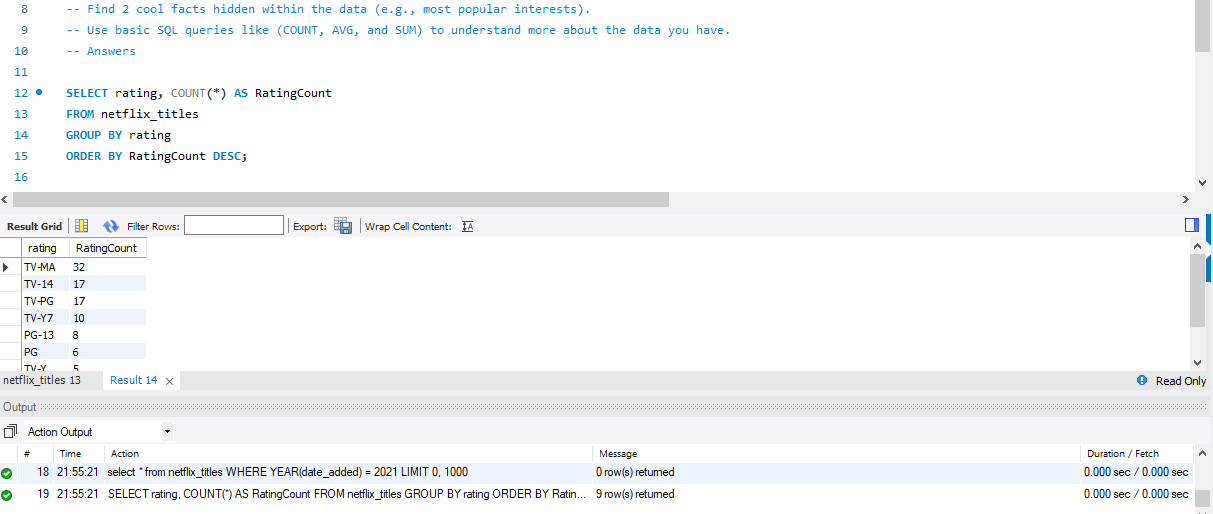
**Question 1 (Data Dive)**

**Briefly explain any difficulties and 1 interesting thing you noticed about your chosen datasets in your README section of your GitHub repository.**

I noticed that SocialMediaUsers and HumanStampede dataset failed to import to my workbench.

Netflix datasets was straight forward, while importing it, i noticed the columns were automatically assigned their data type.

**Question 2 (Data Fun)**



**Explanation:**

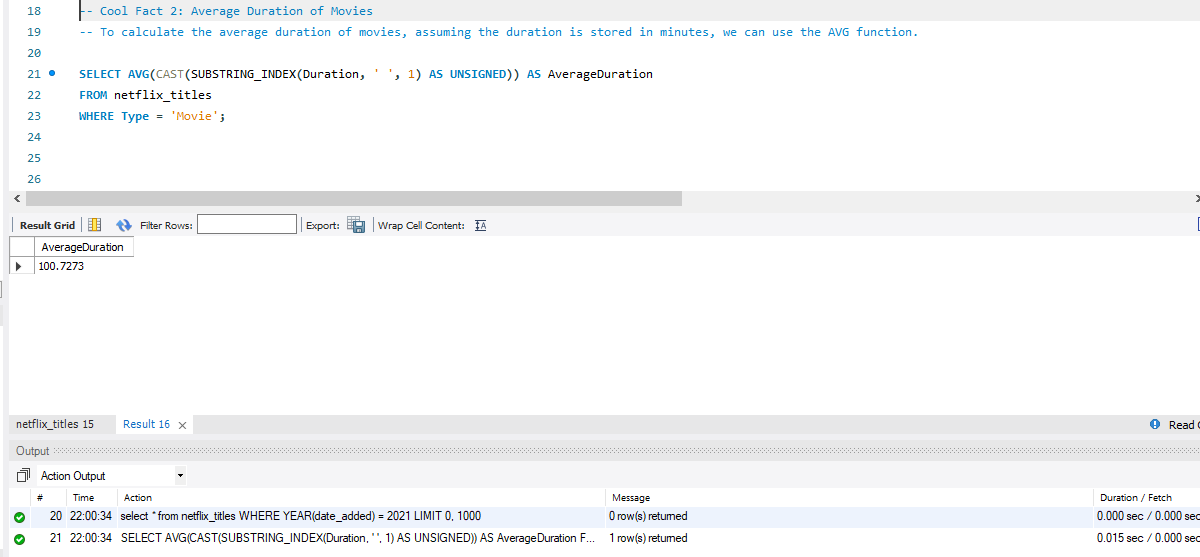
**SELECT rating, COUNT(\*) AS RatingCount:** This selects the rating and counts the number of occurrences of each rating.

