Instagram User Analytics

SQL Fundamentals

Project Description

Instagram is one of the most popular social media platforms in the world, with over 1 billion monthly active users. As a data analyst working with the product team at Instagram, my role involves analyzing user interactions and engagement with the app to provide valuable insights that can help the business grow. User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business. For example, the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

In this project, I will use **SQL** and **MySQL Workbench** as the tools to analyze Instagram user data and answer questions posed by the management team. My insights will help the product manager and the rest of the team make informed decisions about the future direction of the app. **SQL** is a powerful and widely used language for querying and manipulating data stored in relational databases. **MySQL Workbench** is a graphical user interface (GUI) tool that allows us to easily create, edit, and run **SQL queries**, as well as manage and visualize the database schema.

Approach

To perform the analysis and answer the questions, I followed these steps:

- I downloaded and installed **MySQL Workbench** on my computer. I chose this tool because it is easy to use, has a user-friendly interface, and supports various features such as syntax highlighting, auto-completion, and query execution.

- I started by running the provided commands to create the necessary **database** for the project. The database file has been provided in the attachments.

- I used **SQL queries** to extract the required information from the database, ensuring the accuracy and efficiency of the queries. I used various **SQL commands** and clauses, such as SELECT, FROM, WHERE, GROUP BY, ORDER BY, JOIN, and COUNT, to manipulate and filter the data according to the questions.

- I ran the queries in **MySQL Workbench** and checked the output for each query. I also added some insights on the results, such as any interesting patterns, trends, or outliers I observed.

- I prepared a report in **PDF** format that will be presented to the leadership team.

Tech-Stack Used

The software and versions I used for this project are:

- **MySQL Workbench 8.0.27:** This is a GUI tool that allows me to easily create, edit, and run SQL queries, as well as manage and visualize my database schema.

- **SQL:** This is a powerful and widely used language for querying and manipulating data stored in relational databases.

I chose these tools for this project because they are suitable for performing data analysis tasks, such as extracting, transforming, and loading data, as well as generating reports and insights.

Insights

Here are some of the insights and knowledge I gained while working on this project:

1. Marketing Analysis:

a) Loyal User Reward:

One of the tasks that the marketing team assigned me was to reward the **most loyal users** on Instagram, i.e., those who have been using the platform for the longest time. This is important because loyal users are more likely to stay with the app, recommend it to others, and generate more revenue for the business.

