TRAINITY PROJECT NUMBER 2

INSTAGRAM USER ANALYTICS

**MUSKAN BANSAL**

**PROJECT DESCRIPTION:**

When I first read the questions, I was quiet scared to even try but as I gave second and third read, it got simpler. When I finally imported all the data set into my SQL Workbench and ran the first question query that was asked according to my understanding and I got the right output, it gave me confidence. This project helped me build confidence and helped me understand what types of queries a data analyst is required to solve on a daily basis.

This project is about basic understanding of how data analysis is done on a basic level.

The project was divided into two parts:

**1. Marketing**

**2. Investor Metrics**

In the marketing part, I learned different ways a company deals with their services or products to increase the profit while keeping the losses to minumum. Main aim being to optimise the resources they have well enough to get the highest ROI. Rewarding the loyal customers so they stay connected to launching AD Campaigns. Starting new contests to increase the user engagement.

However, in the Metrics part, I learned how imporant it is for a company to know it’s users well, to have the knowledge of the user engagement and also to know about the ill practices if a user is following. Aim is to keep it real and as authentic as possible.

**APPROACH:**

My approch was pretty simple, I went through the provided datasets and the guven installation resources, the video on basic SQL Workbench helped me a lot and that’s how I learned to import the dataset. Even though there was some connecting issue with my SQL port because of which I spent an entire day first figuring out how to resolve that. I was unable to import data or even launch a new one due to porting issue. After troubleshooting, I finally imported the data and was ready to solve the queries.

I read the question twice and the began writing queries. There were times when I could think of the approach but didn’t really know how to execute them to I also went to SQL Documentation.

Question 1, 2 and 4 were easy however I took some time to solve question 3. I learned about the “dayname” function here for which I googled. The last two questions were just as easy.

I have explained below each answer how I have come to that solution and understanding.

