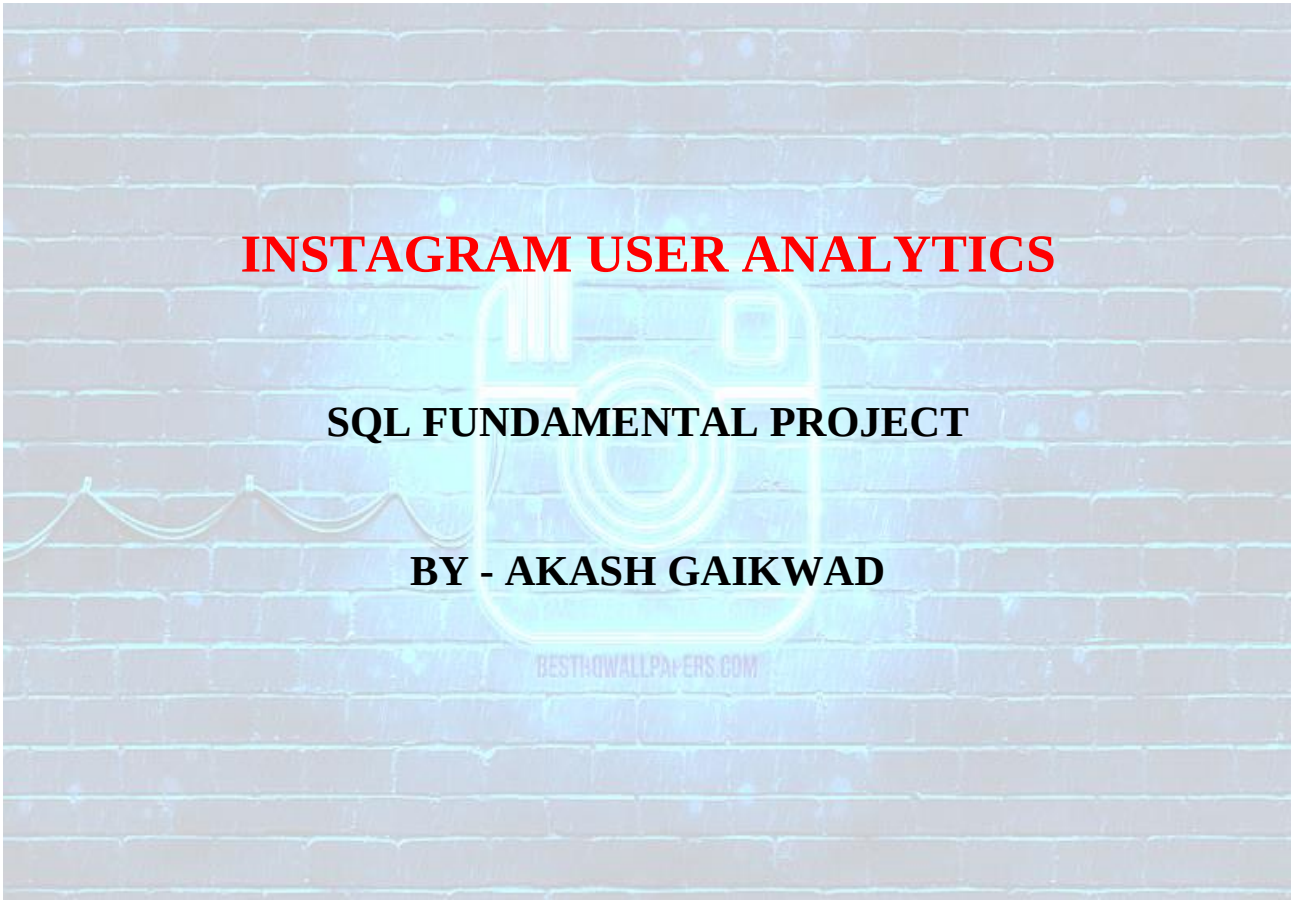


INSTAGRAM USER ANALYTICS

SQL FUNDAMENTAL PROJECT

BY - AKASH GAIKWAD

- ***Description:***

This project of Instagram User Analytics comprises analysis of user data from Instagram. User analysis is a process carried out to derive business insights which in turn helps in development, these insights are used by several teams across the company.

In this project I am working with a product team of Instagram and answer the questions asked by the management team using SQL.

I will be using SQL to derive at solution for the problem statements.

Problem statements:

A) Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

1. **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.
   Your Task: Find the 5 oldest users of the Instagram from the database provided
2. **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.
   Your Task: Find the users who have never posted a single photo on Instagram.
3. **Declaring Contest Winner**: The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
   Your Task: Identify the winner of the contest and provide their details to the team.
4. **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
   Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform.
5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch Ads.
   Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign.