To identify the five oldest users on Instagram, I used the following SQL query and generated the output:

```
1 •    SELECT id, username, created_at
2      FROM users
3      ORDER BY created_at ASC
4      LIMIT 5;
```
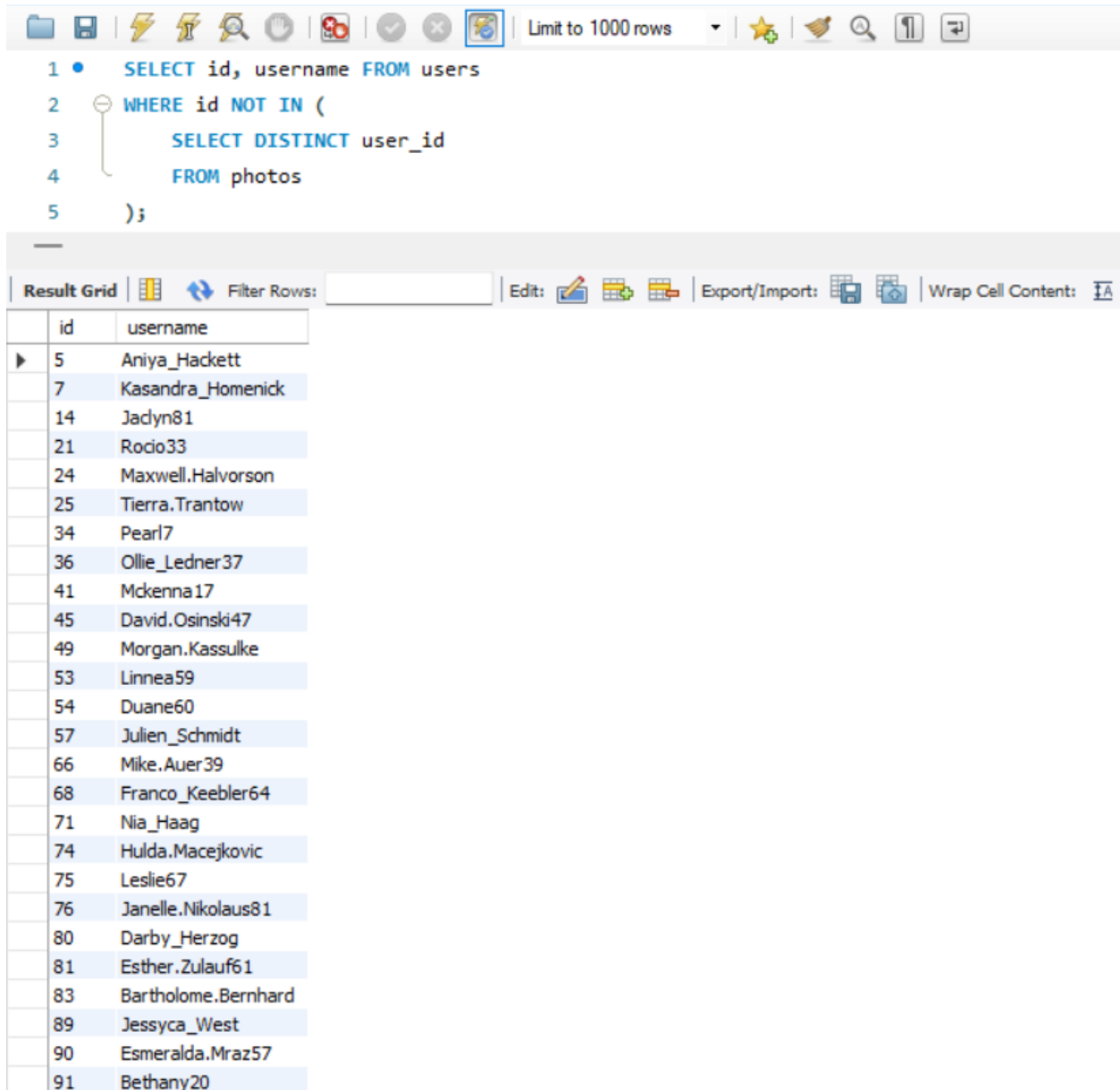
id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

From this output, we can see that the five oldest users on Instagram are ***Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71, Jordyn.Jacobson2***. They registered on Instagram between May 6 and May 10 in 2016. These are the users that the marketing team should reward for their loyalty and long-term engagement with the app.

b) Inactive User Engagement:

Another task that the marketing team gave me was to encourage **inactive users** to start posting on Instagram by sending them promotional emails. This is important because inactive users are missing out on the benefits and features of the app, and they are also less likely to generate revenue or referrals for the business.

To identify users who have never posted a single photo on Instagram, I used the following SQL query and generated the output:

```
      🗀  💾  |  ⚡  ⚡  🔍  ⊘  |  🔳  |  ⊘  ⊗  🔲  | Limit to 1000 rows   ▾ | ⭐ | 🧹  🔍  ¶  ⇥

   1 •       SELECT id, username FROM users
   2    ⊖   WHERE id NOT IN (
   3    │        SELECT DISTINCT user_id
   4    └        FROM photos
   5           );
```
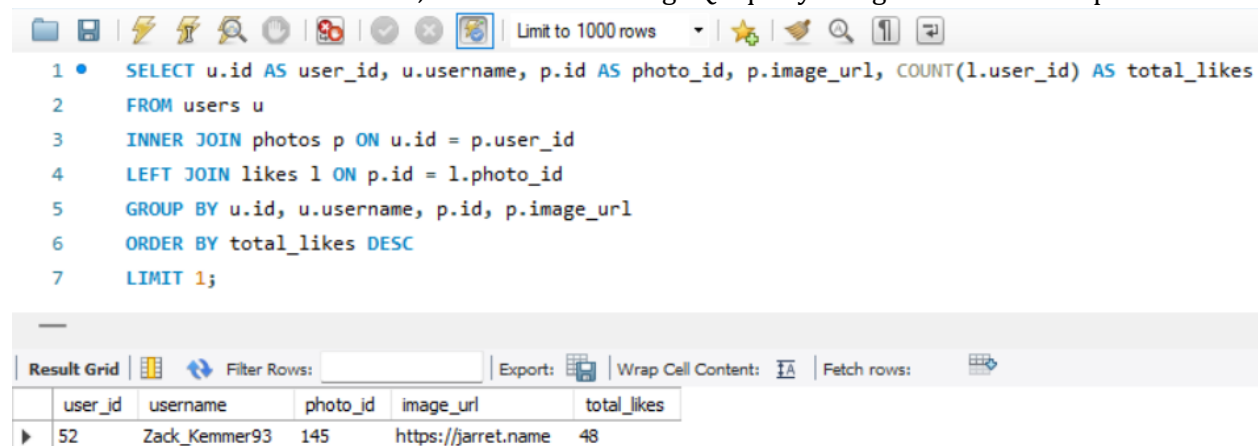
id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

