OUMA ALVINE OTIENO WEEK 15/16

1. Data Dive (10 pts):
   - Pick your dataset and click to download (Social Media Users, Netflix Shows, or Human Stampedes).
   - Import it into MySQL Workbench (learn how!).
   - Briefly explain any difficulties and 1 interesting thing you noticed about your chosen dataset.

DATASET PICKED

Netflix shows

DIFFICULTIES

- **Data Cleaning**: Ensuring the CSV data aligns with the table schema can be challenging, especially with varying data types and missing values.
- **Date Format**: Converting the date_added field from the CSV to the MySQL date format required additional attention.

INTERESTING THING

Diverse Content: The dataset contains a wide variety of genres and types, including movies and TV shows from many different countries, showcasing Netflix's global reach and diverse content library.

2. Data Fun (20 pts):
   - Use simple SQL queries to play with the data.
   - Find 2 cool facts hidden within the data (e.g., most popular interests).
   - Use basic SQL queries like (COUNT, AVG, and SUM) to understand more about the data you have.

SQL QUERIES

```
SELECT show_id FROM netflix_titles;
SELECT show_id FROM netflix_titles WHERE show_id = 's1';

-- dropping a table tha already exists
DROP TABLE IF EXISTS netflix_titles;
CREATE TABLE netflix_titles (
   show_id VARCHAR(50) PRIMARY KEY,
   type VARCHAR(50),
   title VARCHAR(255),
   director VARCHAR(255),
   cast TEXT,
   country VARCHAR(100),
   date_added DATE,
   release_year INT,
   rating VARCHAR(50),
   duration VARCHAR(50),
   listed_in VARCHAR(255),
   description TEXT
);
--inserting data
INSERT INTO netflix_titles (show_id, type, title, director, cast, country, date_added, release_year, rating, duration, listed_in, description) VALUES ('s1', 'Movie', 'Inception', 'Christopher Nolan', 'Leonardo
```

DiCaprio, Joseph Gordon-Levitt', 'United States', '2021-01-01', 2010, 'PG-13', '148 min', 'Sci-Fi, Thriller', 'A thief who steals corporate secrets through the use of dream-sharing technology.');

--Check if the table structure matches your requirements
DESCRIBE netflix_titles;

COOL HIDDEN FACTS

1.Most Common Genre

```
SELECT listed_in, COUNT(*) AS genre_count
FROM netflix_titles
GROUP BY listed_in
ORDER BY genre_count DESC
LIMIT 1;
```

2. Average Release Year of Movies and TV Shows

```
SELECT type, AVG(release_year) AS avg_release_year
FROM netflix_titles
GROUP BY type;
```

Use basic SQL queries like (COUNT, AVG, and SUM) to understand more about the data you have.

```
--Total Number of Titles
SELECT COUNT(*) AS total_titles
FROM netflix_titles;

--Total Number of Movies and TV Shows
SELECT type, COUNT(*) AS count
FROM netflix_titles
GROUP BY type;

--Average Release Year
SELECT type, AVG(release_year) AS avg_release_year
FROM netflix_titles
GROUP BY type;
```

3.Ask Away (30 pts):

- Formulate 2 questions about the data (e.g., what are popular shows in different countries?).
- Write basic SQL queries (WHERE, ORDER BY) to find answers.
- Share what you learned from the answers.

Formulate 2 questions about the data

1.What are the most popular genres in the top 3 countries with the most titles?

```
SELECT country, COUNT(*) AS titles_count
FROM netflix_titles
GROUP BY country
ORDER BY titles_count DESC
LIMIT 3;
```

2.What are the highest-rated movies released in the last 5 years?

```
SELECT title, rating, release_year
FROM netflix_titles
WHERE type = 'Movie' AND release_year BETWEEN 2019 AND 2024
ORDER BY rating DESC;
```

What I have learnt

By running this query, I can discover which recent movies have received the highest ratings.
This insight can inform us about recent trends in movie production and audience preferences.
For example, if many highly-rated movies fall into a specific genre, it may indicate a growing interest in that genre among viewers.