B) Investor Metrics:

Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds.

1. **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts
   Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.
2. **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts.
   Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

I will be finding out answers for the given problem statements using a user database provided in the dashboard and MySQL installed in my personal computer.

Approach:

• The dataset from the dashboard is imported into SQL.

• The analysis is done by using basing SQL functions like SELECT, WHERE, COUNT, AS, FROM, GROUP BY, ORDER BY, etc.

• The displayed answers are added in this project.

• From the data output the question from the management departments are answered.

Tech stack used:

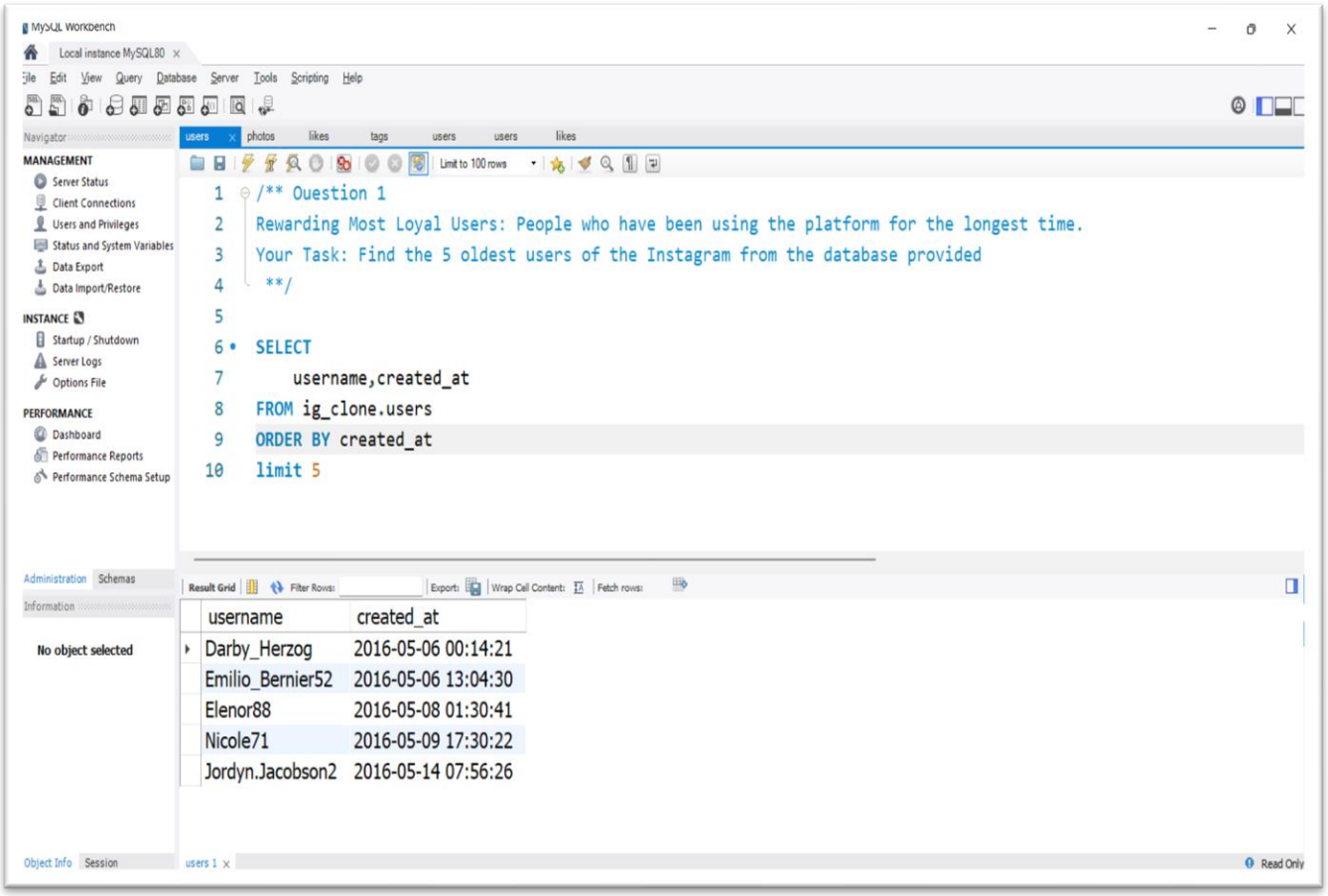
1. MySQL Workbench 8.0 compiled by Visual C++ build 1914, 64- bit.

