**INSTAGRAM USER ANALYTICS PROJECT**

**Description:**

Let’s go through Instagram history briefly. **Kevin Systrom and Mike Krigger** invented an web application called ‘**Burbn’ in 2010** with 25000 users on it’s first day. Social media platform for entertainment and hobbies purpose. Intention behind is to that to **describe moments with small caption**.

Trending now-a-days in youths specially, that they are fascinated towards clicking photos, arts, fashion, hobbies, so on. Its not only for **entertainment** purpose but also for way to earn money such as growing business by advertisement, making connections globally.

Our project is about researching activity of Instagram users. When user posts some picture, videos, he will get reaction over that like comment, likes from his followers. In this project will find out how many users are there in our data, comments on their pictures posted, so on.

**Approach:**

consists of users activity till today, his/her every moment captures or important to study.

There is one point to clear, may be two types of users are there, A is interested in entertainment things called socio-holic and other B is having less interest in social networking. So A will be using Instagram daily or most of the time, rather B will be using just for small use, once in week.

To analyze the Instagram and users behavior A type of people needed most so we can make it clearly. In a given dataset, data is mixed in nature, it contains people from different perspectives, they are using this app for different purposes. Users containing active, inactive, fake users.

**Tech-Stack used :**

Technology stack is used : MYSQL version is 8.0

So given data is in tabular form, structured data that is relational database, to make operations or to solve the queries we generally use SQL language. MySQL is a software used to deal with queries. There are many tools where we can solve problems like MS SQL , oracle. But MySQL is open source platform, easy to handle.