**TECH-STACK Used:** I used MySQL Workbench 8.0 CE to run all the queries.

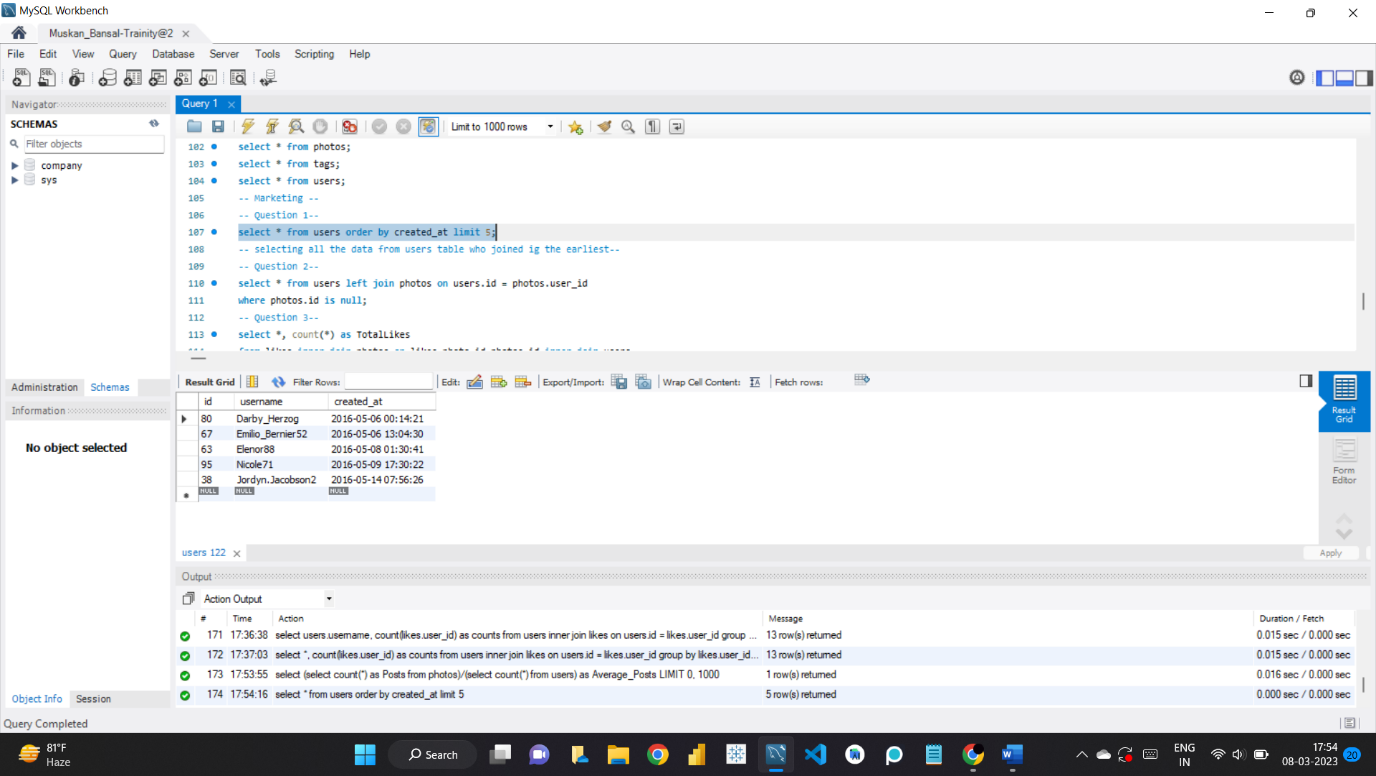
**ANSWERS:**

**-- Marketing --**

-- Question 1--

select \* from users order by created\_at limit 5;

-- selecting all the data of 5 users from ‘users’ table who joined instagram the earliest—



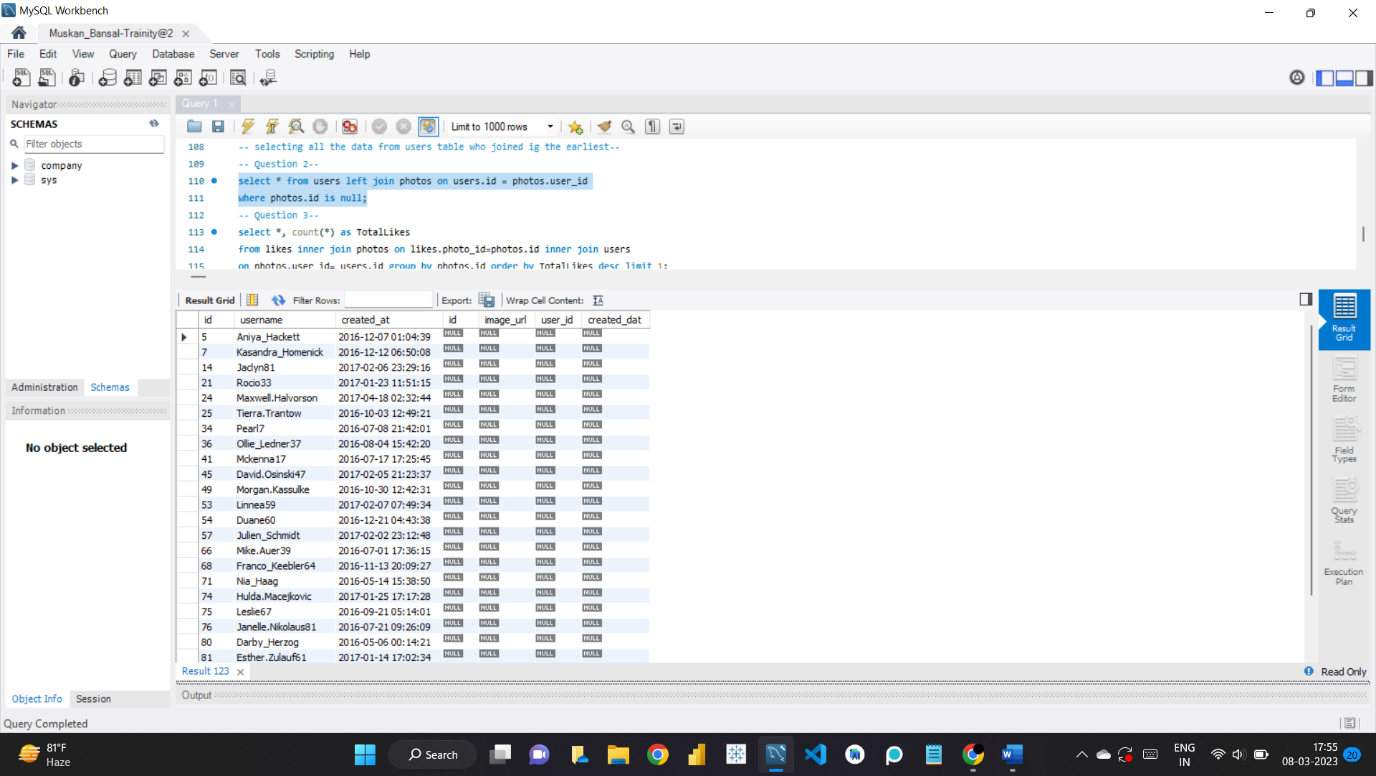
-- Question 2--

select \* from users left join photos on users.id = photos.user\_id

where photos.id is null;