Insights & Solutions:

A. Marketing team:

1) **<u>Rewarding Most Loyal Users:</u>** People who have been using the platform for the longest time.

   Your Task: Find the 5 oldest users of the Instagram from the database provided.



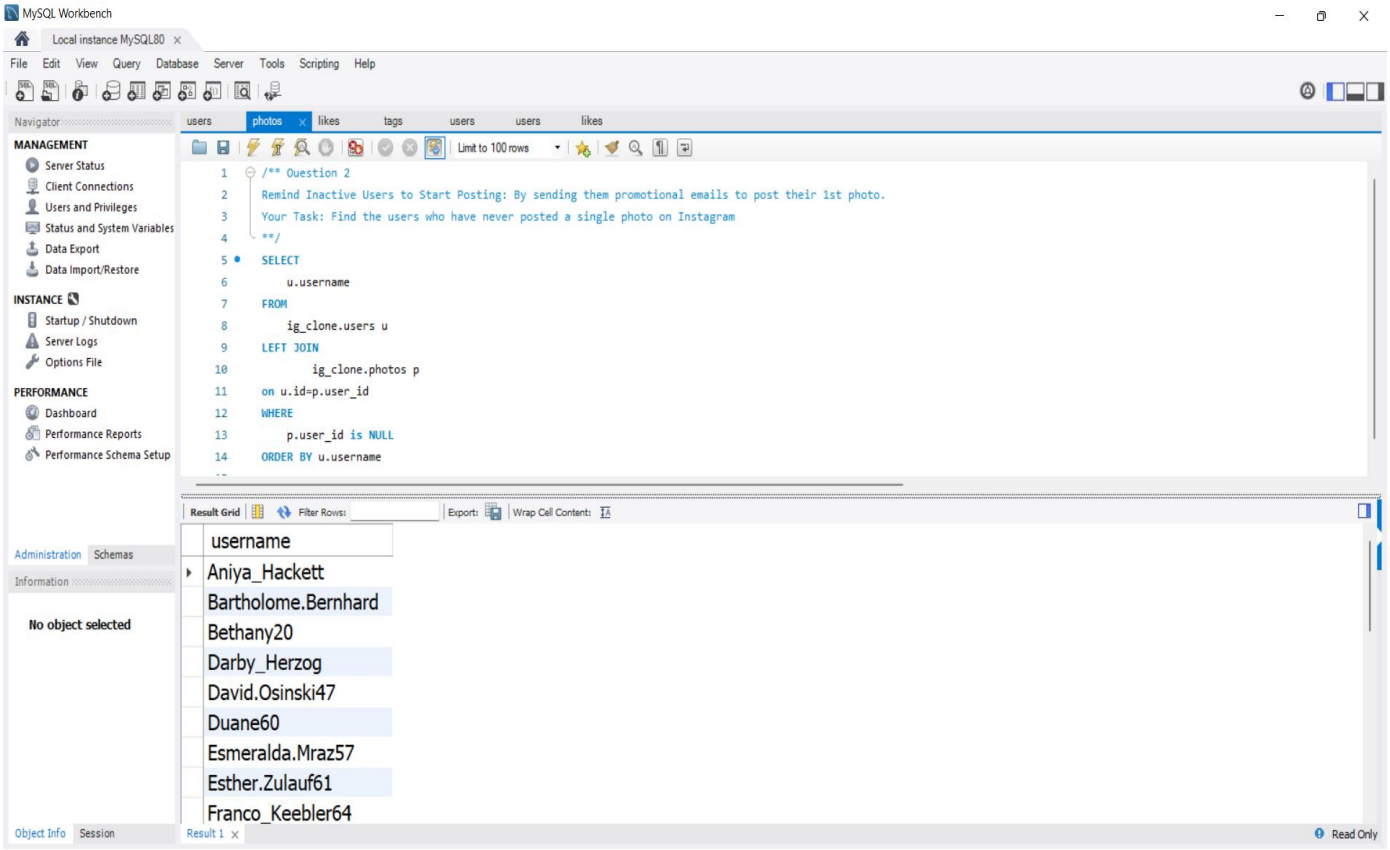
IMG.1

❖ Procedure:

To find the users using Instagram for the longest time, the created_at table from the dataset is used. The ORDER BY is used to find the oldest user, which sorts the table. This leaves us with the insights of oldest users.

2) **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.

Your Task: Find the users who have never posted a single photo on Instagram.



