

## Instagram User Analytics

### Project Description

The purpose of this project is to derive relevant insights from Instagram users' data that will enable various teams in the organization develop tremendously.

Instagram user data is made available in this initiative. The provided data will be used to obtain important information on user interaction and engagement with the App. I must analyze and track users, how they use the digital platform, and recommend the launch of an ad campaign, whether the user is active or not, who is the most loyal user, which hash tag is popular among users, which user is more active and liking more posts, average number of posts per user, whether the user is a bot or not, and so on.

The analysis in this project will assist the product team in deciding whether to launch new features, the development team in improving user experience on the app, the marketing team in deciding whether to launch a new campaign, and so on.

The aforementioned objectives will be met by utilizing SQL approaches and MySQL technologies.

### Approach

- Understanding data: Database named ig\_clone is given in tabular format having total 7 tables with table names given below:

| Table Name | Number of Rows | Number of Columns |
|------------|----------------|-------------------|
| users      | 100            | 3                 |
| photos     | 257            | 4                 |
| comments   | 7488           | 5                 |
| likes      | 8782           | 3                 |
| follows    | 7623           | 3                 |
| tags       | 21             | 3                 |
| photo_tags | 501            | 2                 |

- Understanding of data-types and table schemas;
- Recognition and understanding of Referential-Integrity-Constraint between tables;
- Formation of modular queries using SQL techniques;
- Merging of two or more queries to get actual answers for the given task;
- Capturing results/output.

## Tech-Stack Used

I have executed the query on MY SQL workbench installed on windows 10 operating system, more details are given below:

| Software Details |                              |
|------------------|------------------------------|
| Name:            | Local instance MySQL80       |
| Host:            | localhost                    |
| Port:            | 3306                         |
| Login User:      | root                         |
| Current User:    | root@localhost               |
| SSL cipher:      | SSL not used                 |
| Server           |                              |
| Product:         | MySQL Community Server - GPL |
| Version:         | 8.0.34                       |
| Connector        |                              |
| Version:         | C++ 8.1.0                    |

I choose this MY SQL workbench because of the following reasons:

- It is available for my operating system;
- It takes less memory to install in the system;
- It allows access to data directly;
- Easy to understand and intuitive GUI;
- I can analyse multiple table at once;
- It provides cross-platform support;

## Insights

- The oldest user on Instagram is having id 80, who is on the app for 2653 days and the user with id 67 also engaged with this app for 2653 days.
- The user with id 5,7,14,21,24,25,34 ... has never posted a single photo on Instagram.
- The user with id 52 posted a photo (photo\_id = 145) on Instagram which gets 48 likes which is highest.
- Hashtags with names; smile, beach, party, fun, food; is most commonly used.
- Thursday is the day on which most users registered on the app.
- Some metrics is calculated for Instagram users and found that each user posted on average 2.57 photos on the app.
- It is also found that some bot users exist who likes all the photo post, these users are with id 21,71,5,66,41,14,57,24,76,75,54,91....