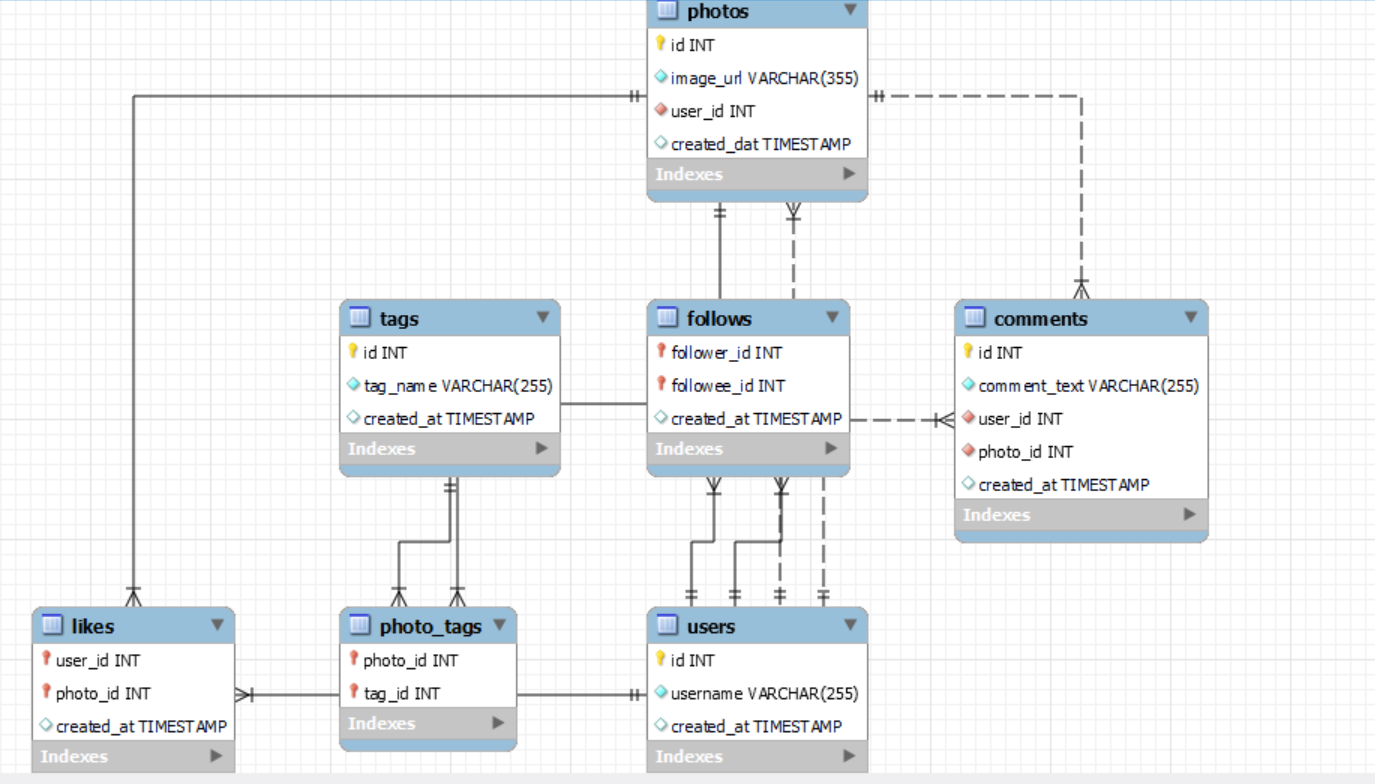
**Insights:** Analysis

Our main purpose is to analyze the given data , we have gathered from Instagram user. It consists of users activity till today, his/her every moment captures or important to study.

It tells about their insights indirectly. How the user posted pictures, videos, making comments, liked or reacts other user’s posts, so on all this information will be analyzed neatly to gain insights more about it.

Suppose we have started analyzing the data, its starts from how many **users are actively uses Instagram daily, weekly, yearly? Have they posted anything on their account? Friend followers and following details, comments we have made on some others pictures, all these data is we have in our dataset**.

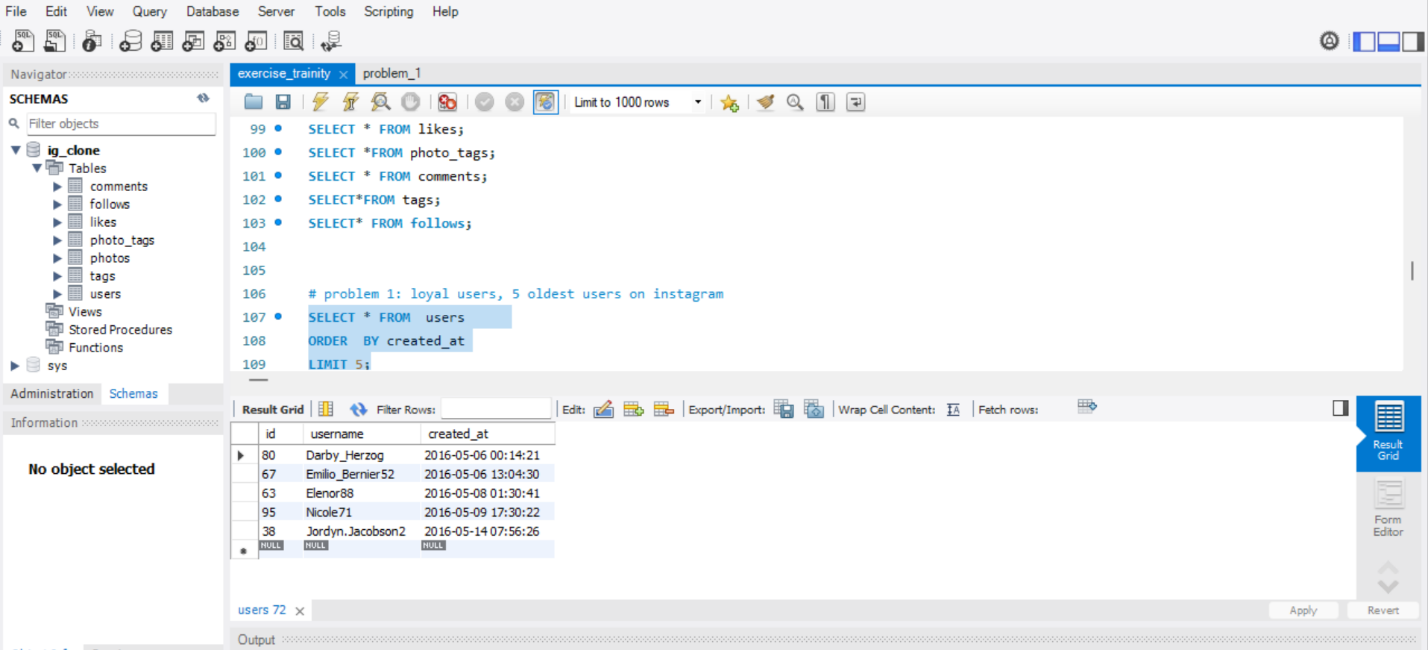
**Schema of Insta\_ig** database, created on MySQL 8.0 version( mysql workbench 🡪 reverse engineer).