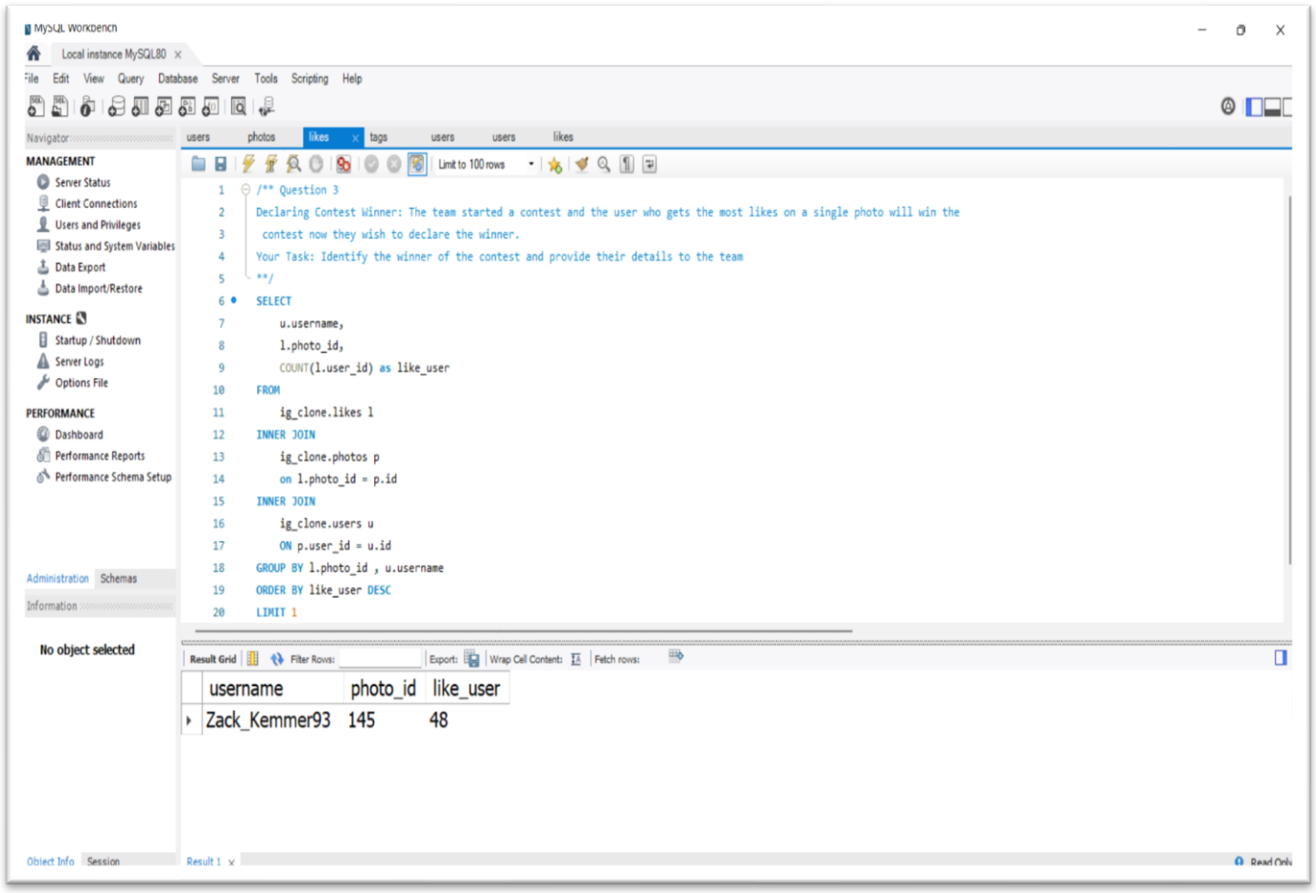
IMG.2

❖ Procedure:

To find inactive users we should find the users who haven't posted a single photo on Instagram, for that we consider users table which contains the user details and the photos table which contains the Instagram post details. By LEFT joining two tables and using WHERE clause the users who haven't posted can be identified. IS null helps us identify the users in users table but not in photos table. This leaves us with the insights of people who haven't posted and been inactive on Instagram.

3) **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.