## Result

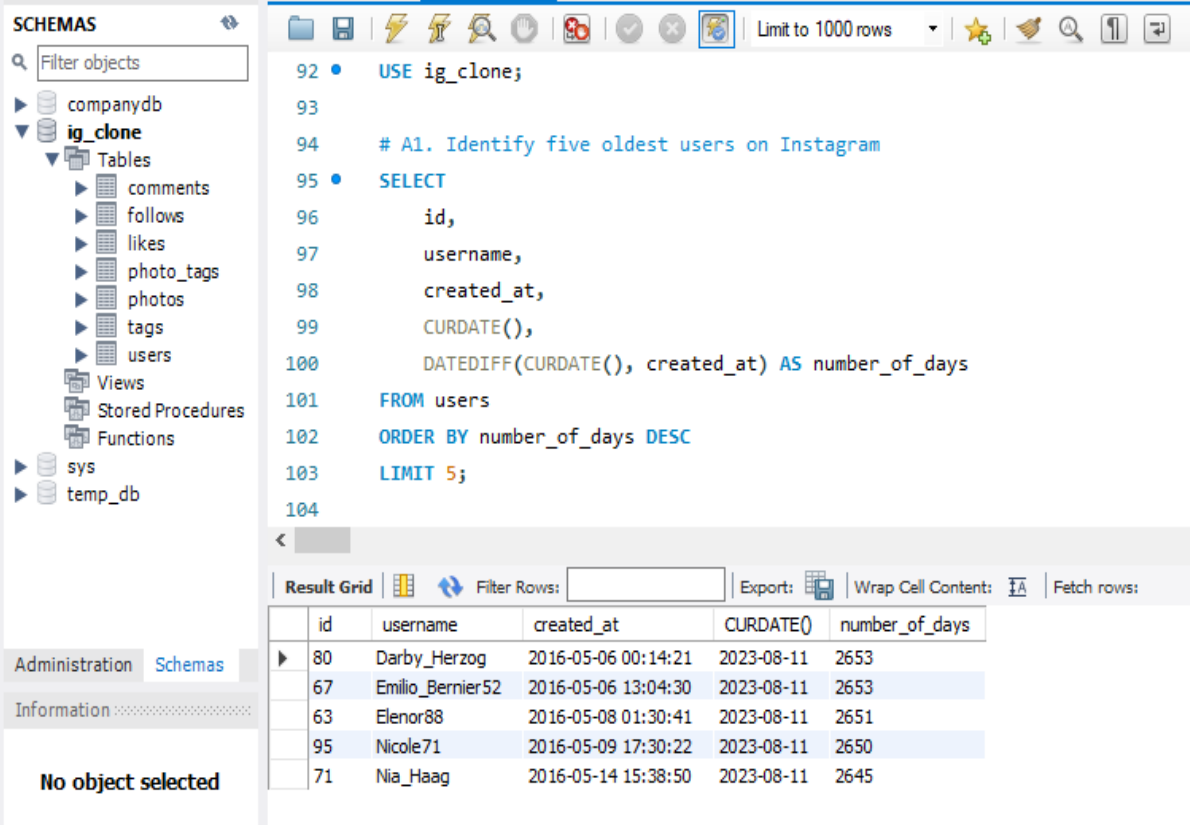
### A) Marketing Analysis:

**Loyal User Reward:** The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

#### Task:

Identify the five oldest users on Instagram from the provided database.

#### Query and Output:



The screenshot shows a database management interface. On the left, the 'SCHEMAS' panel displays a tree view with 'companydb' and 'ig\_clone'. Under 'ig\_clone', there are 'Tables' (comments, follows, likes, photo\_tags, photos, tags, users) and 'Views', 'Stored Procedures', and 'Functions'. The 'users' table is selected. The main area shows a SQL query:

```
92 • USE ig_clone;
93
94 # A1. Identify five oldest users on Instagram
95 • SELECT
96     id,
97     username,
98     created_at,
99     CURDATE(),
100     DATEDIFF(CURDATE(), created_at) AS number_of_days
101 FROM users
102 ORDER BY number_of_days DESC
103 LIMIT 5;
104
```

Below the query, the 'Result Grid' shows the output. The columns are 'id', 'username', 'created\_at', 'CURDATE()', and 'number\_of\_days'. The results are:

| id | username         | created_at          | CURDATE()  | number_of_days |
|----|------------------|---------------------|------------|----------------|
| 80 | Darby_Herzog     | 2016-05-06 00:14:21 | 2023-08-11 | 2653           |
| 67 | Emilio_Bernier52 | 2016-05-06 13:04:30 | 2023-08-11 | 2653           |
| 63 | Elenor88         | 2016-05-08 01:30:41 | 2023-08-11 | 2651           |
| 95 | Nicole71         | 2016-05-09 17:30:22 | 2023-08-11 | 2650           |
| 71 | Nia_Haag         | 2016-05-14 15:38:50 | 2023-08-11 | 2645           |