Schema nothing but the layout structure of our database , to visualize the relationships between them clearly. You can see all this are connected together with one of the column.

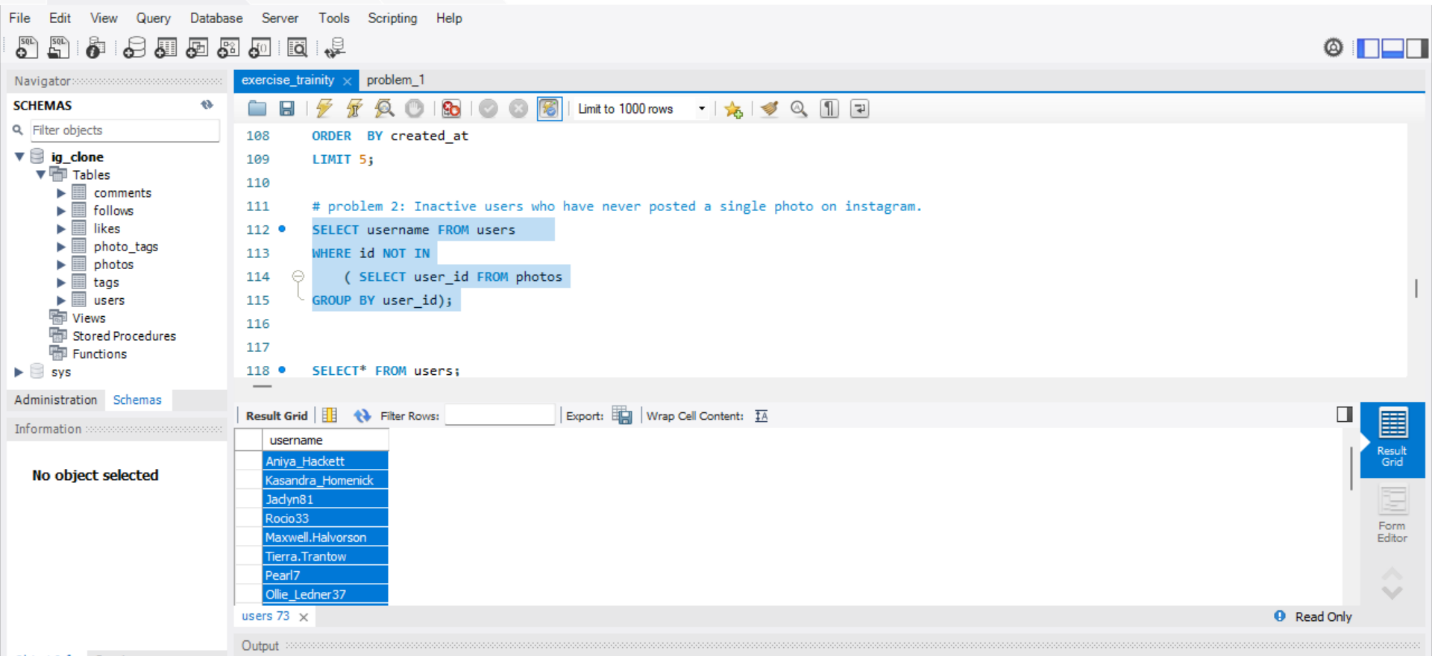
Lets make real time analysis using MYSQL software tool, to understand more visually and quickly. I have some questions raised are as follows:

