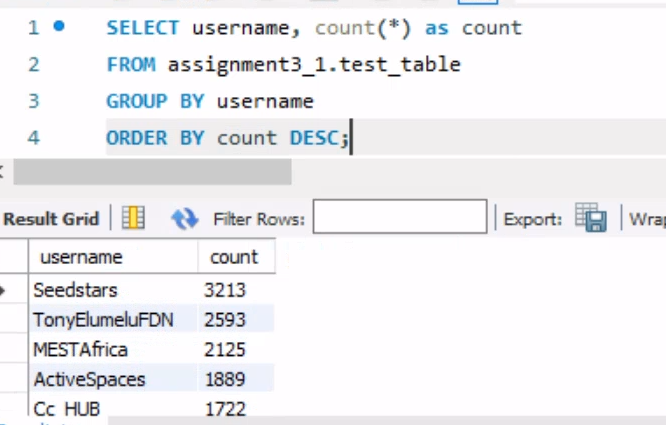
**Alberto Gell**

As a data analyst working for a company that is producing a social media application. The company is taking an innovative approach to social media. Before it develops the interface for messaging, the company wants to investigate the competition. The supervisor has asked us to investigate using a scraped data set. We are to find the average number of messages by a single username, the average number of reshares by a single username, and the time frame that has the highest number of original messages. Using SQL, the most appropriate techniques for investigating the scraped data set are aggregation and grouping functions.

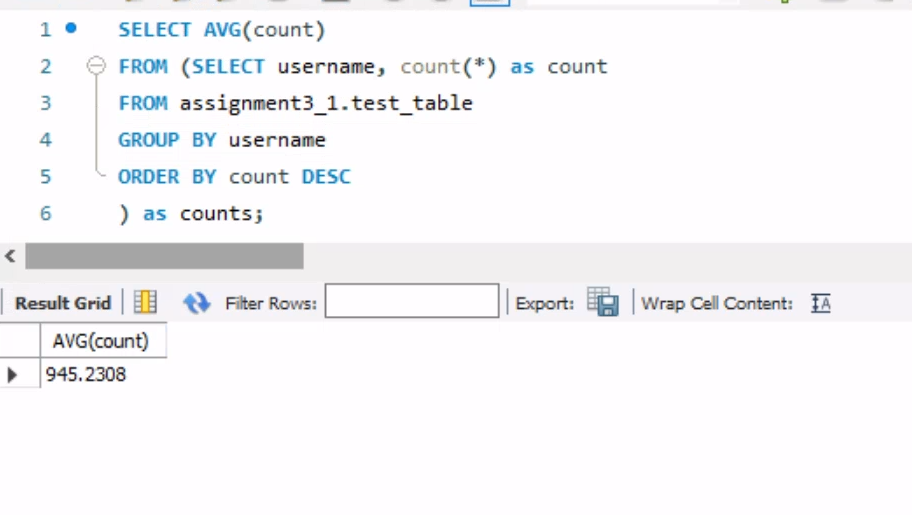
**Count by Username**

The following script counts(aggerates) the number of messages for each username and groups the total number of messages by username in descending order. The query illustrated that the username “Seedstars” has the highest number of messages totaling, 3213



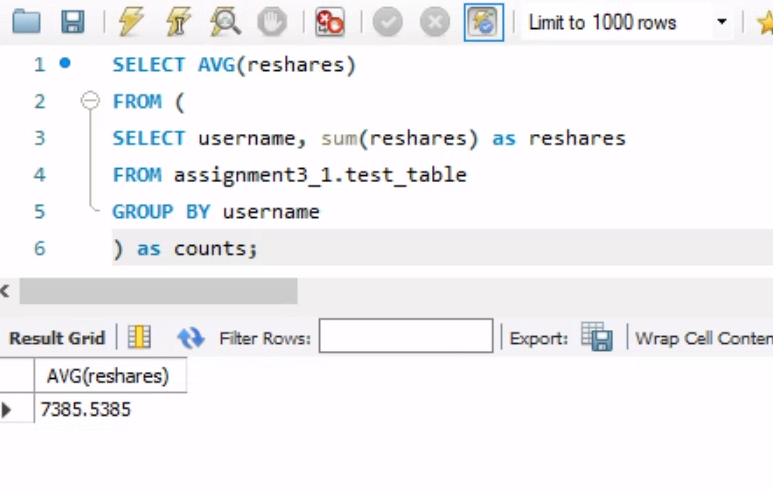
**Average Number of Messages by a Single Username**