--selecting all the data of the users from ‘users’ who have not posted once on instagram—

--here, I used ‘left join’ since I needed data from two tables that are; users and photos—

--I used left join, since I only need to find users who have not posted once—

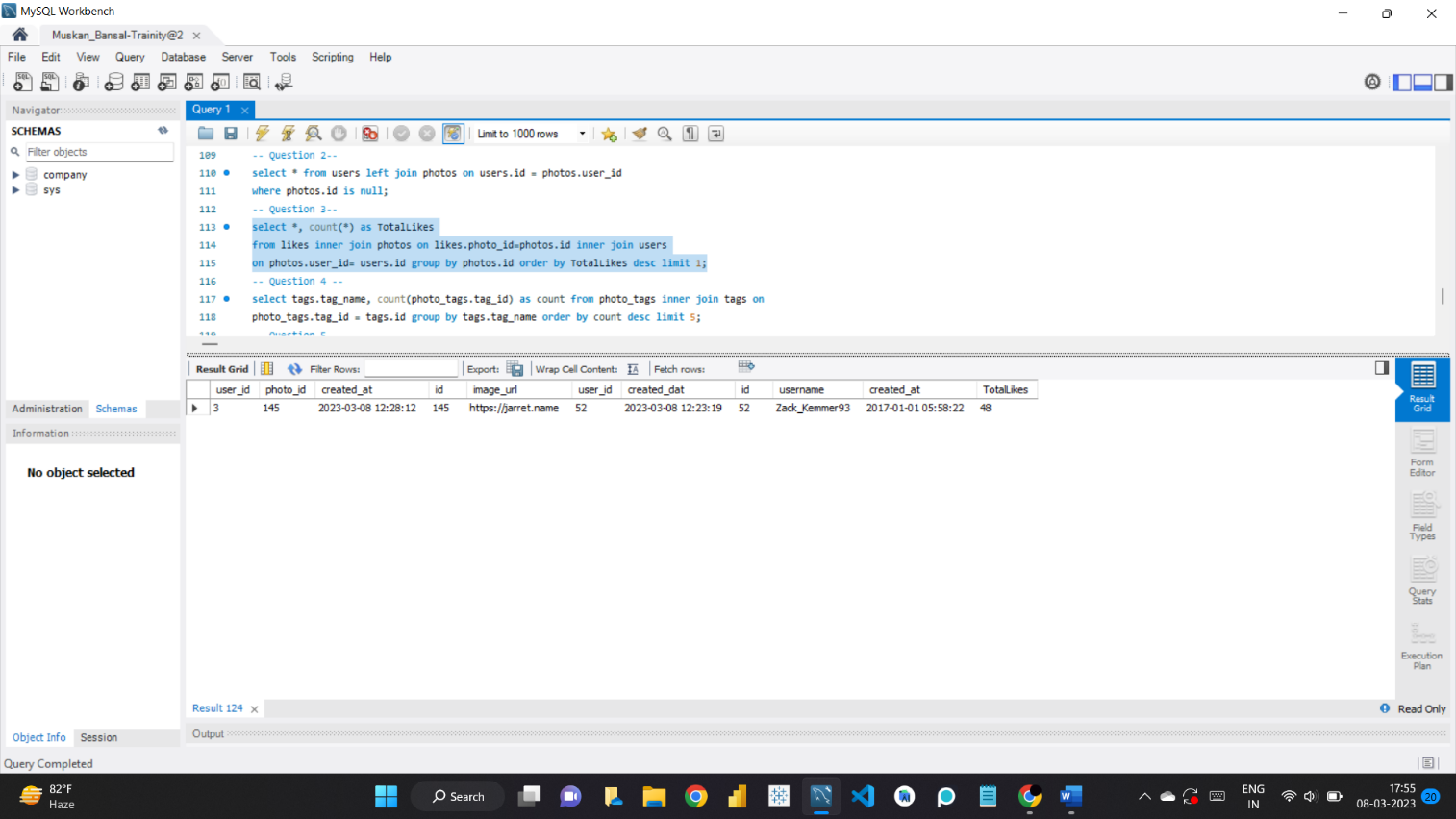
-- Question 3--

select \*, count(\*) as TotalLikes

from likes inner join photos on likes.photo\_id=photos.id inner join users

on photos.user\_id= users.id group by photos.id order by TotalLikes desc limit 1;