#### Result:

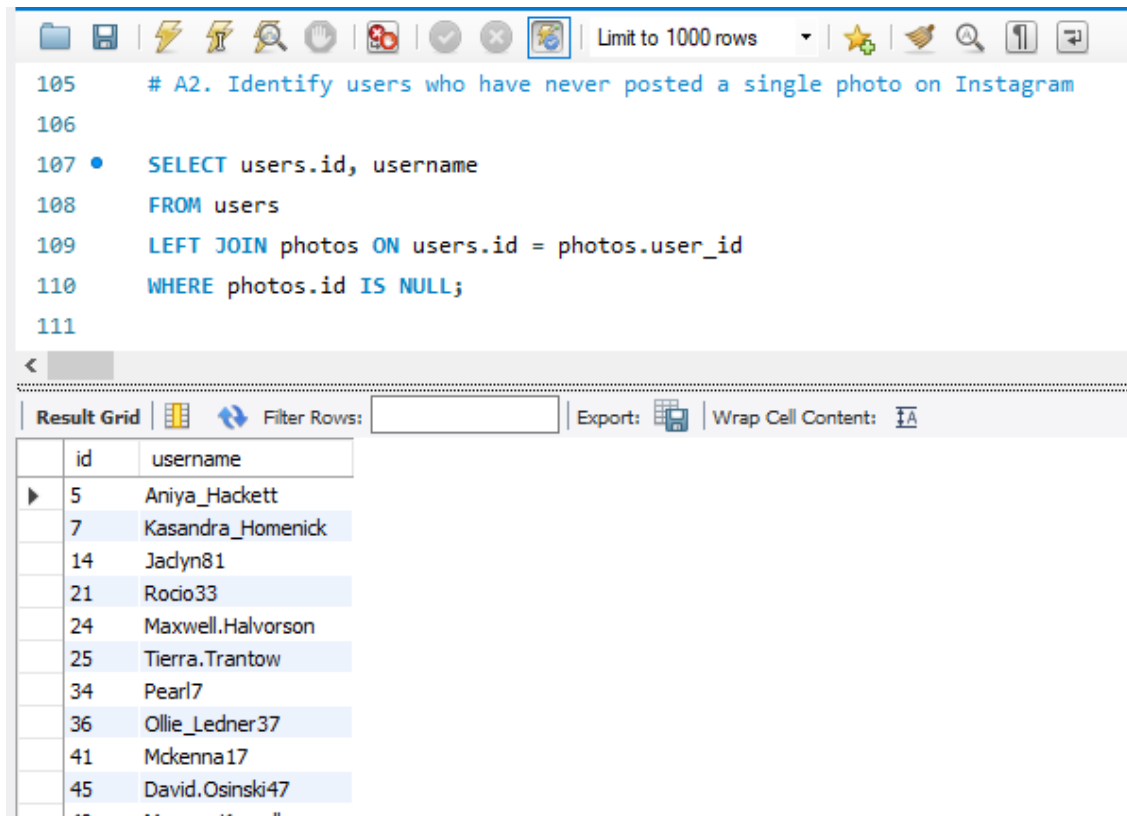
The Loyal user reward will be given to user with id 80 and 67.

**Inactive User Engagement:** The team wants to encourage inactive users to start posting by sending them promotional emails.

**Task:**

Identify users who have never posted a single photo on Instagram.