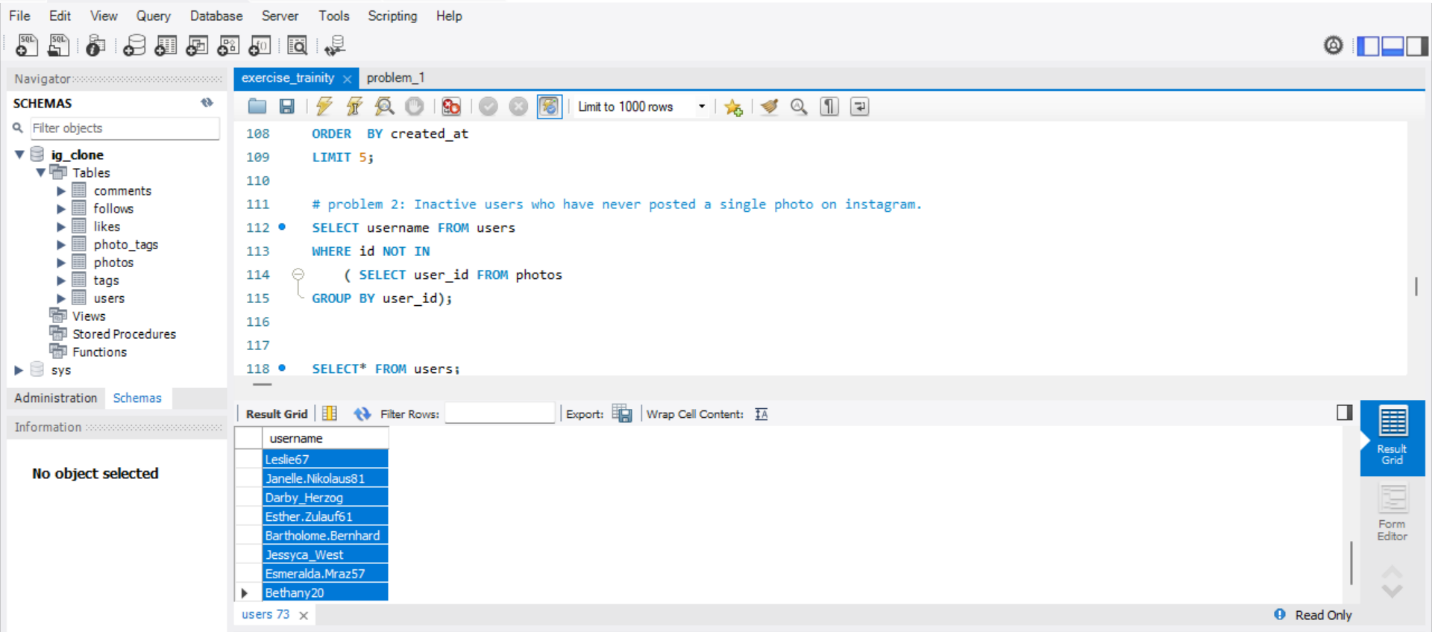
1. In our dataset, will find out **top 5 oldest users on Instagram**, who have created their account early. They will be called as most loyal users, so we have to reward them.



Okay, 1st see there are **total 100 various users** in our dataset. Using dates in **created\_at** column, will get that users who have created account very early compared to others.