Your Task: Identify the winner of the contest and provide their details to the team



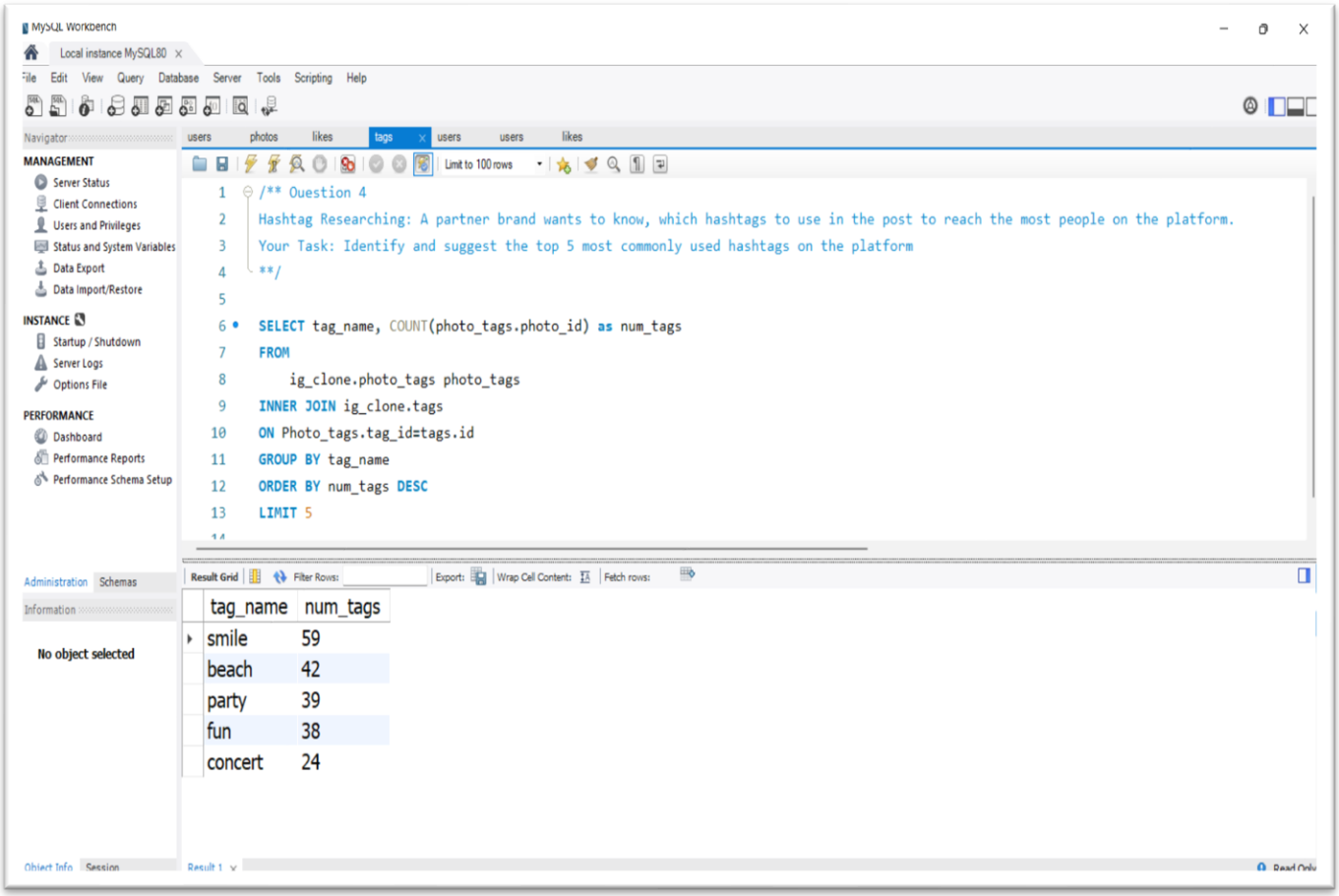
IMG.3

❖ **Procedure:**

To identify the winner, we should COUNT the likes of all the series. So, we join the INNER JOIN likes and photos table. As we use GROUP BY function on photo_id and username, the query groups all the essentials as per the arguments. We use ORDER BY to find the user who got highest likes and LIMIT function to just output one element from the table. By following all these procedures, we can conclude who is the winner of the contest.

