Instagram User Analytics

Project Description:

In order to help the Instagram product team and the product manager answer the questions posed by the management team, it is my project to find out the analytics and provide insights of the given dataset to the management team. MySQL Workbench was used to complete this project. For this project, I will need to find the 5 oldest users to find the users who have

never posted a single photo on Instagram, the contest winner their details, the top 5

most commonly used hashtags on the platform, the day of the week do most users register on, average user posts on Instagram along with the total number of photos on Instagram/total number of users and finally, users (bots) who have liked every single photo.

Approach:

To approach this project, I began by fully comprehending the requirements and objectives of this project. Following this, I began executing the project with an adaptable mindset, exploring diverse approaches to accomplish the task. After evaluating various options, I selected the most suitable approach to continue with the project. I then proceeded to write and implement the queries to produce the correct, comprehensible output. Overall, I am confident with the project’s outcome and consider it a success.

Tech-Stack Used:

MySQL Workbench with the latest released Version 8.0 Community Edition.

I used it because MySQL Workbench is a visual database design tool that combines SQL development, administration, database design, creation, and maintenance into a single development environment.

Insights:

I surmise that the dataset contains a lot of data, and that it is necessary to analyse it before continuing. I also infer that the data is distributed in a specific way, and that understanding the distribution is required before proceeding.

**A. Marketing:**

1. **Rewarding Most Loyal Users:**

Here, I need to find the 5 oldest users of the Instagram from the given database.

For fulfilling the required query, I selected id, username, created\_at column from the users table and order by created\_at column in ascending order with limit of 5

**SQL Statement :**

SELECT

id, username, created\_at

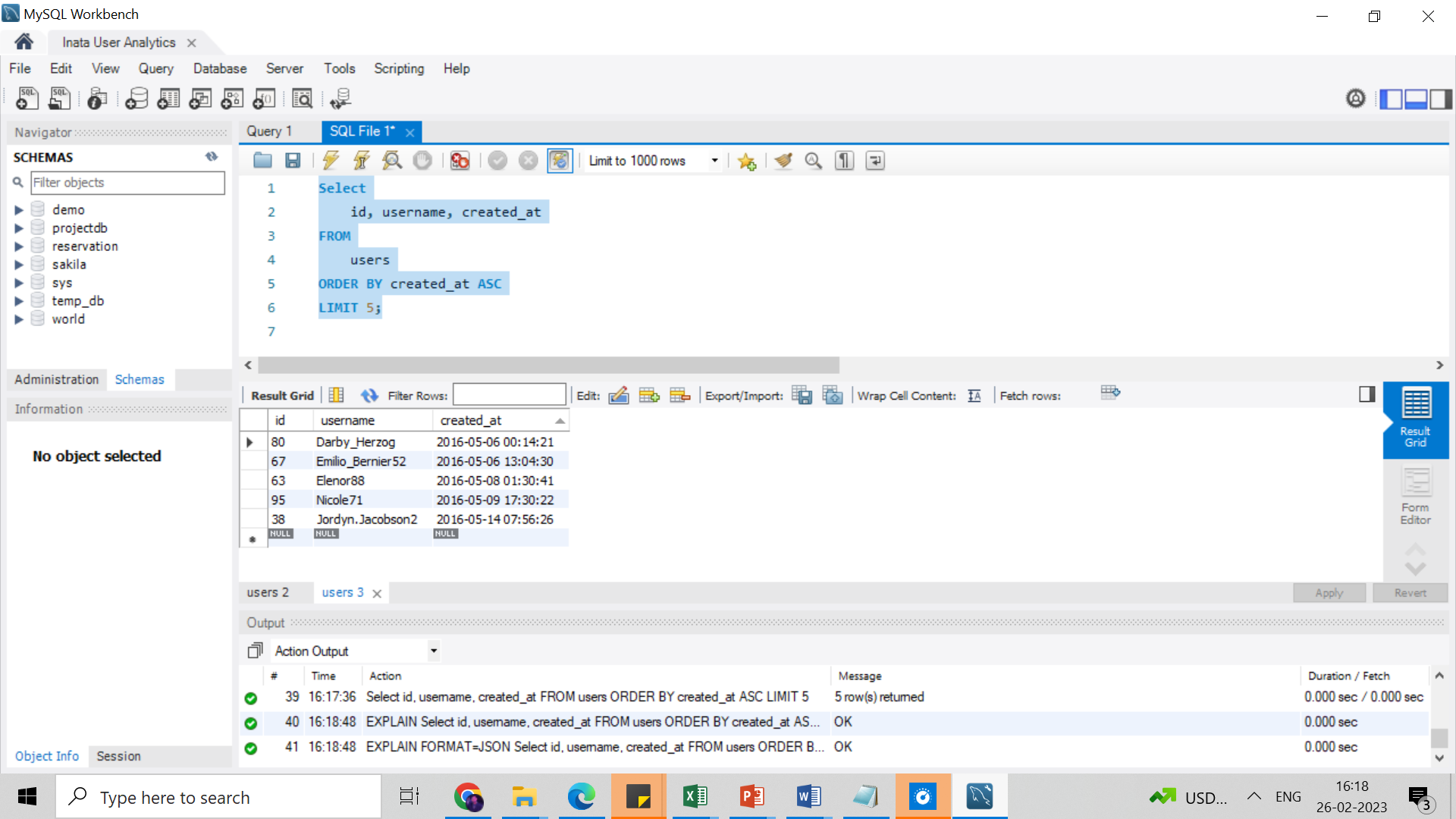
FROM

users

ORDER BY created\_at ASC

LIMIT 5;