--I used ‘inner join’ since I needed to find the data from the tables: likes and photos who have posted and also had the most amount of likes—

--Then I ordered them i descending order and used ‘limit’ since I only need data of the user who got the most likes—

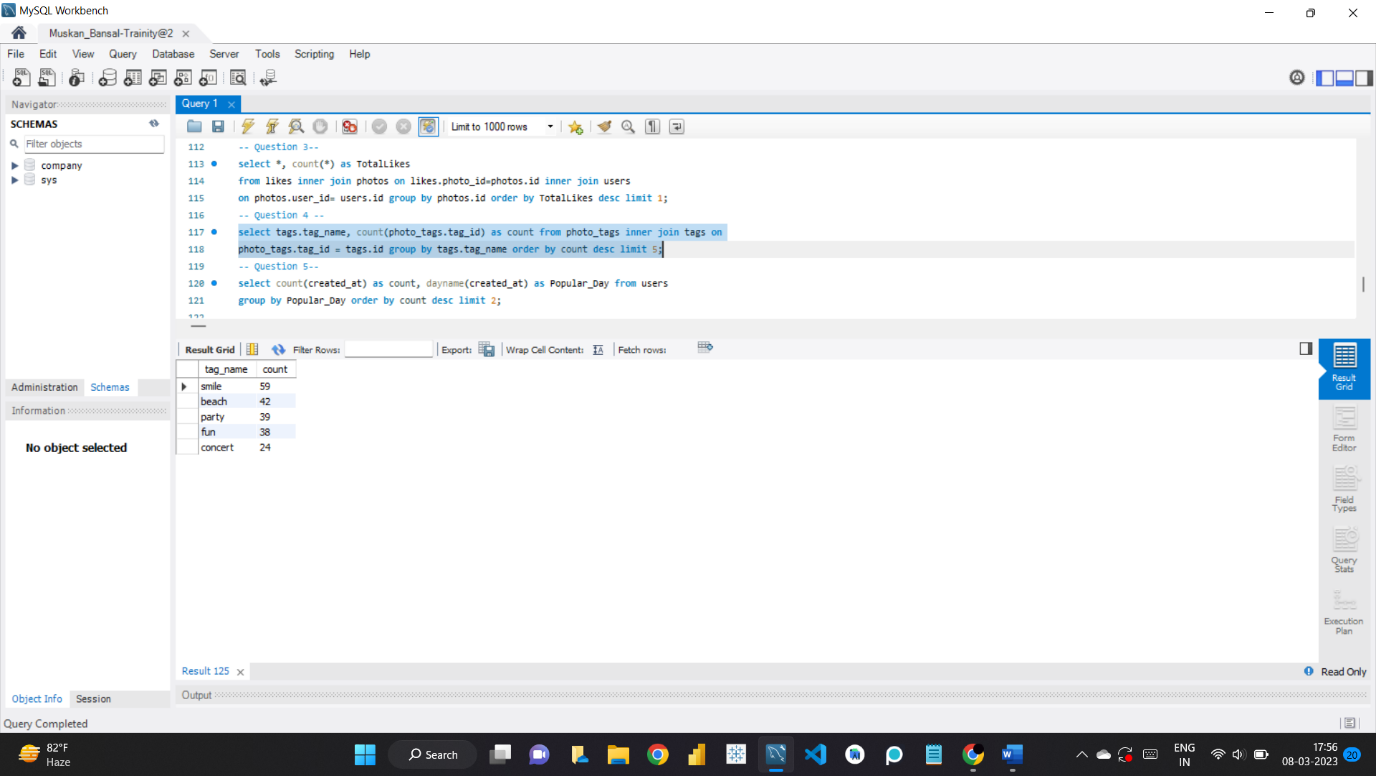
-- Question 4 --

select tags.tag\_name, count(photo\_tags.tag\_id) as count from photo\_tags inner join tags on

photo\_tags.tag\_id = tags.id group by tags.tag\_name order by count desc limit 5;