4) **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform



IMG.4

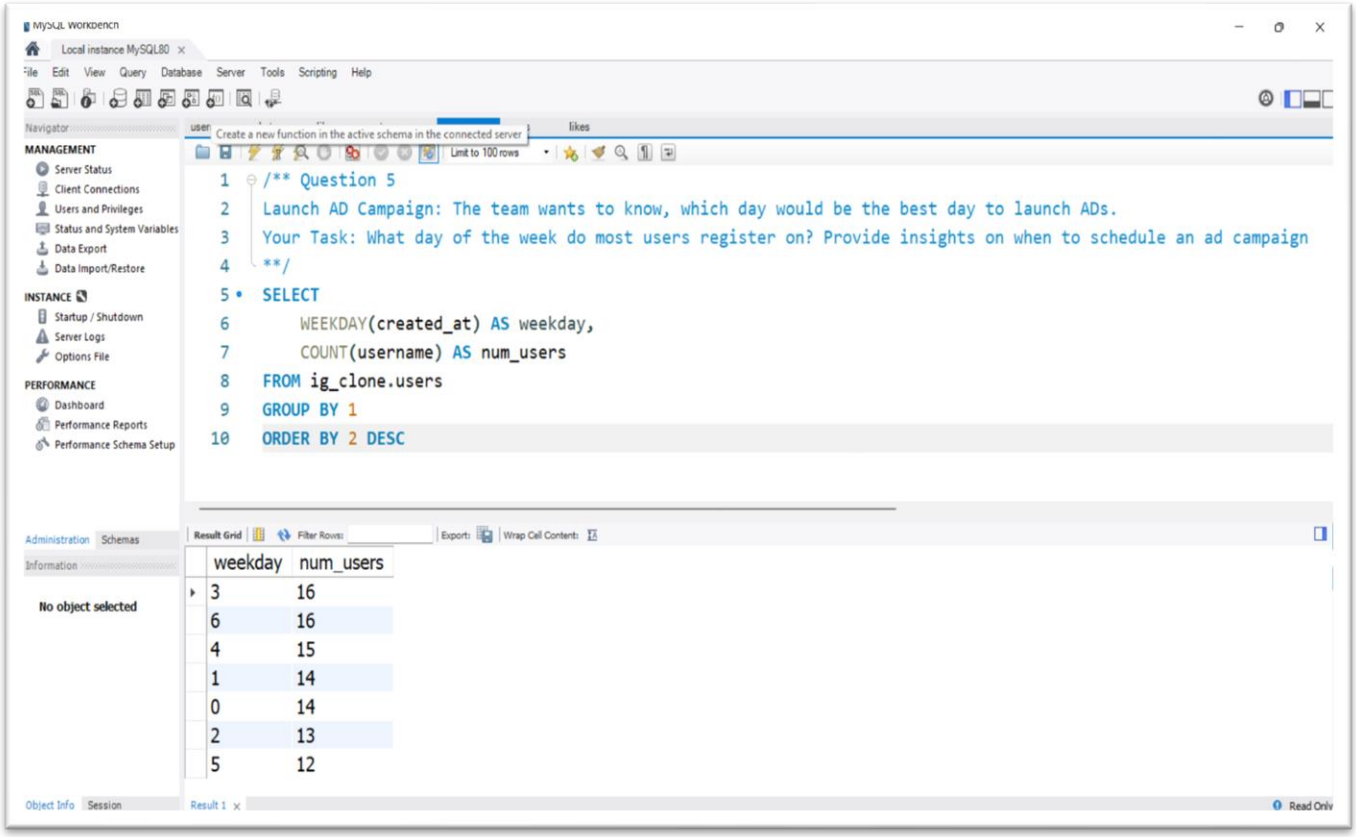
❖ Procedure:

To identify the most commonly used hashtags we need to count number of times the hashtag has been used. For that we need data from both photo_tags and tags table. We INNER JOIN both the tables and use COUNT function to count number of times the hashtag has been used. GROUP BY is used to separate and group the data based on the argument tag_name, which leaves us with a list of hashtags and number of times they've been used. We ORDER BY descending order and LIMIT with 5 to output the top 5 commonly used

Hashtags. By following this procedure, we can find the commonly used hashtags.

5) **Launch AD Campaign**: The team wants to know, which day would be the best day to launch ADs.

   Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign.