**Result:**



**Conclusion:**

Users having IDs 80, 67, 63, 95, and 38 are the oldest users on the platform.

1. **Remind Inactive Users to Start Posting:**

|  |  |
| --- | --- |
| 1 | username |
| 2 | Aniya\_Hackett |
| 3 | Kasandra\_Homenick |
| 4 | Jaclyn81 |
| 5 | Rocio33 |
| 6 | Maxwell.Halvorson |
| 7 | Tierra.Trantow |
| 8 | Pearl7 |
| 9 | Ollie\_Ledner37 |
| 10 | Mckenna17 |
| 11 | David.Osinski47 |
| 12 | Morgan.Kassulke |
| 13 | Linnea59 |
| 14 | Duane60 |
| 15 | Julien\_Schmidt |
| 16 | Mike.Auer39 |
| 17 | Franco\_Keebler64 |
| 18 | Nia\_Haag |
| 19 | Hulda.Macejkovic |
| 20 | Leslie67 |
| 21 | Janelle.Nikolaus81 |
| 22 | Darby\_Herzog |
| 23 | Esther.Zulauf61 |
| 24 | Bartholome.Bernhard |
| 25 | Jessyca\_West |
| 26 | Esmeralda.Mraz57 |
| 27 | Bethany20 |

Here, I need to find the users who have never posted a single photo on Instagram.

For fulfilling the required query, I selected username from users table, then I used left join function to join users table with photos table on user.id = photos.user\_id, then I used where function with condition photos.id is null.

**SQL Statement: Result – User List**

SELECT

username

FROM

users

LEFT JOIN photos

ON users.id = photos.user\_id

WHERE photos.id IS NULL;

**Conclusion:**

This is list of users(usernames) who never posted a single photo on Instagram is given in the user list below. After seeing this data, we can conclude that a total of 26 people present in the database never posted any photos on the platform.

1. **Declaring Contest Winner:**

Here, I need to identify the contest winner and provide their details to the team.

For that, I have selected (username, id) from users, count(\*) as most\_liked from likes

table the I use inner join function to join likes table with users table on users ON

users.id=likes.user\_id then I use group by on photo\_id and then I use order by function

on most\_liked in descending order with limit 1.

