**SQL Aggregate Functions**

**By Tonny Odhiambo**

**Title: Analysis of Netflix Titles Dataset**

**Introduction**

In this report, we delve into an analysis of the Netflix titles dataset, aiming to uncover notable trends and insights regarding the content available on the platform.

**Import Process and Interesting Findings:**

To commence my exploration, I imported the Netflix titles dataset into a SQL database, facilitating efficient analysis. During this process, several compelling observations emerged, shedding light on the diversity and characteristics of Netflix's content library.

I leveraged on my skills in python to do this by writing a script to automate the data importation. The script will be attached with this report in the github repository.

**Cool Facts**:

Among the intriguing discoveries was the revelation of the most popular genres across Netflix titles, showcasing the diverse interests of the platform's audience. Additionally, an examination of the average duration of movies and TV shows provided nuanced insights into viewing preferences and content formats.

**Questions and SQL Queries**:

Formulating targeted questions allowed for a deeper examination of the dataset. By employing SQL queries, we were able to extract pertinent information to address these inquiries, unveiling valuable insights into content distribution and audience engagement**.**

CREATE DATABASE netflix\_db;

USE netflix\_db;

CREATE TABLE netflix\_titles(

show\_id VARCHAR(10) PRIMARY KEY,

type VARCHAR(10),

title VARCHAR(255),

director VARCHAR(255),

cast TEXT,

country VARCHAR(255),

date\_added DATE,

release\_year INT,

rating VARCHAR(10),

duration VARCHAR(50),

listed\_in TEXT,

description TEXT)

select \* from netflix\_titles;

**Two Cool Facts Hidden Within the Data:**

a. **Most Popular Genres:** To find out which genres are the most popular, we can count the occurrences of each genre and then sort them in descending order.

-- Most Popular Genres:

SELECT listed\_in, COUNT(\*) as count

FROM netflix\_titles

GROUP BY listed\_in

ORDER BY count DESC

LIMIT 5;

b. **Average Duration of Movies and TV Shows:** We can calculate the average duration of movies and TV shows separately.

-- Average Duration of Movies and TV Shows:

SELECT type, AVG(duration) as avg\_duration

FROM netflix\_titles

GROUP BY type;

**Two Questions About the Data:**

a. **Which countries have the highest number of TV shows?** To answer this question, we'll count the number of TV shows for each country and then sort them in descending order.

-- Which countries have the highest number of TV shows

SELECT country, COUNT(\*) as num\_tv\_shows

FROM netflix\_titles

WHERE type = 'TV Show'

GROUP BY country

ORDER BY num\_tv\_shows DESC

LIMIT 5;

b. **What are the top 10 longest movies available on Netflix?** We can find the top 10 longest movies by sorting them based on duration in descending order.

-- What are the top 10 longest movies available on Netflix

SELECT title, duration

FROM netflix\_titles

WHERE type = 'Movie'

ORDER BY duration DESC

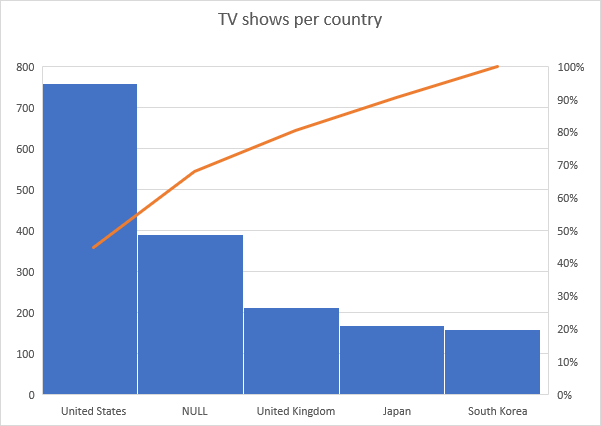
LIMIT 10;

Here is an exported dump from my sql to show all the queries executed.



A Pie chart showing the most popular genres according to the Netflix dataset.

A bar graph showing Which countries have the highest number of TV shows.



A bar graph to show number of titles by country.