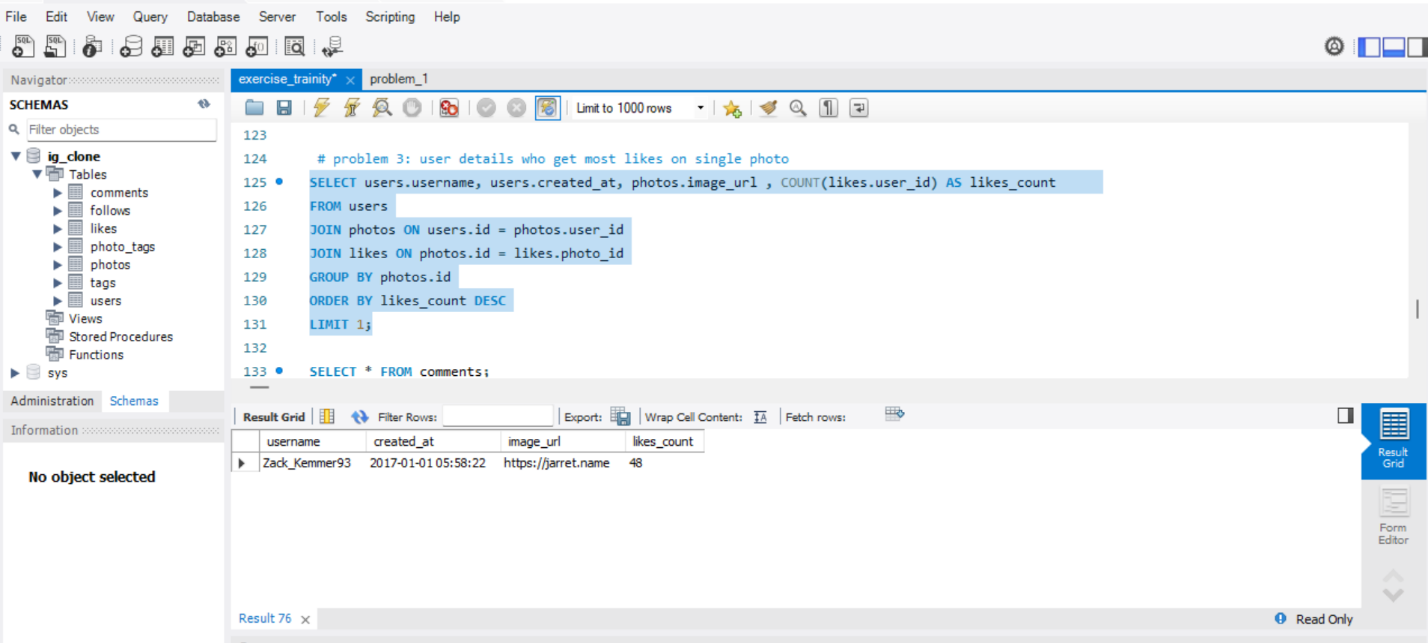
1. Now its time for **inactive users**, who never posted a single photo on Instagram.





So here you can see we have used two tables users and photos. I solved this query using **subqueries** to make this easier**. Out of 100, inactive users are 12** who haven’t posted a picture.

1. We have to declare **contest winner by analyzing who got most likes on single photo.** Also, have to provide the users details.