**SQL Statement:**

SELECT

users.username, users.id, COUNT(\*) AS most\_liked

FROM

likes

INNER JOIN

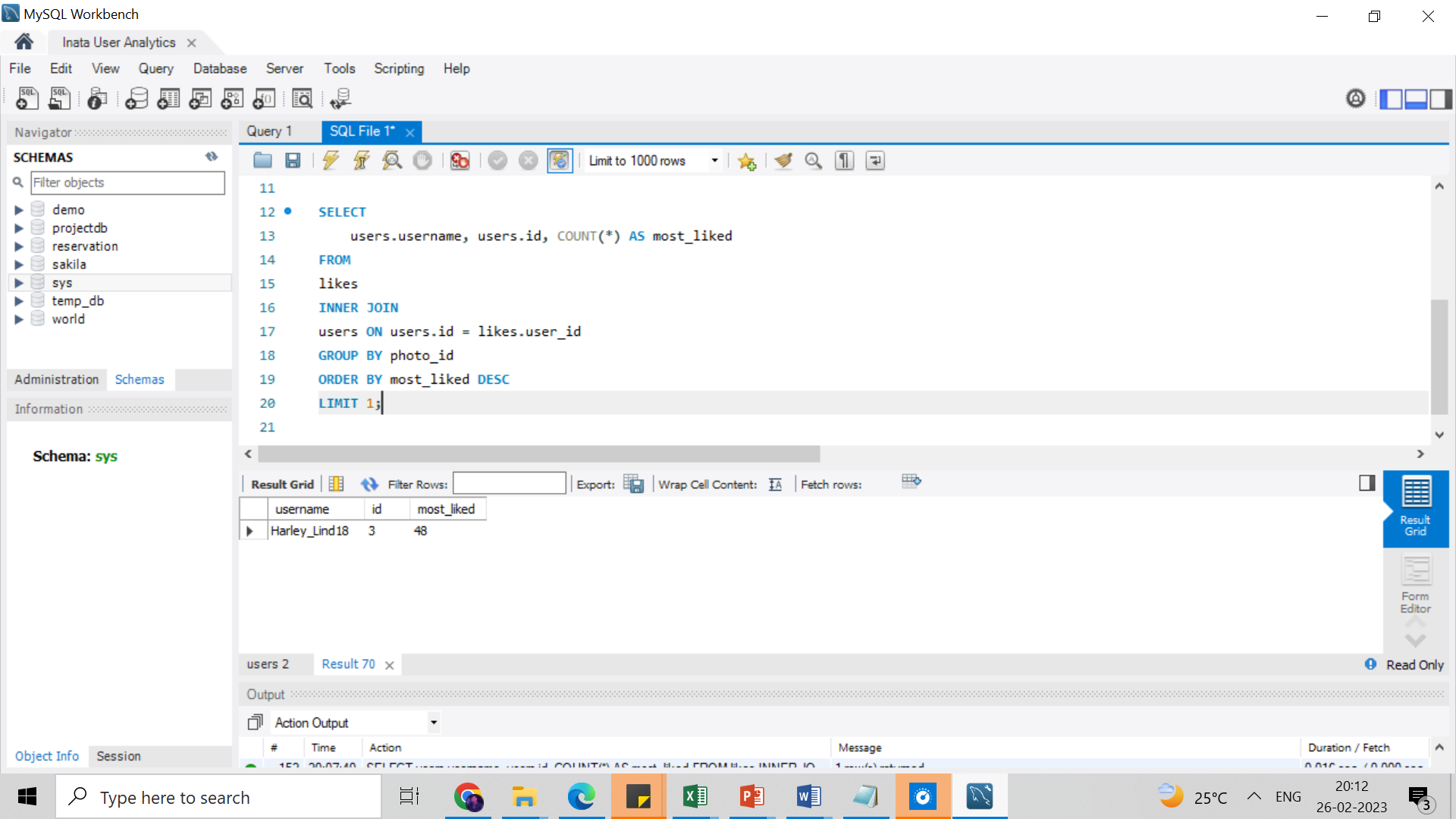
users ON users.id = likes.user\_id

GROUP BY photo\_id

ORDER BY most\_liked DESC

LIMIT 1;

**Result:**



**Conclusion:**

|  |  |  |
| --- | --- | --- |
|  |  |  |

We learned that photo id 3, which gets the most likes of 48, belongs to the user‘Harley\_Lind18’. So, Harley\_Lind18 is the winner of the contest.

1. Hashtag Resercher:

Here, I need to Identify and suggest the platform's top 5 most commonly used hashtags.