**Query and Output:**



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
105 # A2. Identify users who have never posted a single photo on Instagram
106
107 • SELECT users.id, username
108 FROM users
109 LEFT JOIN photos ON users.id = photos.user_id
110 WHERE photos.id IS NULL;
111
```

Below the query editor is a 'Result Grid' showing the output of the query. The grid has two columns: 'id' and 'username'. The results are as follows:

|   | id | username           |
|---|----|--------------------|
| ▶ | 5  | Aniya_Hackett      |
|   | 7  | Kassandra_Homenick |
|   | 14 | Jadlyn81           |
|   | 21 | Rocio33            |
|   | 24 | Maxwell.Halvorson  |
|   | 25 | Tierra.Trantow     |
|   | 34 | Pearl7             |
|   | 36 | Ollie_Ledner37     |
|   | 41 | Mckenna17          |
|   | 45 | David.Osinski47    |
|   | 48 | Maxwell_Kozlowski  |

**Result:**

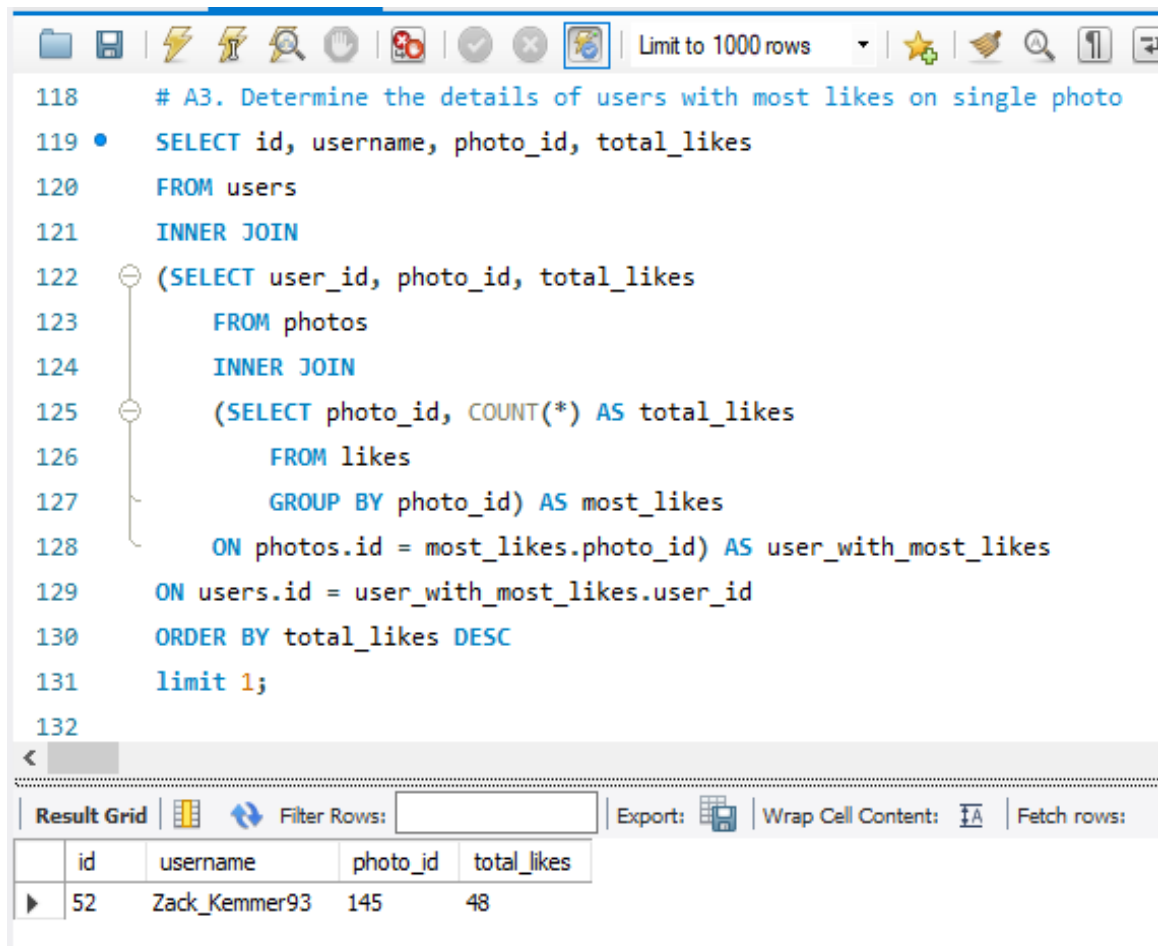
The promotional emails need to be sent to users with id 5,7,14,21,24,25,34,36,41,45, etc.

**Contest Winner Declaration:** The team has organized a contest where the user with the most likes on a single photo wins.

**Task:**

Determine the winner of the contest and provide their details to the team.

**Query and Output:**