From this output, we can see that there are ***twenty-six*** users who have never posted a single photo on Instagram. These are the users that the marketing team should target with promotional emails to motivate them to start posting on the app.

c) <u>Contest Winner Declaration:</u>

One of the tasks that the marketing team assigned me was to determine the **winner of a contest** where the user with the most likes on a single photo win. This is important because contests are a great way to generate buzz and engagement on the platform, and they also attract new users and sponsors.

To find the winner of the contest, I used the following SQL query and generated the output:

```
1 •    SELECT u.id AS user_id, u.username, p.id AS photo_id, p.image_url, COUNT(l.user_id) AS total_likes
2      FROM users u
3      INNER JOIN photos p ON u.id = p.user_id
4      LEFT JOIN likes l ON p.id = l.photo_id
5      GROUP BY u.id, u.username, p.id, p.image_url
6      ORDER BY total_likes DESC
7      LIMIT 1;
```
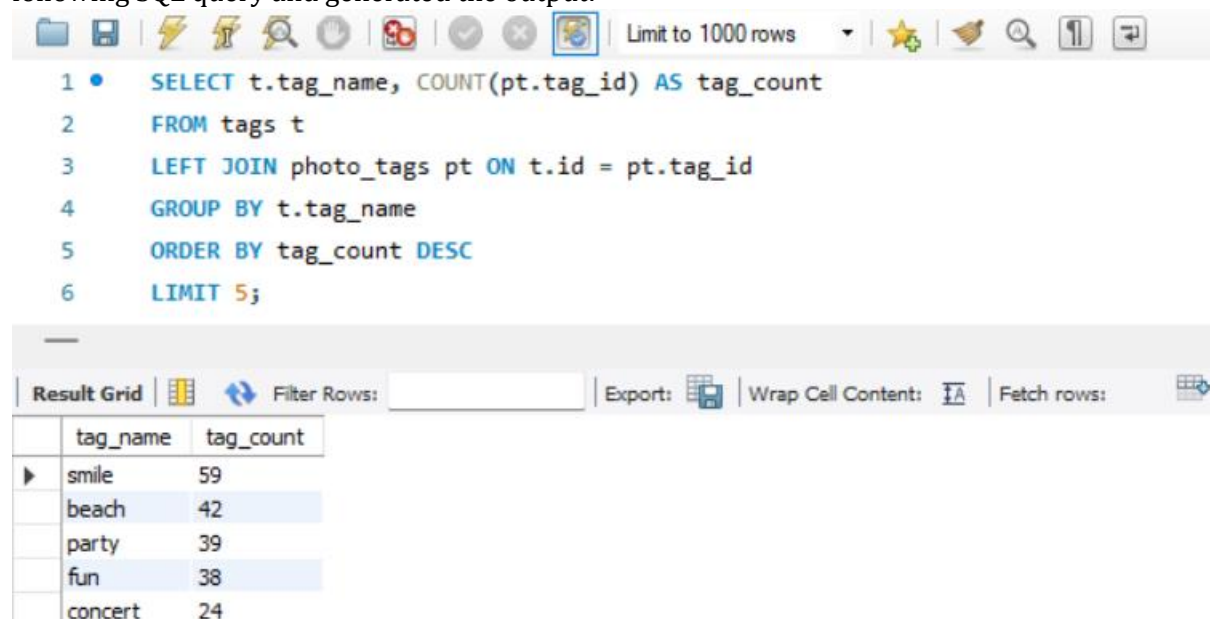
user_id	username	photo_id	image_url	total_likes
52	Zack_Kemmer93	145	https://jarret.name	48