For that, I have selected tag\_id, count (photo\_id) as total\_tags from photo tags table, and then I used join function to join tags table with photo tags table on

tags.id = photo\_tags.tag\_id, I used group by function on tag.id, and then I used order by function on total in descending order with limit 5.

**SQL Statement:**

SELECT

tag\_id, COUNT(photo\_id) AS total\_tags,

tag\_name FROM photo\_tags

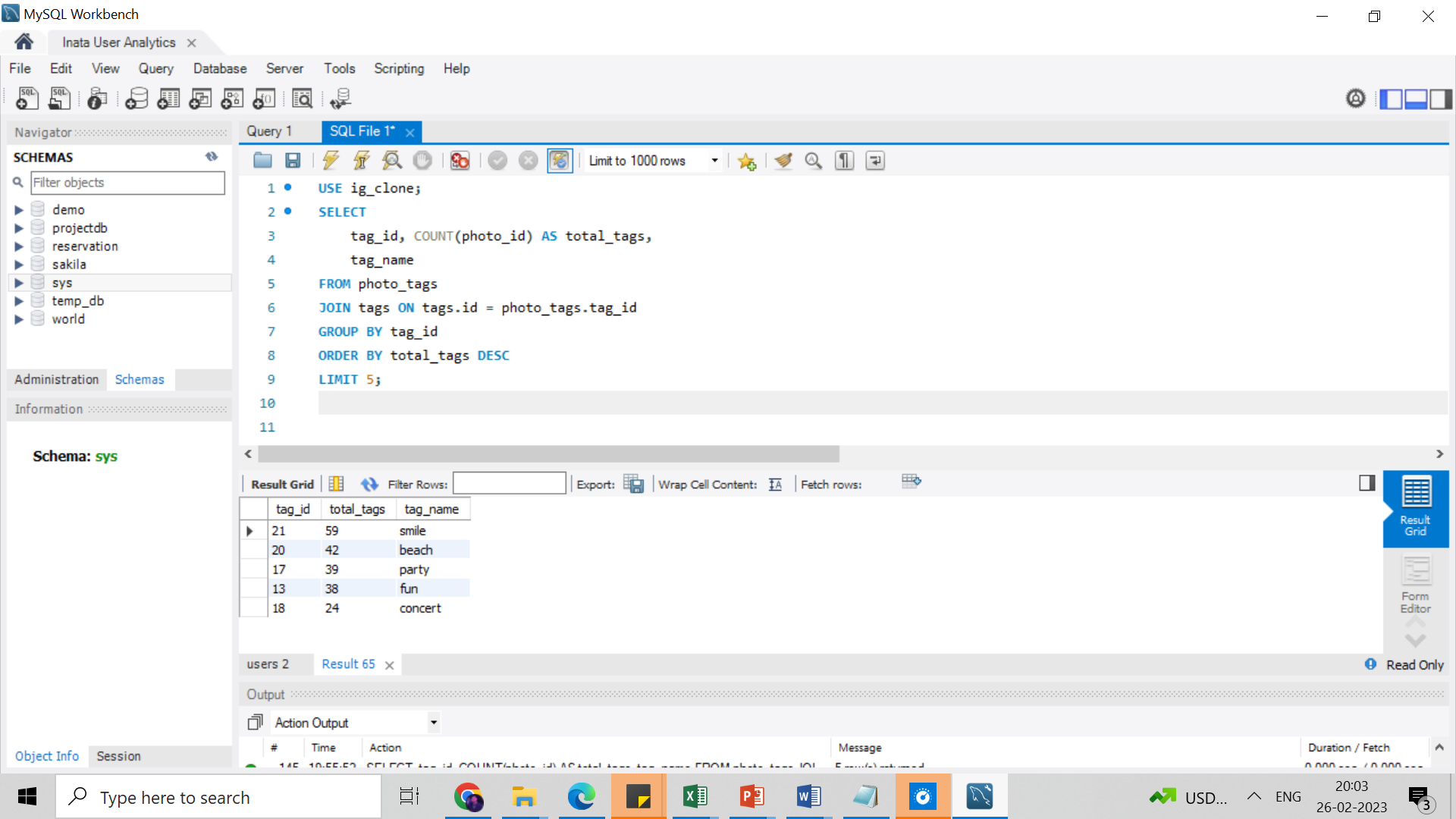
JOIN tags ON tags.id = photo\_tags.tag\_id

