Leah Kiora week15 Assignment

**1.DataDive**

Netflix shows

-- CREATE DATABASE Netflix;

-- USE Netflix;

SELECT \*

FROM netflix\_titles;

**Results;**

CREATE DATABASE Netflix

USE Netflix

SHOW SESSION VARIABLES LIKE 'lower\_case\_table\_names'

SHOW DATABASES

SHOW SESSION VARIABLES LIKE 'lower\_case\_table\_names'

SHOW TABLES FROM `netflix` like 'netflix\_titles'

CREATE TABLE `netflix`.`netflix\_titles` (`show\_id` text, `type` text, `title` text, `director` text, `cast` text, `country` text, `date\_added` text, `release\_year` int, `rating` text, `duration` text, `listed\_in` text, `description` text)

PREPARE stmt FROM 'INSERT INTO `netflix`.`netflix\_titles` (`show\_id`,`type`,`title`,`director`,`cast`,`country`,`date\_added`,`release\_year`,`rating`,`duration`,`listed\_in`,`description`) VALUES(?,?,?,?,?,?,?,?,?,?,?,?)'

DEALLOCATE PREPARE stmt

SELECT \* FROM netflix\_titles LIMIT 0, 1000

I didn’t find any difficulties , and the most interesting thing about this dataset  **Duration and Seasons**: Information about the duration of movies and the number of seasons for TV shows offers insight into the depth of content. For example, shows with multiple seasons like "The Great British Baking Show" suggest popularity and continued viewer interest.

Also **Ratings and Descriptions**: The inclusion of ratings e.g TV-MA, PG-13 and brief descriptions can help in understanding the target audience for each title. This data could be useful for parental controls and content recommendation systems.

**2.DataFun**

**-- Cool Fact 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;

**Results;**

listed\_in & genre\_count

'Action & Adventure, Anime Features, International Movies', '12'

**-- Cool Fact 2: Average Duration of Movies**

SELECT AVG(duration) AS avg\_movie\_duration

FROM netflix\_titles

WHERE type = 'Movie';

**Results;**

avg\_movie\_duration

'100.72727272727273'

**Basic sql quaries(COUNT,AVG and SUM)**

**1.Total Number of Movies and TV Shows**

SELECT type, COUNT(\*) AS count

FROM netflix\_titles

GROUP BY type;

**Results;**

type, count

'Movie', '55'

'TV Show', '45'

**2.Average Duration of All TV Show**

SELECT AVG(CAST(SUBSTRING(duration, 1, LENGTH(duration) - 4) AS DECIMAL)) AS avg\_TVShow\_duration

FROM netflix\_titles

WHERE type = 'TV Show';

**Results;**

avg\_TVShow\_duration

'2.3556'

**3.Sum of Seasons for All TV Shows**

SELECT SUM(CAST(SUBSTRING(duration, 1, LENGTH(duration) - 8) AS DECIMAL)) AS total\_seasons

FROM netflix\_titles

WHERE type = 'TV Show';

**Results;**

total\_seasons

'83'

**3.Ask Away**

Question 1: Which country has the highest number of TV shows?

SELECT country, COUNT(\*) AS tv\_show\_count

FROM netflix\_titles

WHERE type = 'TV Show'

GROUP BY country

ORDER BY tv\_show\_count DESC

LIMIT 1;

**Results;**

Country, tv\_show\_count

**'', '19'**

Question 2: What are the top 5 highest-rated movies?

SELECT title, rating

FROM netflix\_titles

WHERE type = 'Movie'

ORDER BY rating DESC

LIMIT 5;

**Results;**

title, rating

'Firedrake the Silver Dragon', 'TV-Y7'

'Little Singham - Black Shadow', 'TV-Y7'

'Go! Go! Cory Carson: Chrissy Takes the Wheel', 'TV-Y'

'A StoryBots Space Adventure', 'TV-Y'

'Confessions of an Invisible Girl', 'TV-PG'

**What I’ve learnt from the answers**

These insights help us understand content distribution by country and the types of movies that are considered more mature or have higher content ratings.

**MY SQL**

CREATE DATABASE Netflix;

USE Netflix;

SELECT \*

FROM netflix\_titles;

Cool Fact 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;

Cool Fact 2: Average Duration of Movies

SELECT AVG(duration) AS avg\_movie\_duration

FROM netflix\_titles

WHERE type = 'Movie';

-- Total Number of Movies and TV Shows

SELECT type, COUNT(\*) AS count

FROM netflix\_titles

GROUP BY type;

Average Duration of All TV Show

SELECT AVG(CAST(SUBSTRING(duration, 1, LENGTH(duration) - 4) AS DECIMAL)) AS avg\_TVShow\_duration

FROM netflix\_titles

WHERE type = 'TV Show';

Sum of Seasons for All TV Shows

SELECT SUM(CAST(SUBSTRING(duration, 1, LENGTH(duration) - 8) AS DECIMAL)) AS total\_seasons

FROM netflix\_titles

WHERE type = 'TV Show';

Question 1: Which country has the highest number of TV shows?

SELECT country, COUNT(\*) AS tv\_show\_count

FROM netflix\_titles

WHERE type = 'TV Show'

GROUP BY country

ORDER BY tv\_show\_count DESC

LIMIT 1;

Question 2: What are the top 5 highest-rated movies?

SELECT title, rating

FROM netflix\_titles

WHERE type = 'Movie'

ORDER BY rating DESC

LIMIT 5;