**DANIEL-CALEB CHERUIYOT**

**cheruiyotdanielcaleb@gmail.com**

**Picked the Netflix Titles CSV File**

**Difficulties**

Inconsistent Data Entries. The dataset contains inconsistent or missing entries, especially in columns like Director, Cast, and Country, where some fields are blank.

Categorical Diversity. The Listed in column contains multiple genres per entry, which can make it challenging to filter or group data by a single genre.

**Interesting Observation:**

The dataset contains a variety of content from different countries. This diversity can be explored to see which genres or types of content are prevalent in different regions.

2. Data Fun (20 pts):

a) Most Popular Genres - Used the COUNT function to determine which genres are most popular.

SELECT listed\_in, COUNT(\*) AS Count

FROM netflix\_titles

GROUP BY listed\_in

ORDER BY Count DESC

LIMIT 5;

b) Average Duration of Shows

SELECT AVG(duration) AS Avgduration

FROM netflix\_titles;

3. Ask Away (30 pts):

a) What are the top-rated shows in the dataset? (Ordered by rating)

SELECT title, rating

FROM netflix\_titles

ORDER BY rating DESC

LIMIT 10;

*----You’ll get the titles and ratings of the highest-rated shows.*

b) What are the top 5 countries producing the most Netflix shows?

SELECT country, COUNT(\*) AS ShowCount

FROM netflix\_titles

GROUP BY country

ORDER BY ShowCount DESC

LIMIT 5;

*--- This query will reveal the top 5 countries with the most shows.*