```
118 # A3. Determine the details of users with most likes on single photo
119 • SELECT id, username, photo_id, total_likes
120 FROM users
121 INNER JOIN
122 (SELECT user_id, photo_id, total_likes
123 FROM photos
124 INNER JOIN
125 (SELECT photo_id, COUNT(*) AS total_likes
126 FROM likes
127 GROUP BY photo_id) AS most_likes
128 ON photos.id = most_likes.photo_id) AS user_with_most_likes
129 ON users.id = user_with_most_likes.user_id
130 ORDER BY total_likes DESC
131 limit 1;
132
```

Result Grid

|   | id | username      | photo_id | total_likes |
|---|----|---------------|----------|-------------|
| ▶ | 52 | Zack_Kemmer93 | 145      | 48          |

**Result:**

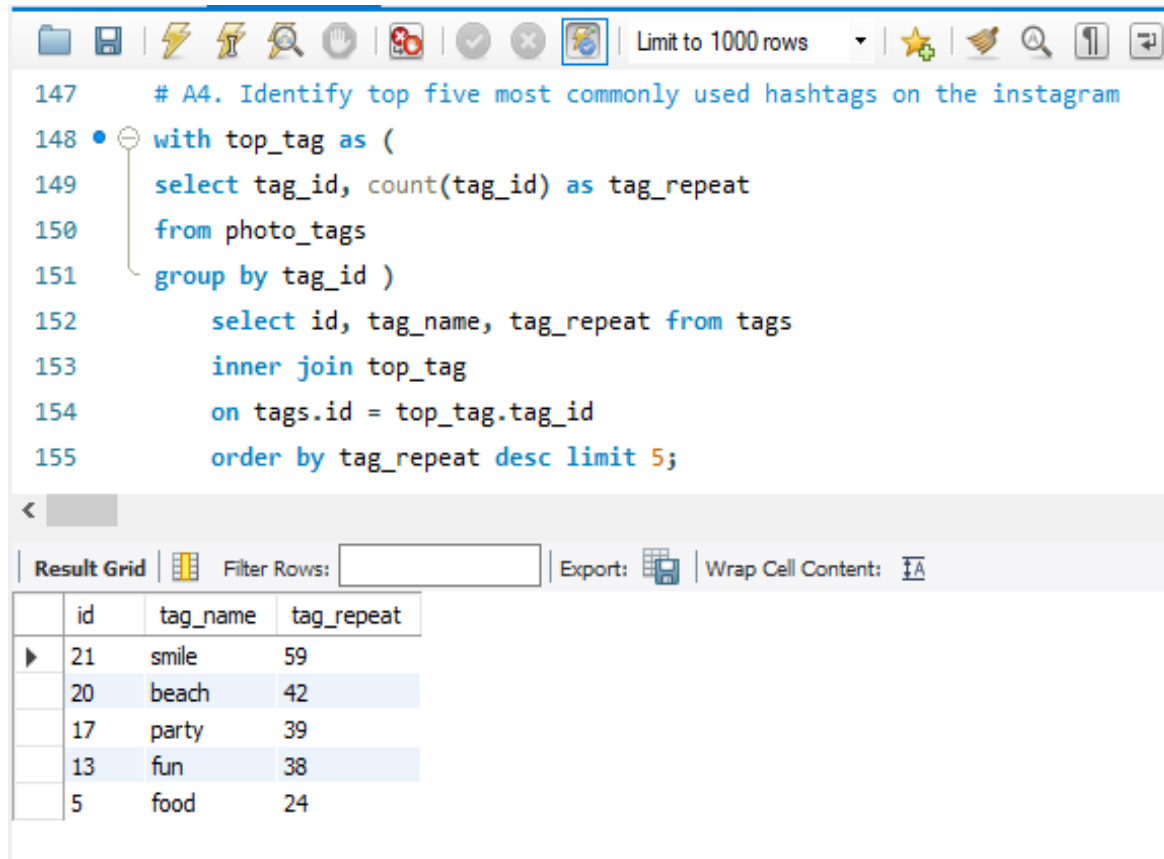
The winner of the contest is the user with id **52** who posted a photo (photo\_id =145) which got total 48 likes.

**Hashtag Research:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

**Task:**

Identify and suggest the top five most commonly used hashtags on the platform.

**Query and Output:**