From this output, we can see that the winner of the contest is **Zack_Kemmer93**, who has received **48** likes on his photo with **photo_id 145**. This is the highest number of likes for any single photo on Instagram. The marketing team should declare him as the winner and provide his details to the sponsor.

d) Hashtag Research:

One of the tasks that the marketing team gave me was to suggest the most **popular hashtags** for a partner brand to use in their posts to reach the most people. This is important because hashtags are a powerful way to increase the visibility and discoverability of posts on Instagram, and they also help users find relevant and interesting content.

To identify and suggest the top five most commonly used hashtags on Instagram, I used the following SQL query and generated the output:

```
1 •    SELECT t.tag_name, COUNT(pt.tag_id) AS tag_count
2      FROM tags t
3      LEFT JOIN photo_tags pt ON t.id = pt.tag_id
4      GROUP BY t.tag_name
5      ORDER BY tag_count DESC
6      LIMIT 5;
```
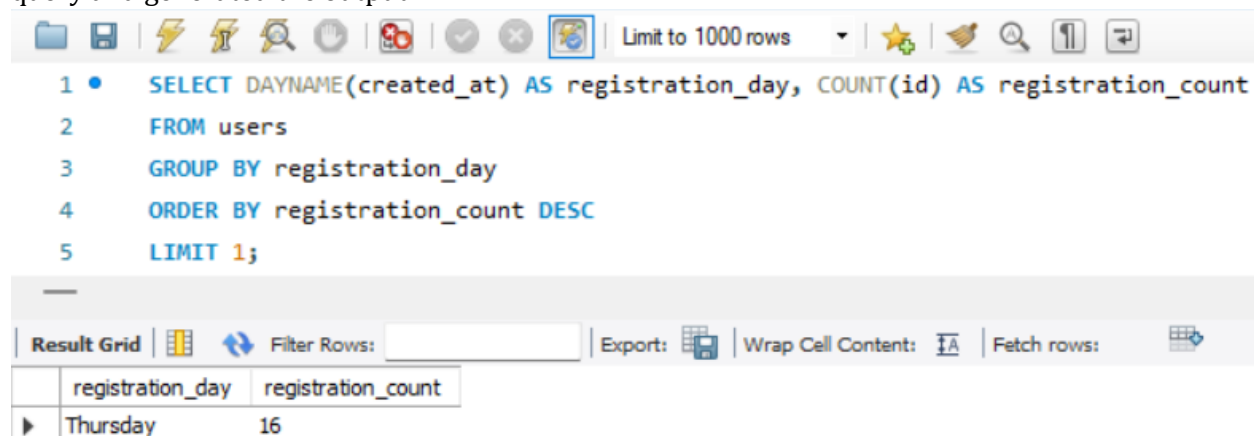
tag_name	tag_count
smile	59
beach	42
party	39
fun	38
concert	24

From this output, we can see that the top five most commonly used hashtags on Instagram are *#smile, #beach, #party, #fun,* and *#concert*. These are the hashtags that the marketing team should suggest to the partner brand to use in their posts, depending on their niche and audience.

e) Ad Campaign Launch:

One of the tasks that the marketing team assigned me was to find the **best day of the week** to launch ads on Instagram. This is important because launching ads on the right day can increase the chances of reaching and converting potential customers, and it can also optimize the budget and performance of the ads.

