite

Dataset:

Used a structured dataset (`netflix_data`) to explore and analyze Netflix titles using SQL queries in **DB Browser for SQLite**.

Objectives:

- Learn to query structured data using `SELECT`, `WHERE`, `GROUP BY`, etc.
- Perform joins, subqueries, and aggregation functions.
- Extract useful insights.
- Practice SQL views, joins, and subqueries.

Tools:

- **SQLite** (DB Browser for SQLite)
- SQL (Structured Query Language)

Step-by-Step SQL Tasks

1. View Dataset Structure & Contents (select)

```
SELECT * FROM netflix_data;
PRAGMA table_info(netflix_data);
```

- Inspected the table and columns.
- Verified fields like `title`, `type`, `release_year`, `country`, `rating`, `duration`.

2. Movies After 2020 (where, order by)

```
SELECT title, release_year, type
FROM netflix_data
WHERE type = 'Movie' AND release_year > 2020
ORDER BY release_year DESC;
```

- Lists movies released after 2020.

3. Count Movies and TV Shows (group by)

```
SELECT type, COUNT(*) AS count
FROM netflix_data
GROUP BY type;
```

- Shows total number of Movies and TV Shows.

4. Created New Table for Rating Audiences (create table, insert into)

```
DROP TABLE IF EXISTS ratings_info;
CREATE TABLE ratings_info (rating TEXT PRIMARY KEY, audience TEXT);
```

```
INSERT INTO ratings_info (rating, audience) VALUES
('PG', 'Family'),
('R', 'Adults'),
('TV-MA', 'Mature'),
('TV-Y', 'Kids');
```

- Created a helper table to label audiences for content ratings.

5. Joins – Combine Movie Titles with Audience (left join, inner join)

```
-- LEFT JOIN
SELECT netflix_data.title, netflix_data.rating, ratings_info.audience
FROM netflix_data
LEFT JOIN ratings_info ON netflix_data.rating = ratings_info.rating
LIMIT 15;

-- INNER JOIN
SELECT n.title, n.rating, r.audience
FROM netflix_data AS n
INNER JOIN ratings_info AS r ON n.rating = r.rating;
```

- Merged `netflix_data` with `ratings_info` using both INNER and LEFT JOIN.

6. Longest Movie Duration (Subquery)

```
SELECT title, duration
FROM netflix_data
WHERE type = 'Movie'
ORDER BY CAST(SUBSTR(duration, 1, INSTR(duration, ' ') - 1) AS INTEGER)
DESC
LIMIT 1;
```

- Found the movie with the longest duration using string manipulation and sorting.

7. Top 5 Countries with Most Netflix Titles

```
SELECT country, COUNT(*) AS total
FROM netflix_data
GROUP BY country
ORDER BY total DESC
LIMIT 5;
```

- Shows top 5 countries with the most content in the dataset.

8. Create and Query a View for Recent Movies (create view)

```
DROP VIEW IF EXISTS recent_movies;

CREATE VIEW recent_movies AS
SELECT title, release_year, country
FROM netflix_data
WHERE type = 'Movie';

SELECT * FROM recent_movies
ORDER BY release_year DESC
```

```
LIMIT 5;
```

- Created a reusable view to easily access recent movies.

9. SUM – Total Number of Movies

```
SELECT SUM(CASE WHEN type = 'Movie' THEN 1 ELSE 0 END) AS total_movies
FROM netflix_data;
```

- Counted total number of movies using `SUM + CASE WHEN`.

AVG – Average Movie Release Year

```
SELECT AVG(release_year) AS avg_movie_year
FROM netflix_data
WHERE type = 'Movie';
```

- Calculated average release year for all movies.

Insights:

- USA has the most content in the dataset.
- A large number of shows are rated for mature audiences (`TV-MA`).
- Netflix releases a steady stream of content, with recent years being particularly active.