--In this question I needed to find the most used hashtags and for that I used count() function—

--then I ordered the tags from most used to least used and applied limit as I only needed the 5 most used hashtags—



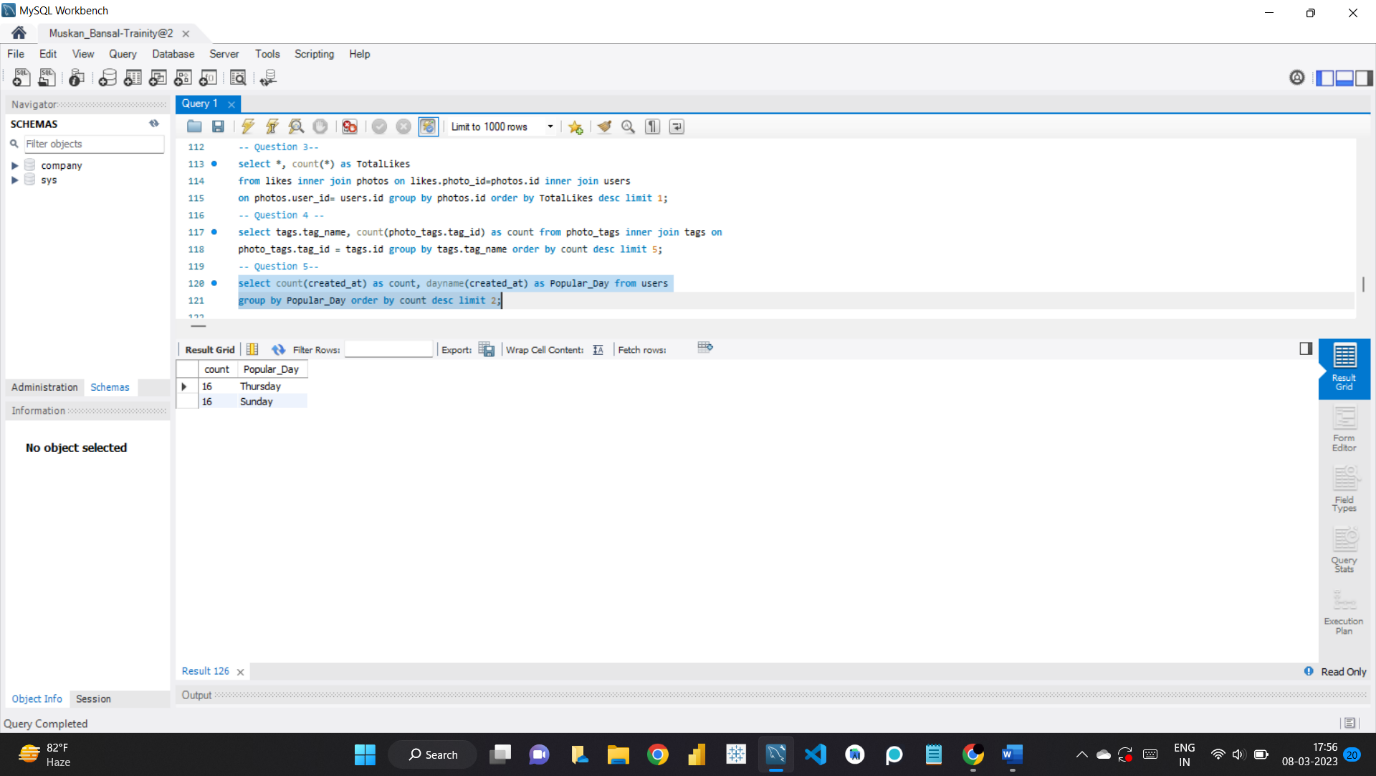
-- Question 5--

select count(created\_at) as count, dayname(created\_at) as Popular\_Day from users

group by Popular\_Day order by count desc limit 2;

--This question was straightforward. I used ‘dayname()’ function to find the days where most users registers—

--counted the days and arranged in descending order and then put a limit of 2 as I only needed to find the two most popular days where users have registered the most--