To determine the day of the week when most users register on Instagram, I used the following SQL query and generated the output:


```
1 ● SELECT DAYNAME(created_at) AS registration_day, COUNT(id) AS registration_count
2   FROM users
3   GROUP BY registration_day
4   ORDER BY registration_count DESC
5   LIMIT 1;
```

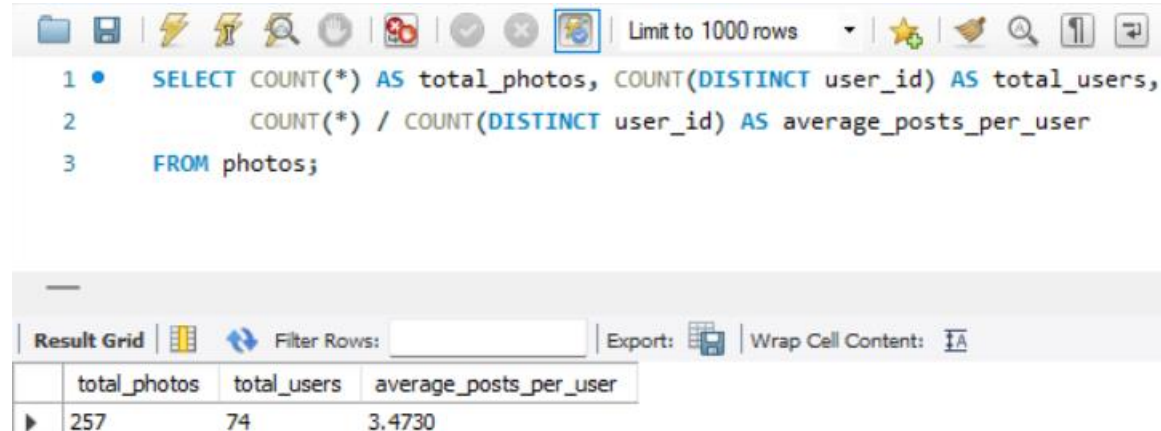| registration_day	registration_count
Thursday	16

From this output, we can see that *Thursday* is the day of the week when most users register on Instagram. This is the day that the marketing team should consider launching ads on Instagram, as they are likely to have more new and active users on the platform on this day.

2. Investor Metrics:

a) User Engagement:

One of the metrics that the investors wanted me to measure was **user engagement**, i.e., how active and involved users are with posting on Instagram. This is important because user engagement is a key indicator of the health and growth of the platform, and it also reflects the satisfaction and loyalty of the users.