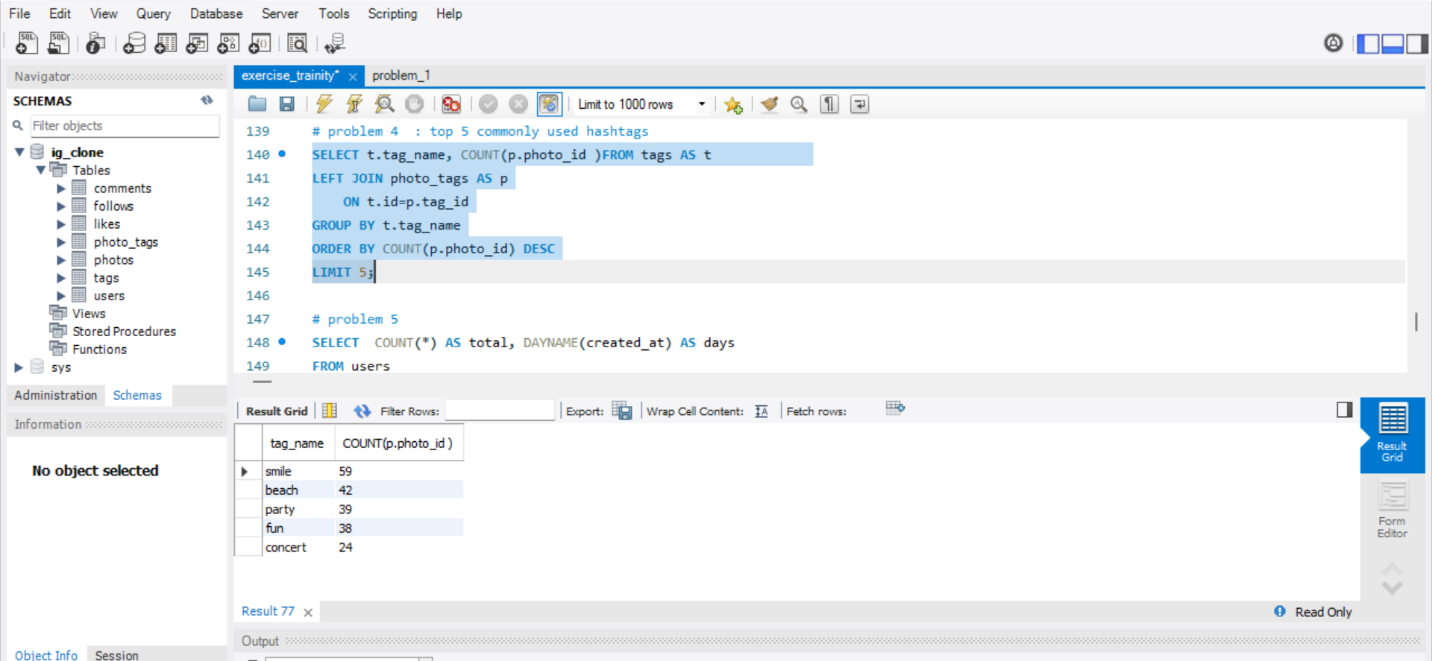


Winner is Zack\_Kemmer93 having 48 likes on his photo, followed by 'Malinda\_Streich' and Adelle96 both having 43 likes as runner-up.

How we can find winner, we want **likes** table along with **photos**, and also users.

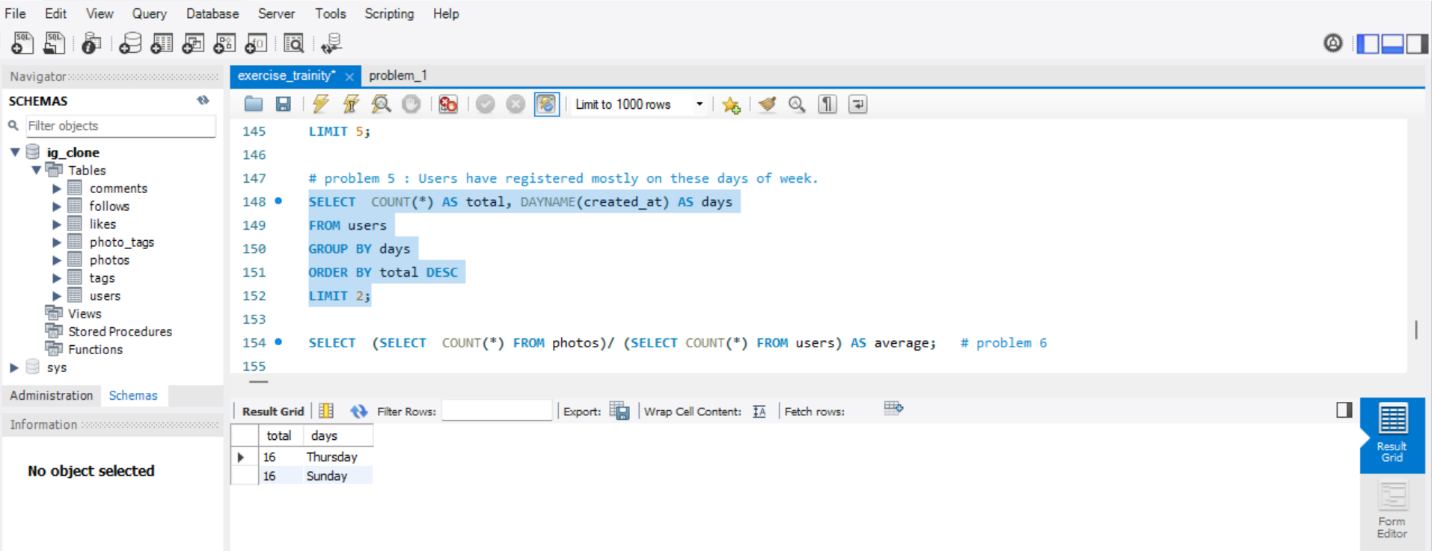
Twist is that we want single photo of user, but users can upload more than one photo, that’s why we join this tables as above.