```
147 # A4. Identify top five most commonly used hashtags on the instagram
148 with top_tag as (
149     select tag_id, count(tag_id) as tag_repeat
150     from photo_tags
151     group by tag_id )
152     select id, tag_name, tag_repeat from tags
153     inner join top_tag
154     on tags.id = top_tag.tag_id
155     order by tag_repeat desc limit 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

|   | id | tag_name | tag_repeat |
|---|----|----------|------------|
| ▶ | 21 | smile    | 59         |
|   | 20 | beach    | 42         |
|   | 17 | party    | 39         |
|   | 13 | fun      | 38         |
|   | 5  | food     | 24         |

**Result:**

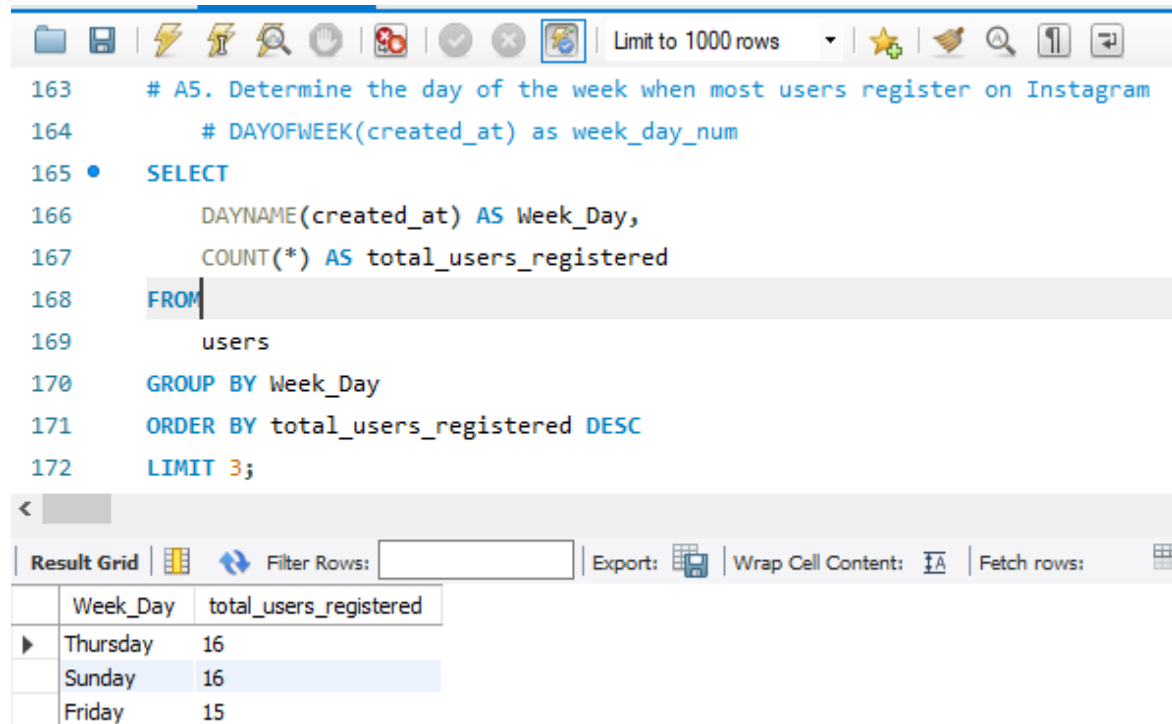
The partner brand can use the hashtags smile, beach, party, fun, food in their post which will definitely help them to reach more people.