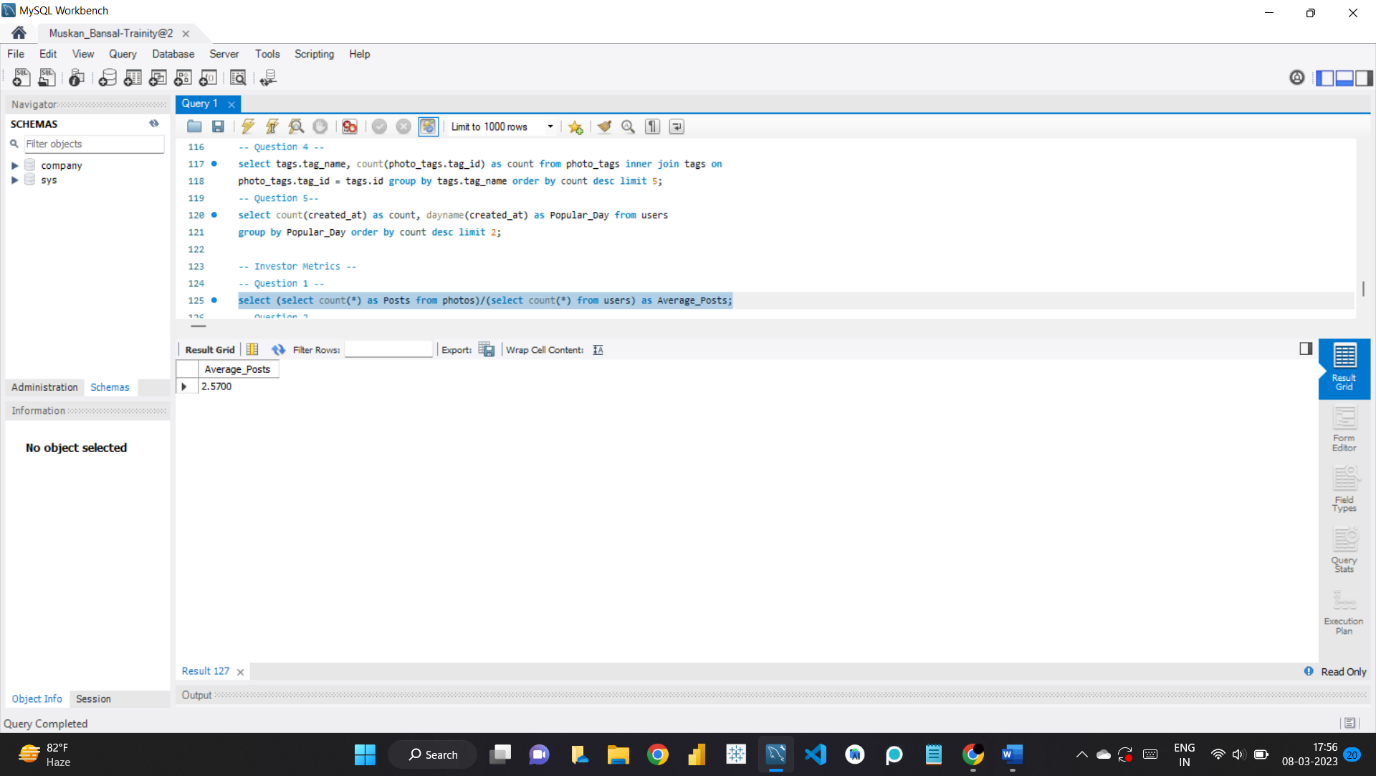
**-- Investor Metrics --**

-- Question 1 --

select (select count(\*) as Posts from photos)/(select count(\*) from users) as Average\_Posts;