GROUP BY tag\_id

ORDER BY total\_tags DESC

LIMIT 5;

**Result:**



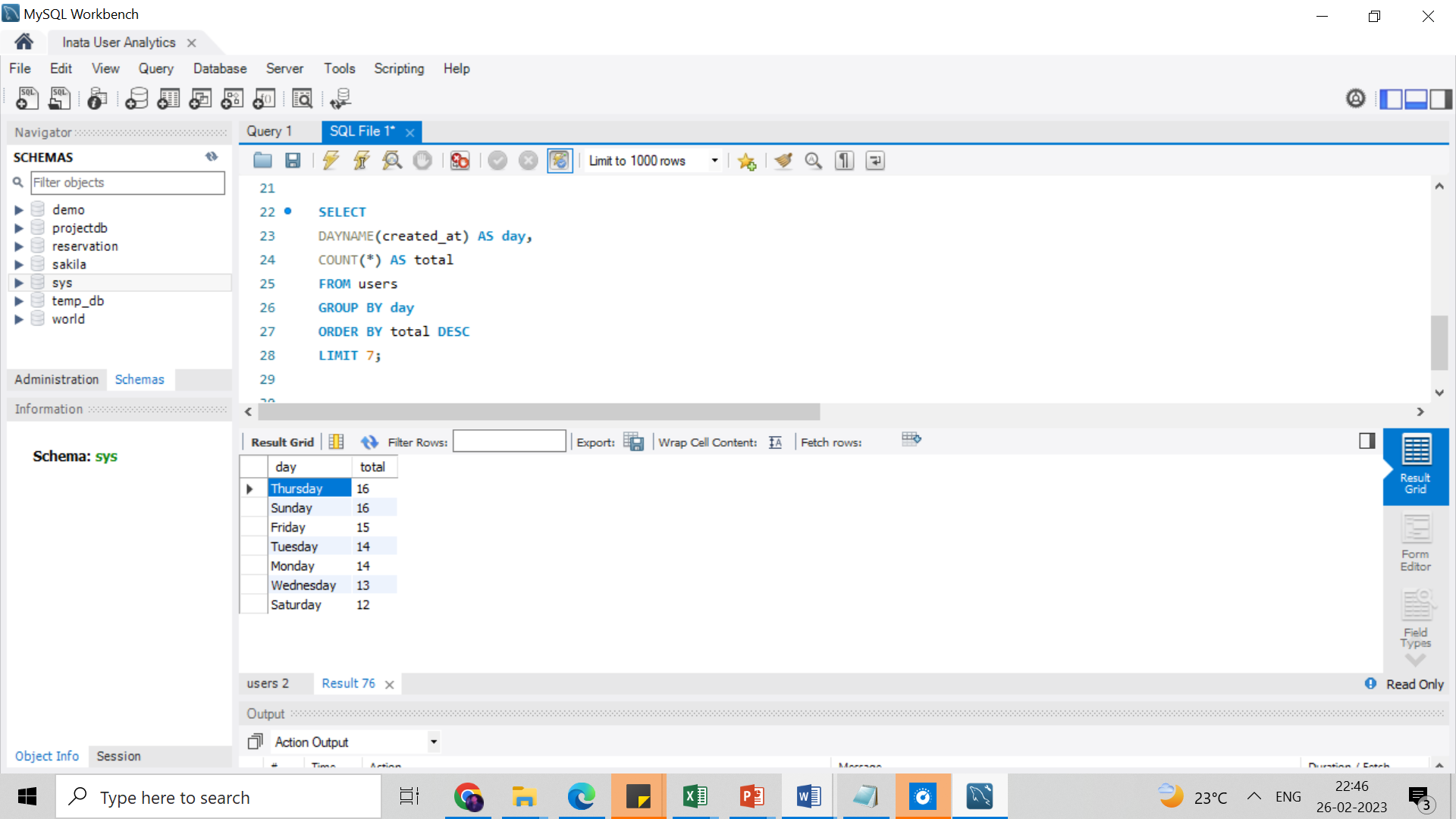
**Conclusion:**

The top 5 popular tags used on Instagram are – smile, beach, party, fun, and concert, with total tag counts of 59, 42, 39, 38, 24, respectively.

1. **Launch AD Campaign:**

Here, I need to Identify on what day of the week do most users register on and provide insights to the management team on when to schedule an ad campaign.

**SQL Statement: Result:**

SELECT

DAYNAME(created\_at) AS day,

COUNT(\*) AS total

FROM users

GROUP BY day

ORDER BY total DESC;

**CONCLUSION:**

With the above query, we learned that most users (16) registered on Thursday and Sunday. So, either is the best day to schedule an ad campaign. Since Thursday is on top, we suggest moving forward with Thursday as the best day to schedule an ad campaign.