**FROM netflix\_titles:** This specifies the table to query.

**GROUP BY rating:** This groups the results by the rating.

**ORDER BY RatingCount DESC:** This orders the results in descending order based on the count, showing the most common rating first.

**Average Duration of Movies**



**Explanation:**

**SELECT AVG(CAST(SUBSTRING\_INDEX(Duration, ' ', 1) AS UNSIGNED)) AS AverageDuration:** This calculates the average duration of movies. The SUBSTRING\_INDEX(Duration, ' ', 1) extracts the numeric part of the duration string (e.g., "90" from "90 min"), and CAST converts it to an integer.

**FROM netflix\_titles:** This specifies the table to query.

**WHERE Type = 'Movie':** This filters the results to only include movies.

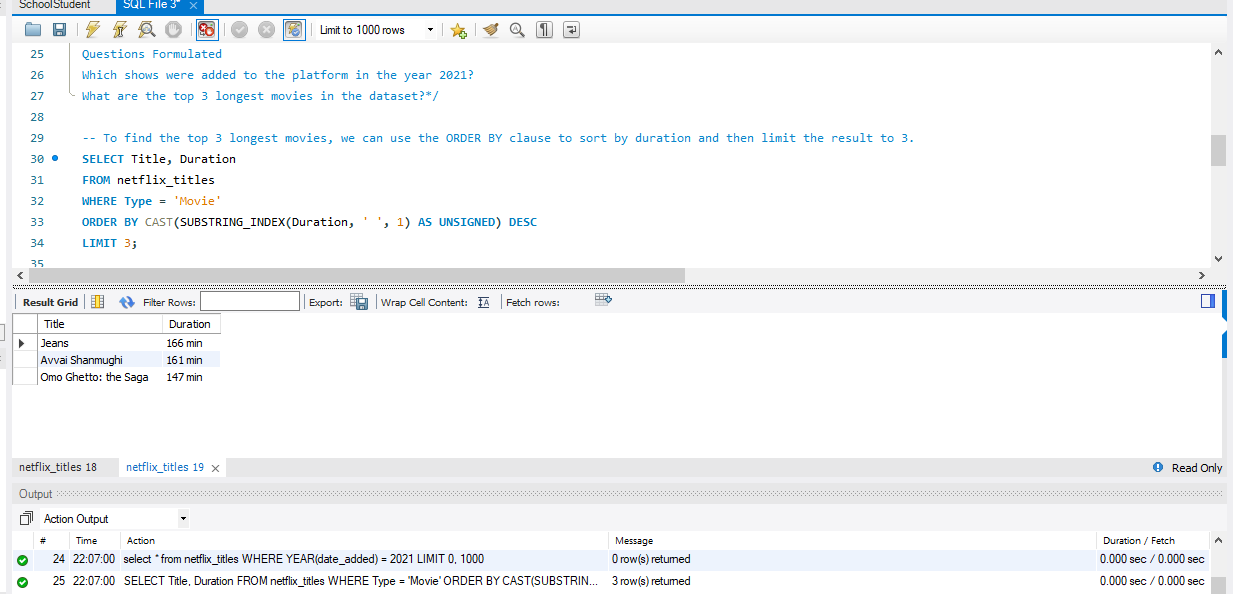
**Overview**

**Most Common Rating:** The most common rating among the shows is TV-MA, with a total count of 7.

**Average Duration of Movies:** The average duration of movies in the datasets is approximately 100.72 minutes.

**Question 3 (Ask away)**

**What are the top 3 longest movies in the datasets?**



**Explanation:**

**SELECT Title, Duration:** This selects the title and duration of the movies.

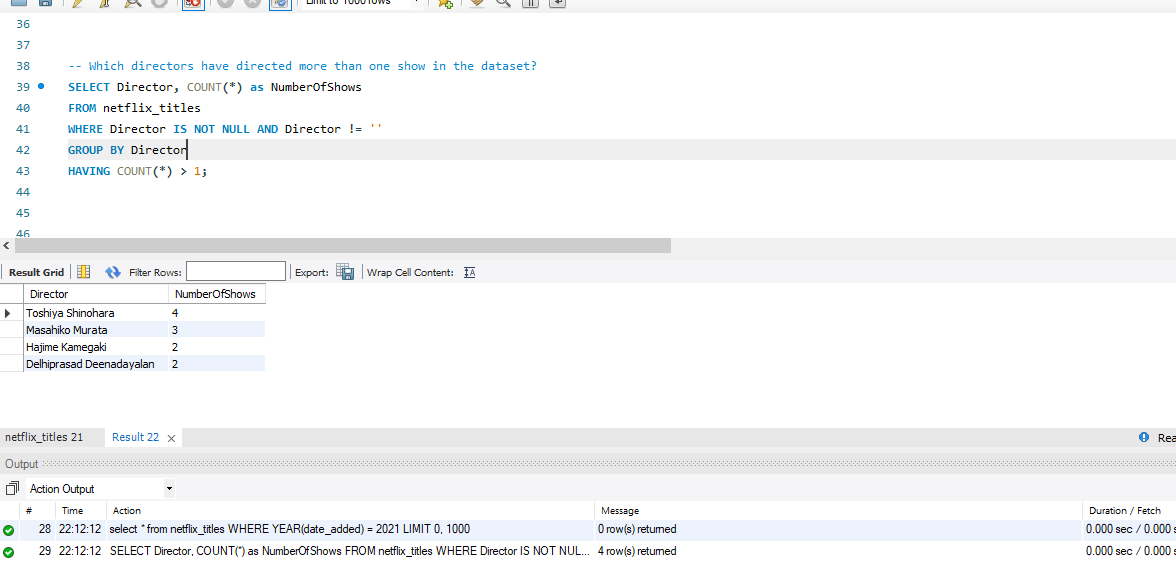
**FROM Shows:** This specifies the table to query.

**WHERE Type = 'Movie':** This filters the results to only include movies.

**ORDER BY CAST(SUBSTRING\_INDEX(Duration, ' ', 1) AS UNSIGNED) DESC:** This sorts the movies by duration in descending order.

**LIMIT 3:** This limits the result to the top 3 longest movies.

**Which directors have directed more than one show in the dataset?**



**Explanation:**

**SELECT Director, COUNT(\*) as NumberOfShows:** This selects the director and the count of their shows, with an alias NumberOfShows for readability.

**FROM Shows:** This specifies the table to query.

**WHERE Director IS NOT NULL AND Director != '':** This ensures only non-null and non-empty directors are considered.

**GROUP BY Director:** This groups the results by director.

**HAVING COUNT(\*) > 1:** This filters the groups to include only those with more than one show.

**LEARNINGS**

From the result, we can learn that certain directors are prolific within the datasets, having directed multiple shows. This insight could be useful for identifying key creators who contribute significantly to the platform's content library.