**Ad Campaign Launch:** The team wants to know the best day of the week to launch ads.

**Task:**

Determine the day of the week when most users register on Instagram.

**Query and Output:**



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
163 # A5. Determine the day of the week when most users register on Instagram
164 # DAYOFWEEK(created_at) as week_day_num
165 • SELECT
166     DAYNAME(created_at) AS Week_Day,
167     COUNT(*) AS total_users_registered
168 FROM
169     users
170 GROUP BY Week_Day
171 ORDER BY total_users_registered DESC
172 LIMIT 3;
```

Below the query editor, the results are displayed in a table with columns 'Week\_Day' and 'total\_users\_registered'.

| Week_Day | total_users_registered |
|----------|------------------------|
| Thursday | 16                     |
| Sunday   | 16                     |
| Friday   | 15                     |

**Result:**

**Thursday** and **Saturday** can be the best day to launch the campaign for advertisement

## B) Investor Metrics:

**User Engagement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

### Task:

Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

### Query and Output:

Average number of post per user is same as total number of photos divided by total numbers of users.



```
181
182 # B1.2. Total number of photos on Instagram divided by total number of users.
183 • SELECT
184     ROUND((SELECT COUNT(*)
185             FROM photos) / (SELECT COUNT(*)
186                             FROM users), 2) AS photo_per_user;
187
```

Result Grid

| photo_per_user |
|----------------|
| 2.57           |

### Result:

It is found that every user posts 2.57 photo on Instagram

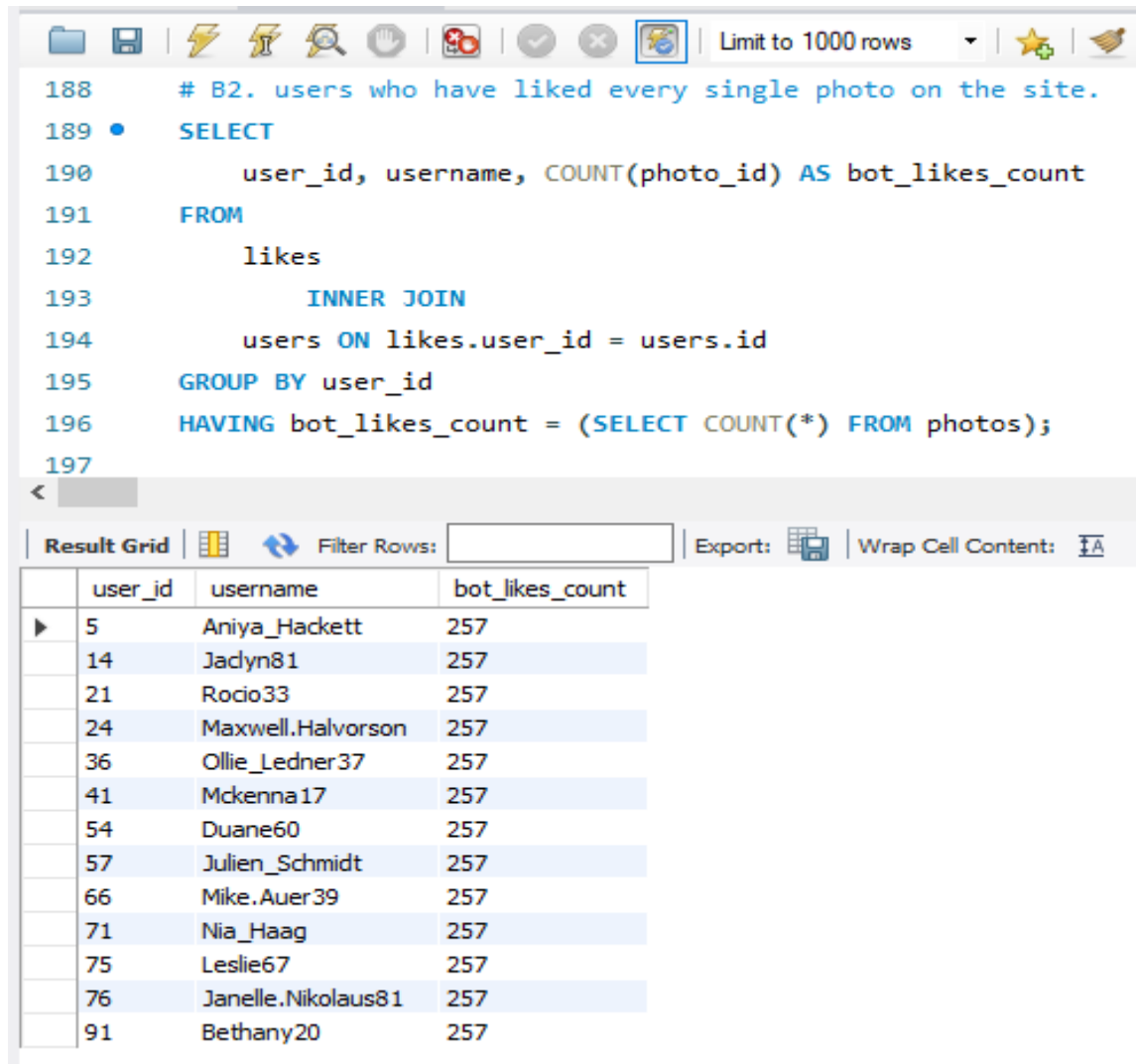


**Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.

**Task:**

Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

**Query and Output:**