1. **Hashtag researching**: need to find which **hashtags are used mostly in posts on Instagram.**



Here u cane see, **smile** hashtag have used by most people in their posts. Followed by beach, party, fun, concert etc. For this **photo\_tags** as well as **tags** tables are used.

1. **Launch AD campaign**: Have to find which **day is best day in week to launch ads**, it means on which days the users are active most of the time.



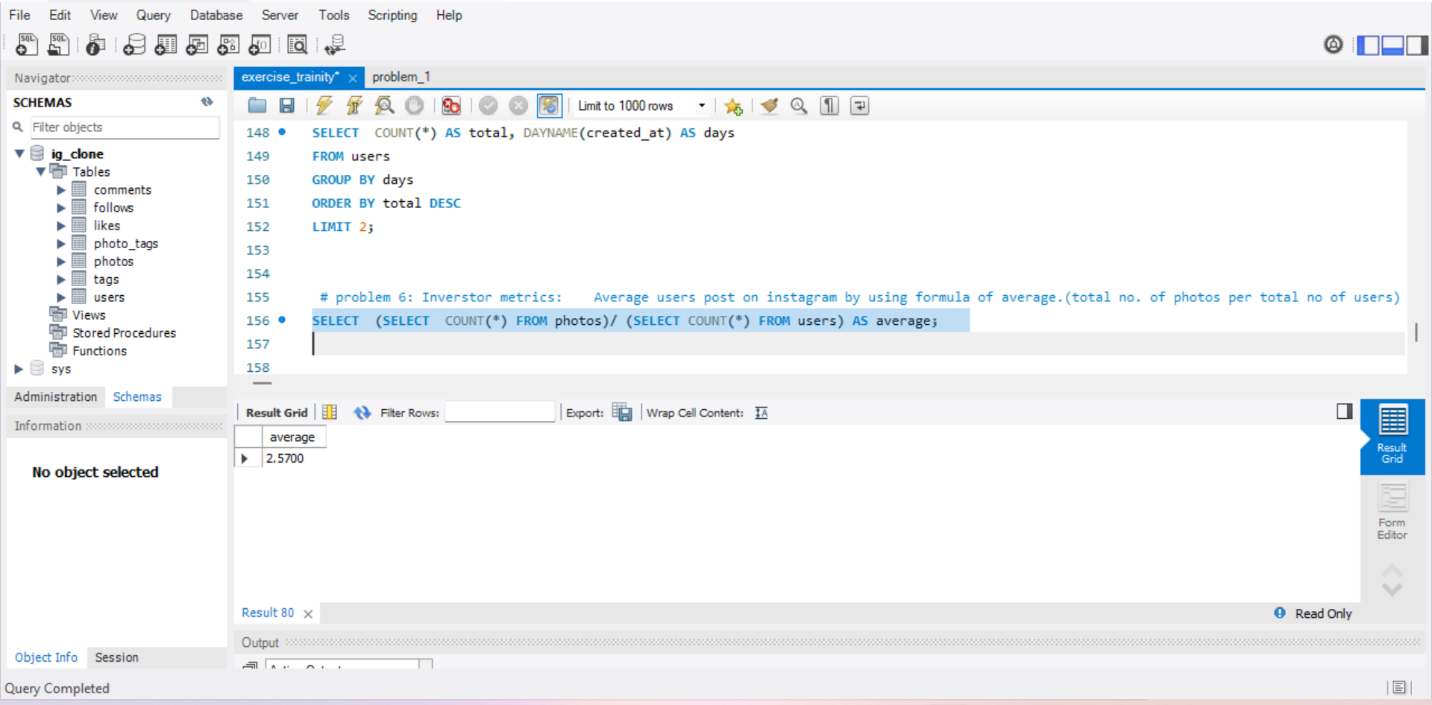
As **the created\_at column contains date , so we converted date to day by DAYNAME.**

On Thursday and Sunday, users have registered or logged in mostly like **highest 16 users** out of 100, followed by **Friday about 15 users,** and so on.