IMG.5

❖ **Procedure:**

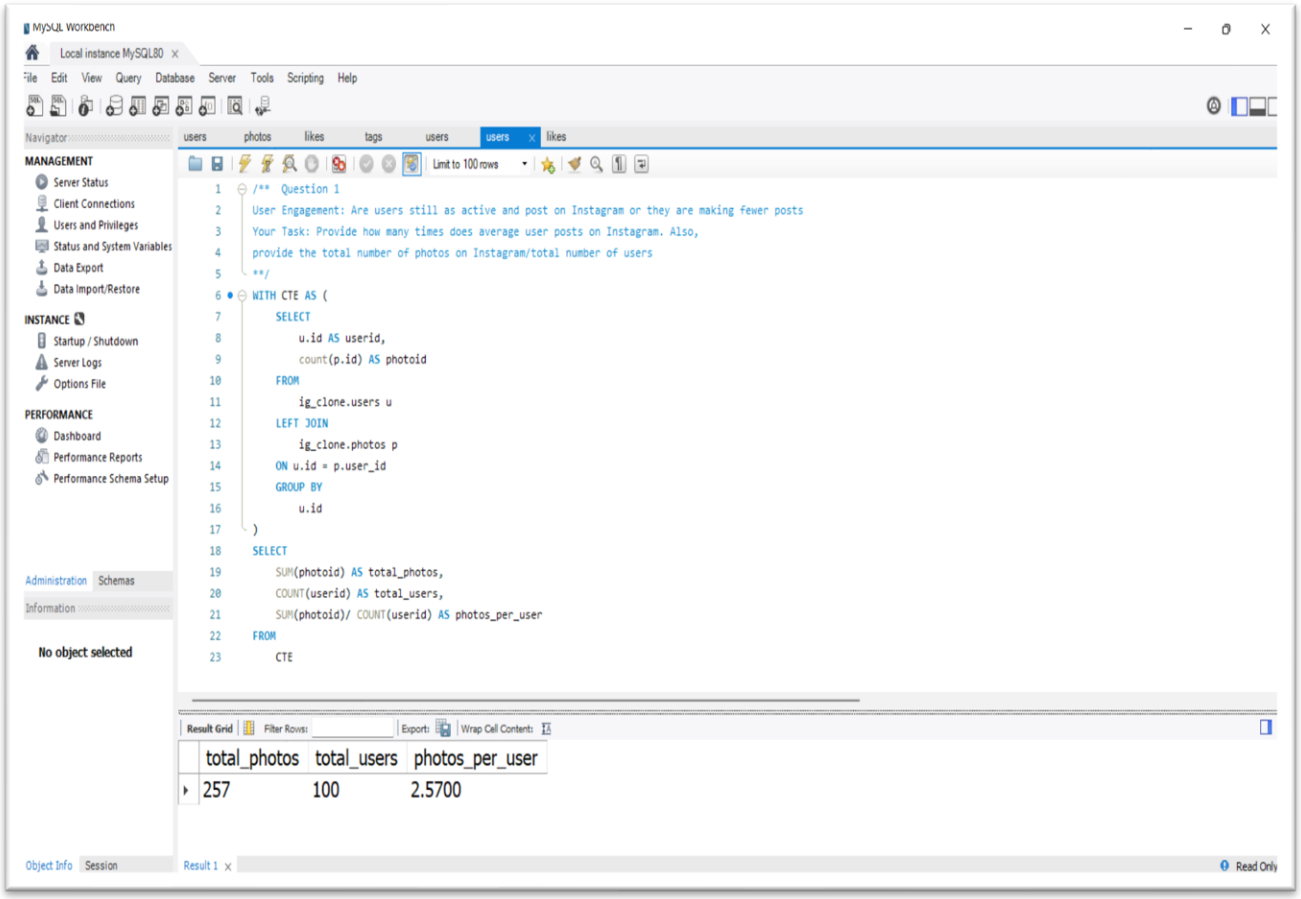
We need to find in which day of the week the users register most on. To find that we need to get data from users table and I'll be using WEEKDAY in MySQL. By using WEEKDAY functions on created_at table which has the data we can get the day on which the user has registered and COUNT to find the number of users registered on that particular day. We use GROUP BY on weekday and get the grouped output and use ORDER BY to sort it in descending order.By following the procedure, we can find the perfect day to launch the AD campaign. But in my views instead of finding the users register

date we can find on which does users post more photos, which in turn means more activity. This would give us great result

B) Investor Metrics:

1) **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts.

   Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.