**B) Investor Metrics:**

**1. User Engagement:**

Here I need to find how many times does an average user posts on Instagram, and

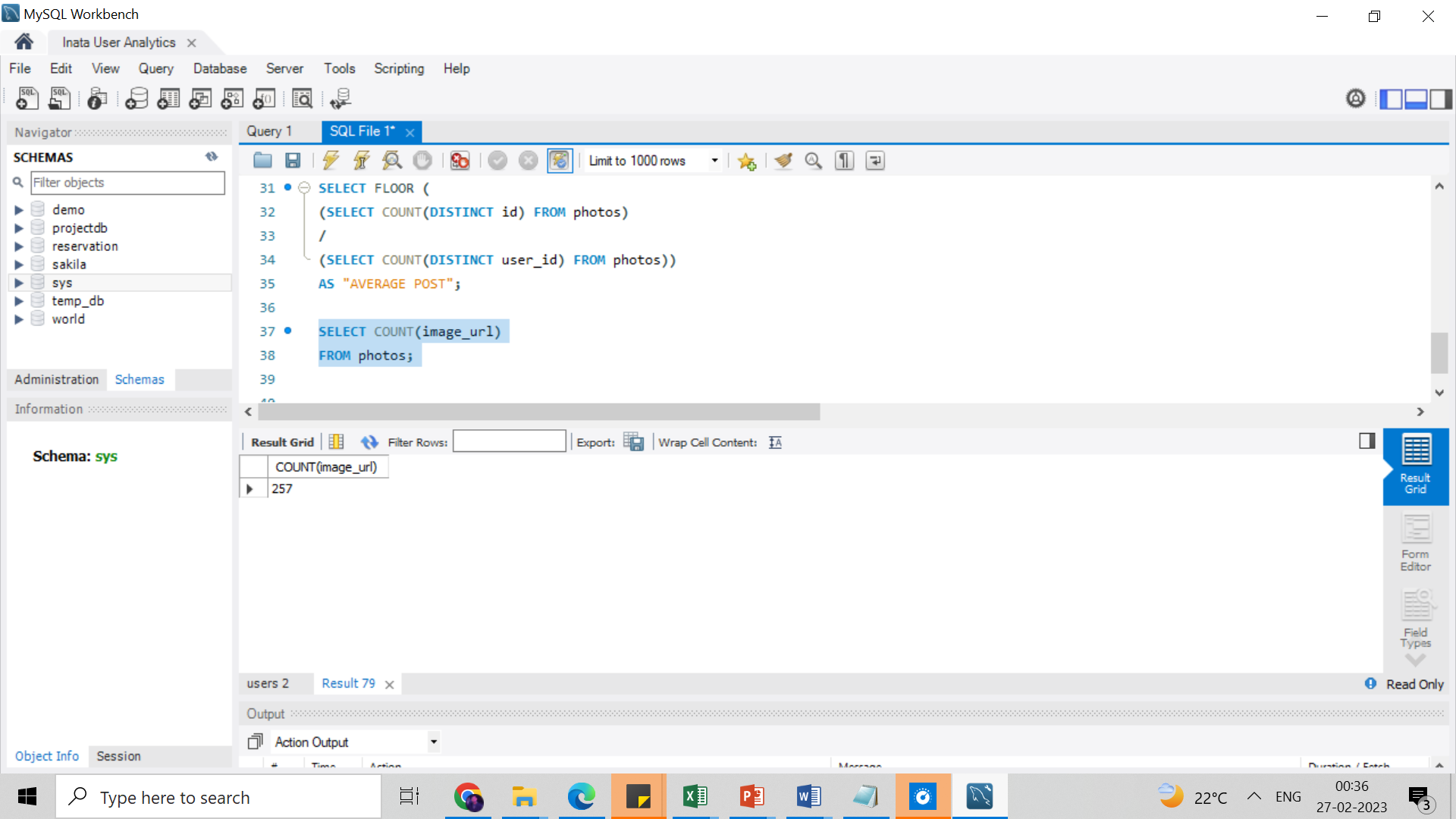
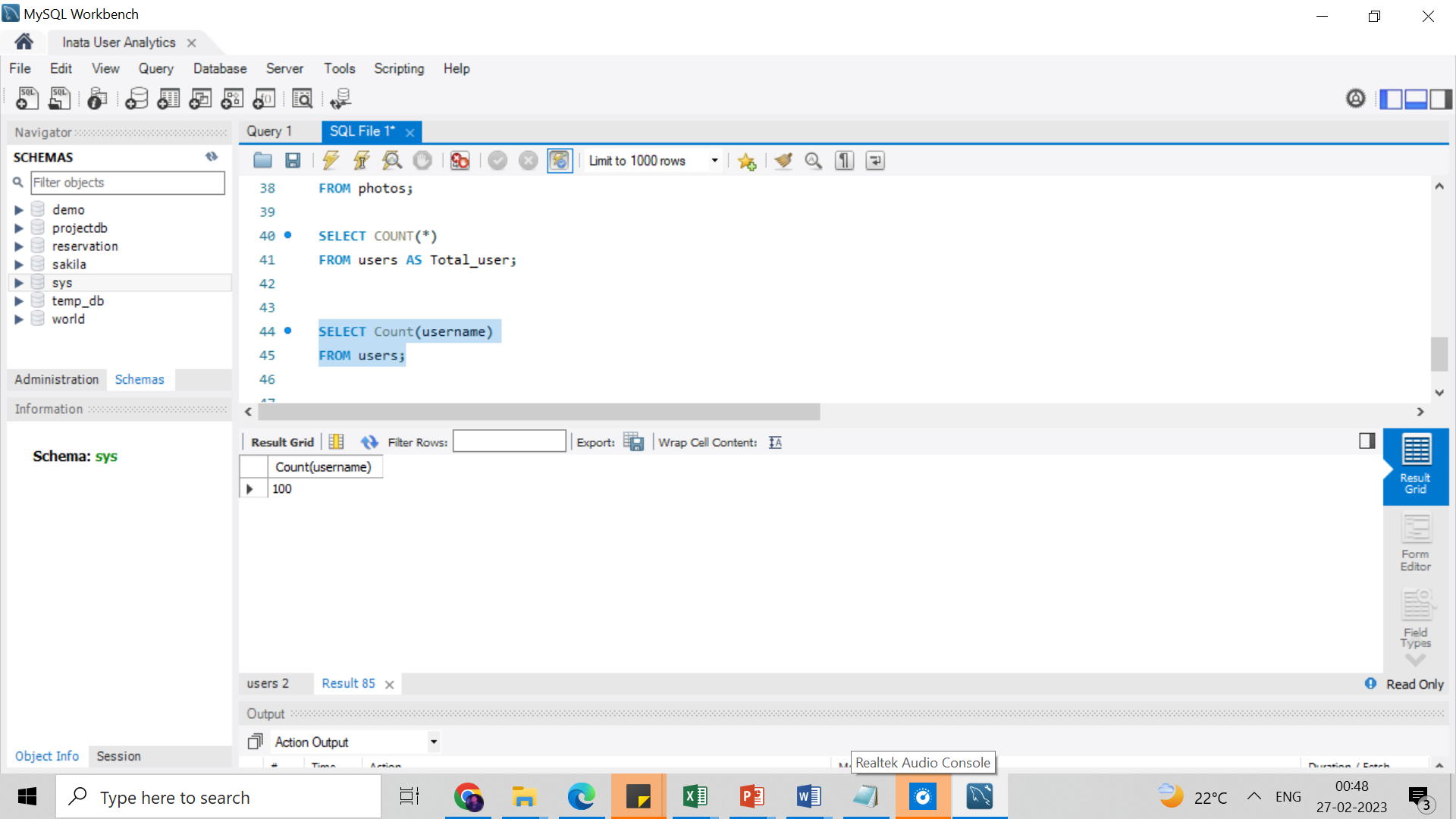
provide the total number of photos on Instagram/total number of users.

This problem can be solved in two parts:

* 1. 1st Part: Finding the total no. of photos and total no. of users on the platform.

Here, I selected count command to count the numbers of images form the table photos.