In accordance with **Investor metrics**, we have 2 question:

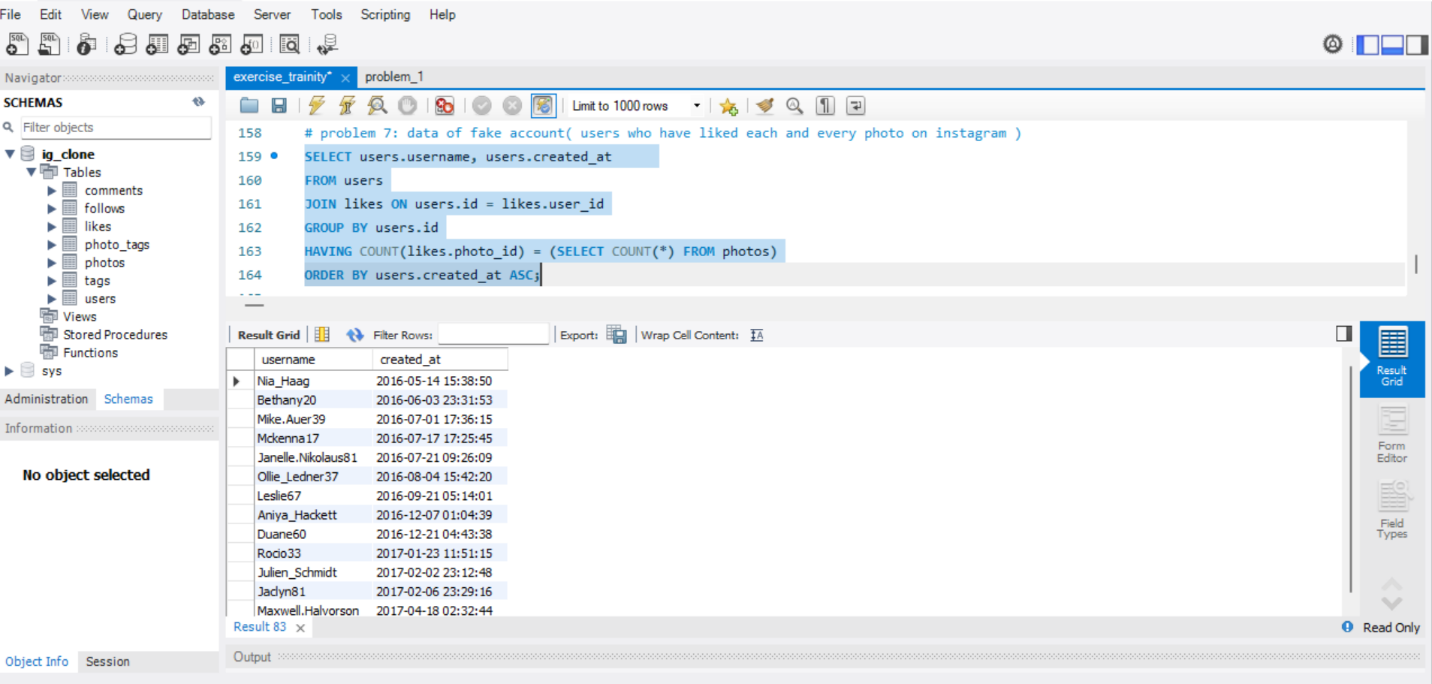
1. For user **engagement activity**, are users active or not? Are they actively or frequently posted or logged in to account or not?

From the data will find average rate of user posts on Instagram.



Average is nothing but total number of photos per total number of users. As total users are 100 and total pictures posted overall is 257, avg comes 2.57

1. Related to **fake accounts**: User who have liked every single photo onsite will give us dummy and fake users.



Here total fake accounts are 12 out of 100 users.

**Result:**

In this way we have analyze the given data using sql language. Purpose to find and asks as many questions to our data, so that it will give all hidden answers. We have captured every detail moment of each user to gain more insight.