```
188 # B2. users who have liked every single photo on the site.
189 • SELECT
190     user_id, username, COUNT(photo_id) AS bot_likes_count
191 FROM
192     likes
193     INNER JOIN
194     users ON likes.user_id = users.id
195 GROUP BY user_id
196 HAVING bot_likes_count = (SELECT COUNT(*) FROM photos);
197
```

Result Grid

|   | user_id | username           | bot_likes_count |
|---|---------|--------------------|-----------------|
| ▶ | 5       | Aniya_Hackett      | 257             |
|   | 14      | Jadyn81            | 257             |
|   | 21      | Rocio33            | 257             |
|   | 24      | Maxwell.Halvorson  | 257             |
|   | 36      | Ollie_Ledner37     | 257             |
|   | 41      | Mckenna17          | 257             |
|   | 54      | Duane60            | 257             |
|   | 57      | Julien_Schmidt     | 257             |
|   | 66      | Mike.Auer39        | 257             |
|   | 71      | Nia_Haag           | 257             |
|   | 75      | Leslie67           | 257             |
|   | 76      | Janelle.Nikolaus81 | 257             |
|   | 91      | Bethany20          | 257             |

**Result:**

Yes, there are many bot/dummy accounts on Instagram which liked each and every photo on Instagram. These users are having user id 5, 14, 21, 24, 36, 41, 54, 57, 66, 71, 75, 76, 91, etc.

**Conclusion:**

The "Instagram User Analytics" project is carried out utilizing SQL queries on MYSQL workbench. This initiative has tracked Instagram users' insights and examined their involvement with the app.

By completing this assignment, I learned about SQL advanced queries and query optimization. I thoroughly comprehended the data and learned how to link it using a join query. I investigated strategies for working with SQL's date type capability. I've learned about window functions and common table expressions(CTE).

**Drive Link:**

<https://drive.google.com/drive/folders/16WQejFng45F-w82ufgMYXXkQveqAv4ur?usp=sharing>