**SQL statement: Result:**

SELECT COUNT(image\_url)

FROM photos;

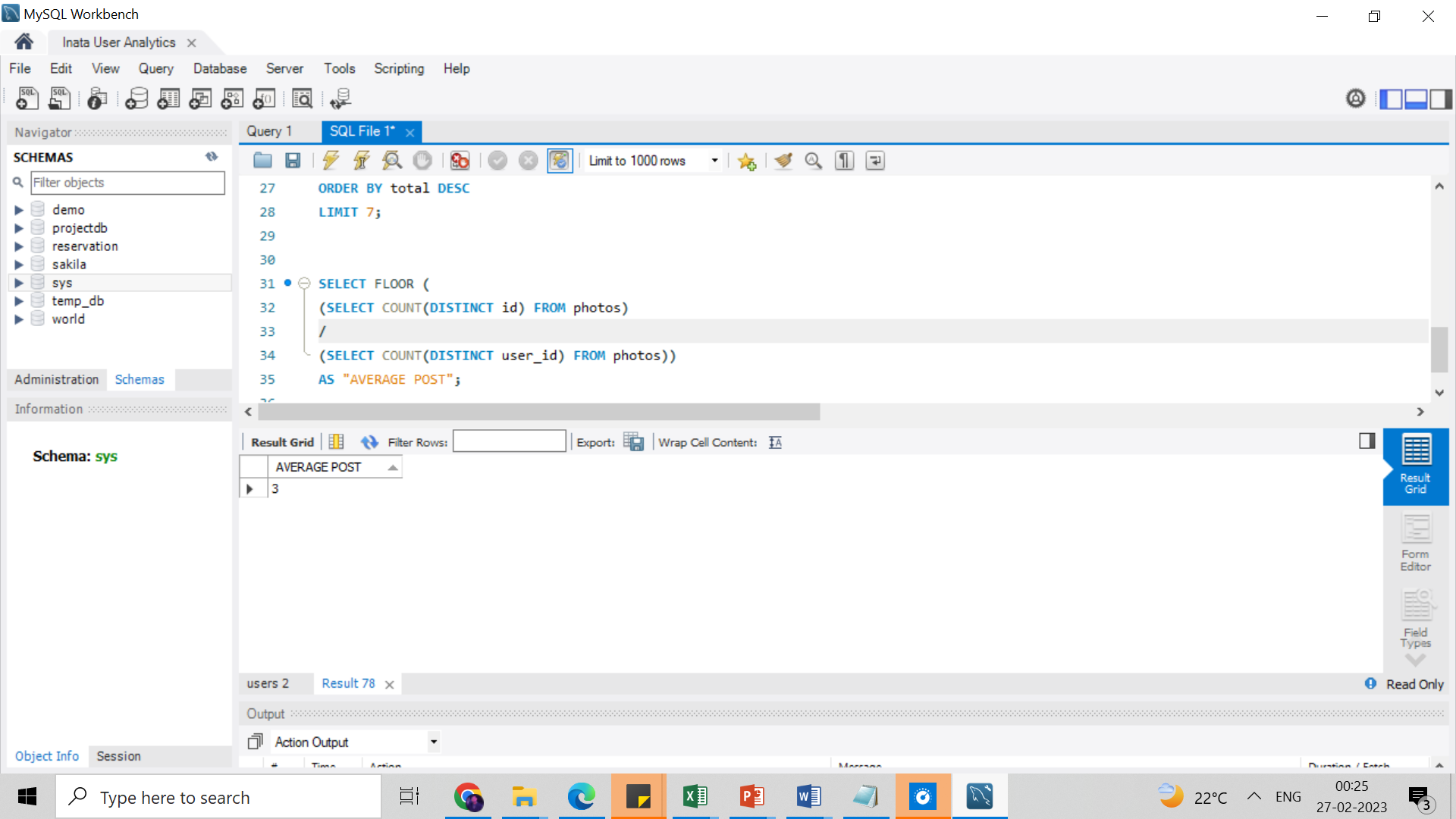
SELECT Count(username)

FROM users;

* 1. 2nd Part: Finding the average user post on the platform.

Here, I selected Floor command returns the largest integer value that is smaller than or equal to a number. Then I proceeded with count command of the distinct ids from the photos table and divided with the another selected another count command of distinct user\_id from table photos and passed it to show output as average post.

**SQL statement: Result:**

SELECT FLOOR (

(SELECT COUNT(DISTINCT id) FROM photos)

/

(SELECT COUNT(DISTINCT user\_id) FROM photos))

AS "AVERAGE POST";

**Conclusion:**

From the above result, we can conclude that an average posts per user is around 3.