The following script calculates the average number of messages per username. The inner query calculates the count of messages for each username and groups them accordingly. The outer query then calculates the average of these counts. Based on the aggregated results the average number of messages by a single username is 945



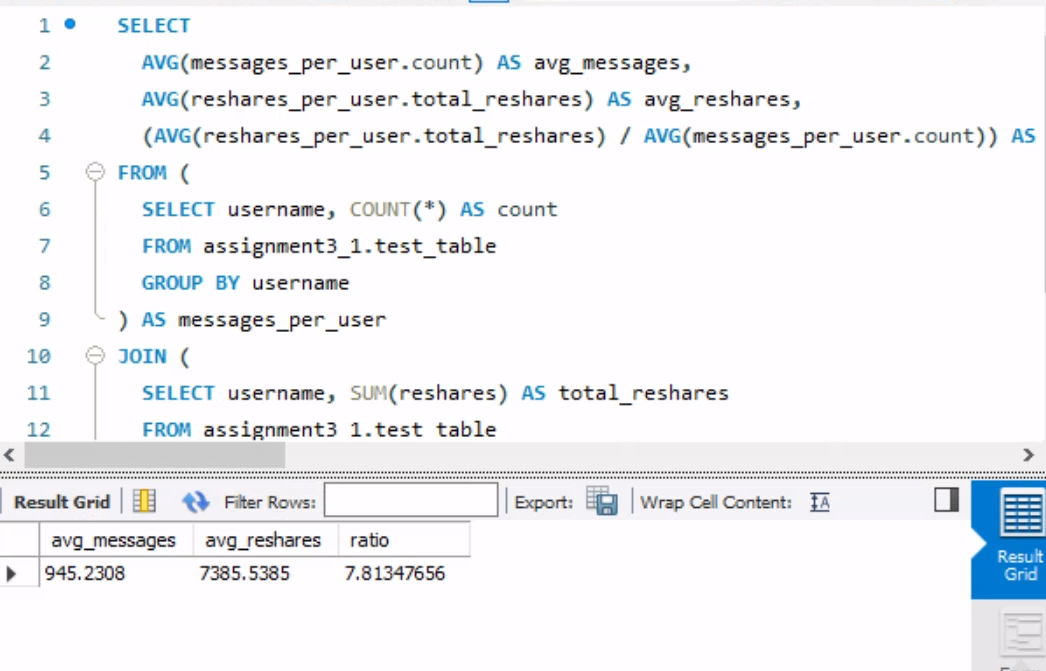
**The Average Number of Reshares by a Single Username**