--agerage number of posts per user—

--I used alias using ‘as’ and give the alias name ‘Average\_Posts’—



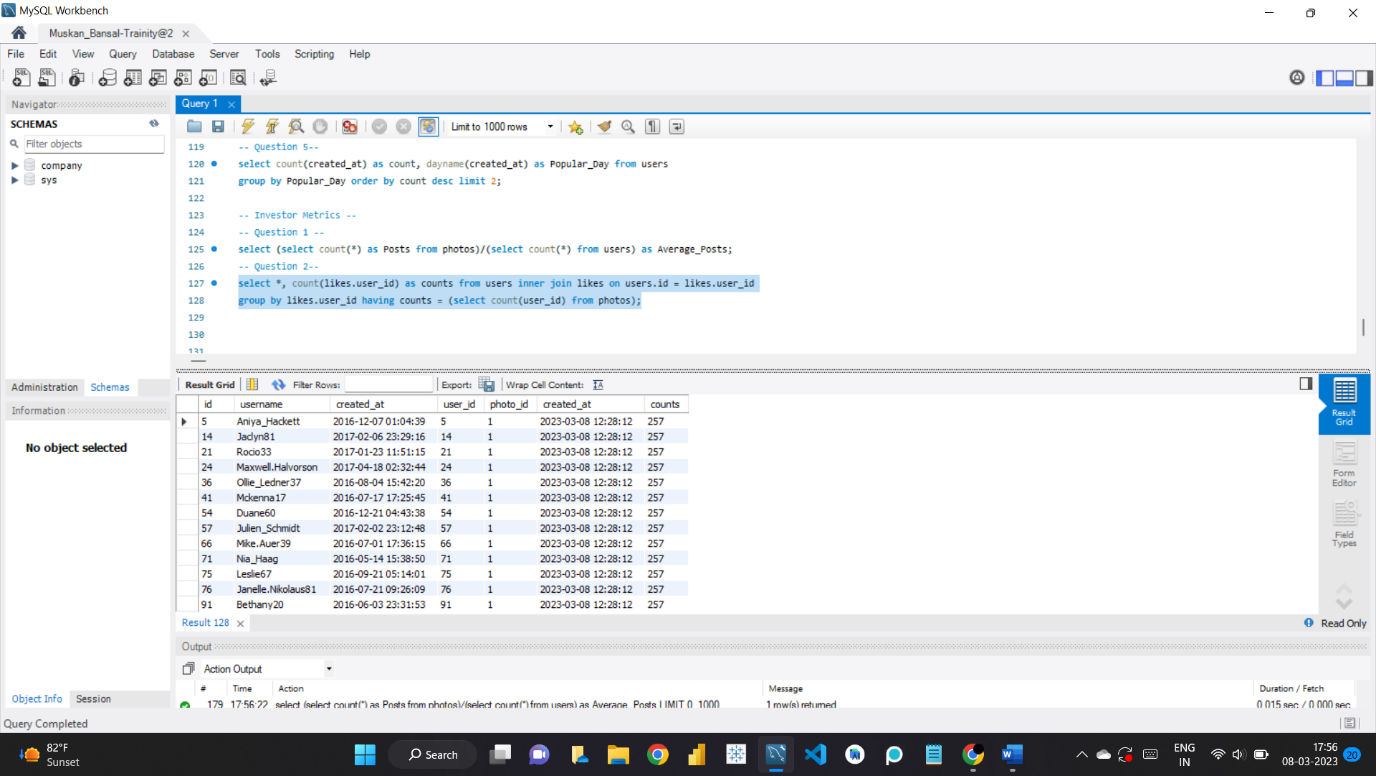
-- Question 2--

select \*, count(likes.user\_id) as counts from users inner join likes on users.id = likes.user\_id

group by likes.user\_id having counts = (select count(user\_id) from photos);

--here, I extracted all the data of those accounts that have liked all the posts, meaning those are bots—

--I used inner join as I only needed data of those accounts who have liked some posts—



**INSIGHTS:**

1. I learned about SQL Workbench.
2. I got to know about functions like ‘dayname’.
3. I understood how company has to deal with multiple queries to optimise their profits the most.
4. I learned why Marketing is so important and not just fenced to spreading information about a certain product/services.
5. I understood why it is important to reward users and why a company sends mails to uses and how they get the information regarding who to send.
6. I learned how company decides when to launch an AS Campaign or why hashtags could be an important information.
7. I also learned, why user engagement is important and how company finds the data of users using certain queries, they can easily get the exact data they want out of large amount of data within seconds which is not possible humanly.

**RESULT:**

Solving queries and completing this project gave me confidence. I learned about new functions and how to troubleshoot if and when port is not connecting. I learned the basic of SQL Workbench that how it has 3 major components and why it is used in so many places, it has wide variety of functions, can easily import/export data and a lot more. The three major components being:

**1. MySQL Connections**

**2. Models**

**3. Migration**

I solved all the queries on my own with the amount of dataset given, which I know was not large however it was my first time working on even this small amout of data. I am ecstatic that I opted for this course and looking forward for the future projects.

If there is any mistake or update or suggention, you can help me letting know for my future projects **via mail:** muskannbansall577@gmail.com