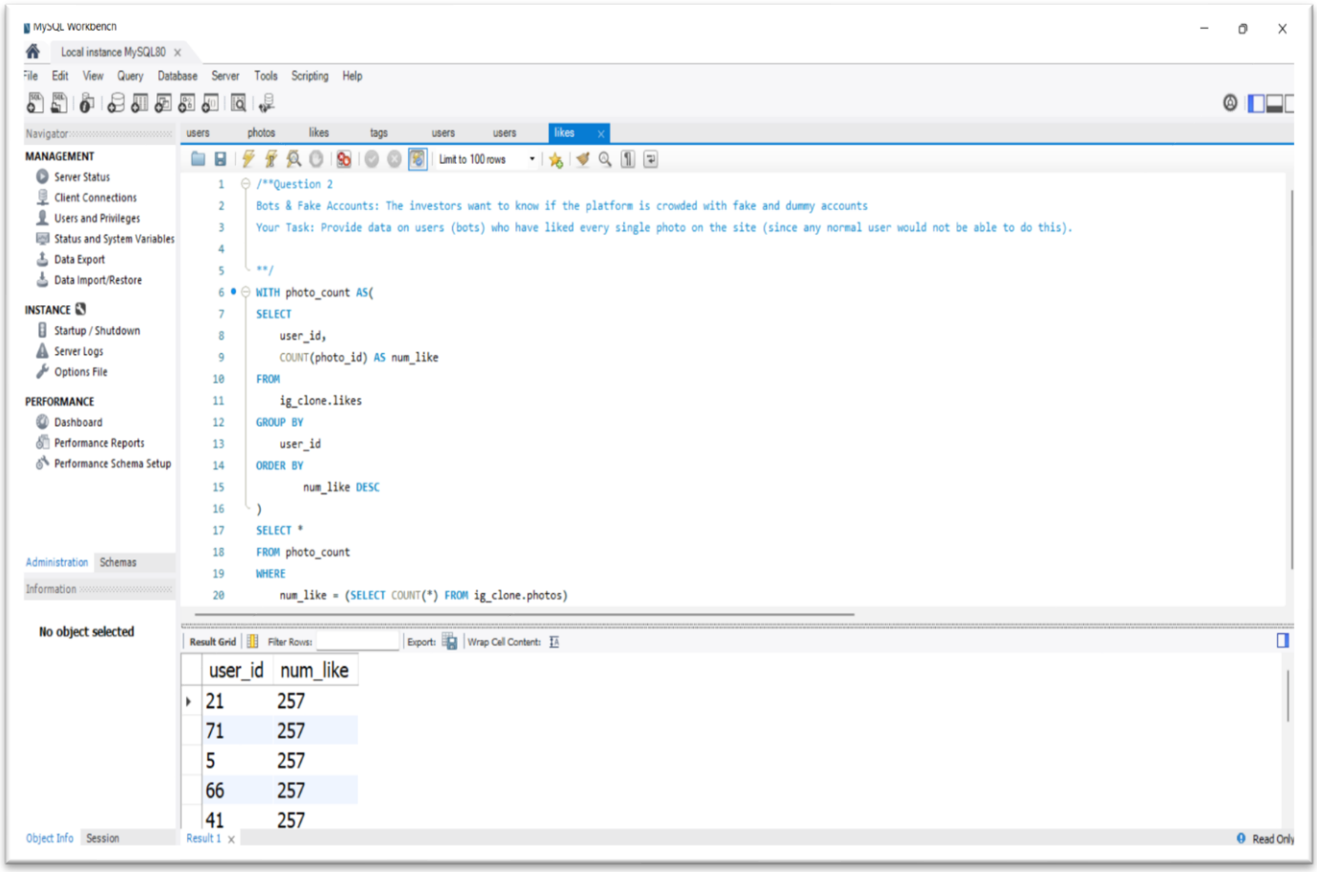
IMG.6

❖ Procedure:

In this section we will be getting multiple insights, the first one is getting the post per user. We will be counting the photo_id to get number of times the user posts and group it to 12 get individual results. We use CTE as temporary result set and from there we find the total number of photos and users using aggregate functions like SUM. The given problem statements are answered accordingly.

2) **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts

   Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).



IMG.7

- **Procedure:**

  To find the fake account we can use the mentioned method to find the accounts which likes every photo, which is an un usual activity mostly carried out fake bots. For that we use the Likes table from the dataset. We'll be counting all the likes and group it based on the user_id to get output. We order by descending order to find the users who liked all the photos.

  By following this procedure, we got a list of accounts which we consider as fake accounts. We can also find user accounts which comments on every post, because it is also a suspicious activity. This may provide us with more efficient results.

- ***Insights & Knowledge gained***
  - The dataset provide in the dashboard is in query format and which is suitable for MySQL
  - I did some searching and successfully modified the query and created a database in MySQL Workbench.
  - This has helped me to know more about creating and handling databases because I am a fresher in this field.
  - I have gained more hands-on experience in this project because, I have handled a real life like situation.
  - It helped me to understand the GROUP BY function well, as I was a little confused about that when I started the project.
  - I have gained knowledge about the aggregate functions in SQL like COUNT, SUM, etc.
  - We can provide more accurate answers for problem statement 5, by also checking the users who comment on all the posts. Some of the fake bots are programmed in such a way to comment but not Like the post. By doing this we may even find accounts like that and eliminate them.

- ***Result***

  By completing this project, I have come to know about the importance of Structured Query Language (SQL) in data analysis. We can use SQL in relational databases and provide with insights which makes our life easier. Instagram user analytics is like a real-life project we get in our work, this project has helped me to understand how the database creation and analysis works.

  I have successfully answered and attached the outputs of each and every problem statement I have worked on MySQL Workbench.