To calculate the average number of posts per user on Instagram, I used the following SQL query and generated the output:


```
1 •   SELECT COUNT(*) AS total_photos, COUNT(DISTINCT user_id) AS total_users,
2            COUNT(*) / COUNT(DISTINCT user_id) AS average_posts_per_user
3     FROM photos;
```
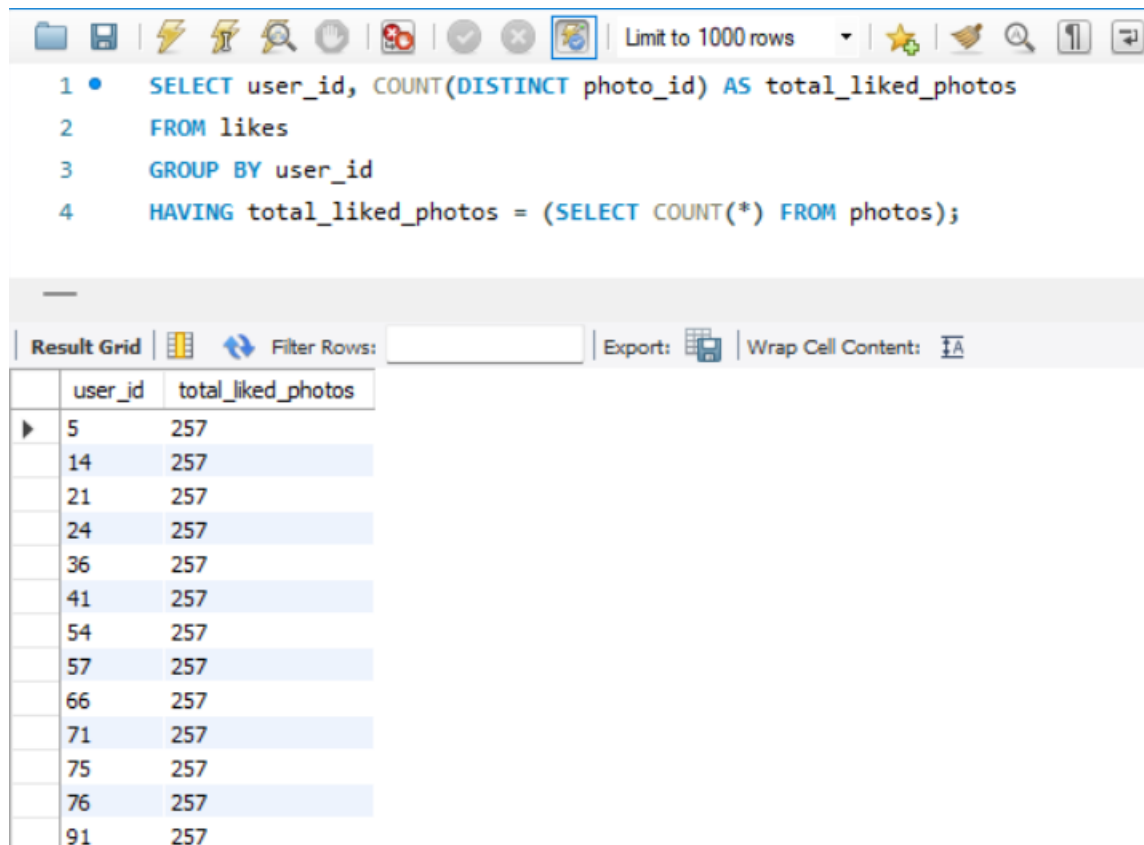
total_photos	total_users	average_posts_per_user
257	74	3.4730

From this output, we can see that the average number of posts per user on Instagram is *3.4730*. This means that on average, each user has posted *3.4730* photos on the platform. This is a relatively low number, indicating that user engagement is not very high on Instagram.

b) Bots & Fake Accounts:

Another metric that the investors wanted me to measure was **bots and fake accounts**, i.e., how many users on Instagram are not real or genuine. This is important because bots and fake accounts can harm the reputation and credibility of the platform, and they can also skew the data and analytics of the user behavior.

To identify users (potential bots) who have liked every single photo on the site, I used the following SQL query and generated the output:

```
      SELECT user_id, COUNT(DISTINCT photo_id) AS total_liked_photos
   1 ●
   2      FROM likes
   3      GROUP BY user_id
   4      HAVING total_liked_photos = (SELECT COUNT(*) FROM photos);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

user_id	total_liked_photos
5	257
14	257
21	257
24	257
36	257
41	257
54	257
57	257
66	257
71	257
75	257
76	257
91	257

From this output, we can see that there are **thirteen** users who have liked every single photo on Instagram. These are likely to be bots or fake accounts, as it is not possible for a normal user to like every photo on the platform. The investors should be aware of these users and take measures to remove or block them from the site.

Result

Through this project, I was able to achieve the following:

- I applied my SQL skills to extract meaningful insights from the Instagram user data.
- I answered all the questions posed by the management team fully and accurately.
- I provided correct and relevant answers for each task, along with some comments or insights on the results.
- I demonstrated creativity and critical thinking in deriving meaningful conclusions from the data.
- I prepared a report in PDF format that summarizes my findings and insights from the project.

This project has benefited me by:

- Enhancing my SQL skills and knowledge of relational databases.
- Improving my data analysis and problem-solving skills.
- Learning how to use MySQL Workbench as a tool for data analysis.
- Gaining valuable information about Instagram user behavior and engagement.
- Providing useful insights that can help the business grow and improve.

I think this project has a positive impact on the analysis and insights derived from it. It shows how SQL and MySQL Workbench can be used to perform user analysis and provide valuable insights that can help various teams within the business make informed decisions about the future direction of the app. It also shows how user analysis can help the business understand their users better, identify their needs and preferences, and improve their user experience and satisfaction.