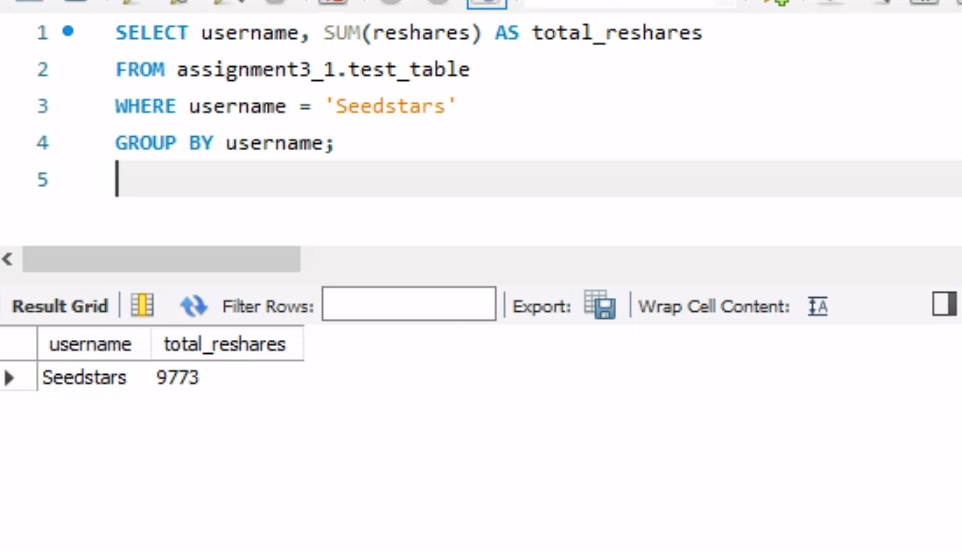
The following script calculates the average number of reshares per username, the results determined that the average number of reshares per username is 7385



**Ratio**

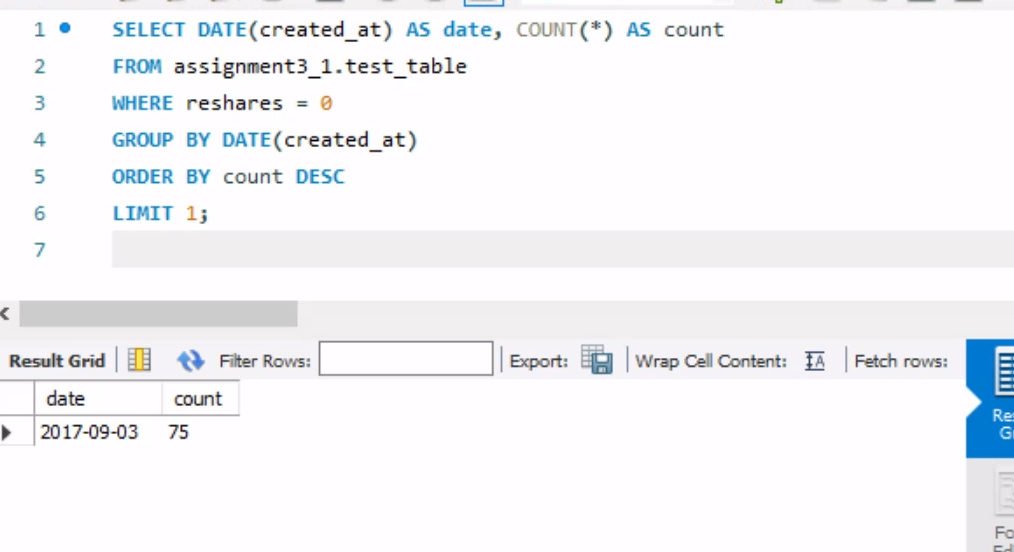
To validate the average number of reshares per username we wanted to determine what is the ratio between messages per username and reshares. The ratio is 1 to 7 To gain further insight into the analysis we hypothesized that the reshare ratio should thus be highest for the username with the highest messages. The analysis determined that username “Seedstars” has 1 to 10 ratio.





**Time Frame with the Highest Number of Original Messages**

This query identifies the date with the highest number of original messages. It uses WHERE reshares = 0 to filter out only the original messages, then groups by date and counts the messages, ordering by the count in descending order and limiting the result to the top one.



**Why were these techniques chosen?**

The scripts were tailored to fit the exact requirements, such as calculating averages and summing reshares. Two additional scripts were added to identify the peak time frame for original messages, which was not initially provided and the ratio between messages and reshares. To validate the results the queries executed in MySQL Workbench and the output was verified. These techniques are standard for handling and summarizing large datasets in relational databases. They are efficient and widely used for tasks involving counts, averages, and sums