1. **Bots & Fake Accounts:**

Here, I need to find and provide data on users (bots) who have liked every single photo on

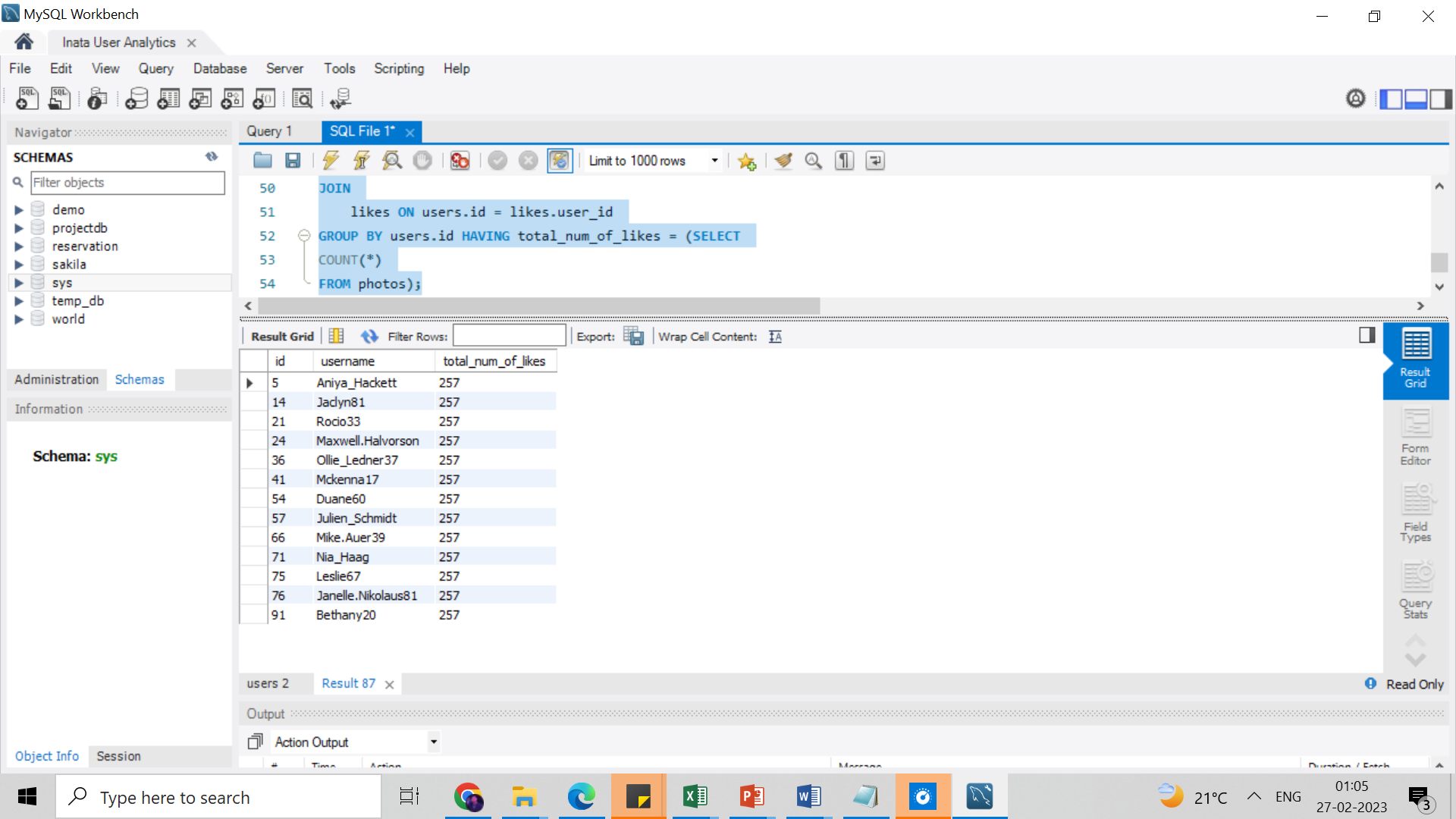
the site.

For that, I first selected user.id, username, and count function as total\_likes\_by\_user from users table, then I used join function to join users table with likes table on

users.id=likes.user\_id, later I used group by function for user.id and having function with

condition total\_likes\_by\_user = select statement for count (\*) from photos table.

**SQL statement: Result:**

SELECT

users.id, username, COUNT(users.id) AS total\_num\_of\_likes

FROM users

JOIN

likes ON users.id = likes.user\_id

GROUP BY users.id HAVING total\_num\_of\_likes = (SELECT

COUNT(\*)

FROM photos);

**Conclusion:**

There is a total of 257 photos available on the platform, and from the above result we can conclude that around 13 users come in the category of bots as they have liked every single picture available on the platform.

**Result:**

While working on the project on Instagram user analytics, I made significant progress. Understanding the fundamentals of SQL and learning how to use it effectively have been very helpful to me. I was able to examine Instagram users' data and their posts. This has made it easier for me to comprehend how Instagram users behave and interact with one another. It has significantly aided me in developing my abilities and knowledge. I now have a better understanding of how to analyse data using SQL. I've also improved my ability to concisely and clearly explain my findings to others. I believe that this project has greatly benefited me and improved my communication skills.

**Drive Link:**

<https://drive.google.com/file/d/1ei9ds8efr8mhSrL1zeFtZ_IsoGPK9g2u/view?usp=sharing>