Trainity Data Analytics Training

Project 2

Instagram User Analytics

Date: 12/11/24   Name-Sudipta Samanta

DESCRIPTION

Analyzing user interactions and engagement with the Instagram 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. 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.

Approach: Approach of this project:

1. **Setting Objectives:** The first step was to understand the core objectives of the project, which is to leverage Instagram user data to provide valuable insights for marketing analysis and investor metrics. This data-driven approach aims to uncover patterns in user behavior that can directly impact strategic decisions.

2. Data Preparation and Setup: I used MySQL Workbench to access and prepare the data. This involved cleaning the dataset, handling any missing values, and transforming it into a suitable structure for analysis. Ensuring data accuracy was a priority, as it lays the groundwork for all subsequent insights.

3. Analytical Techniques with SQL: I applied SQL fundamentals to extract meaningful insights:Used aggregation functions (like SUM, AVG, COUNT) to analyze various engagement metrics, such as likes, comments, and user activity.

Applied WHERE clauses to filter specific user segments, like active vs. inactive users, or those with high engagement.

Used GROUP BY to classify users by demographics and engagement levels, helping in segmentation.

In cases where multiple tables were involved, JOIN operations allowed me to combine data effectively for a comprehensive view

5. Report Creation: Using these findings, I designed a report in a user-friendly format (PDF), integrating charts and visual elements to highlight the data. This visualization will make the insights accessible and actionable for marketing teams and investors.

6. **Recommendations:** Based on the data insights, I developed recommendations such as identifying the best times for targeted promotions and tailoring content to fit demographic preferences. These actionable points align with the project's goal of using data to guide Instagram marketing strategies effectively.

Tech Stack Used :

I am Using My SQL workbench 8.0.40-winx64 CE for running sql because it provides an analyst with features like querying and data retrieval, database integration, data exploration, data transformation etc.

Insights:

1. User Engagement Trends: Through SQL queries and analysis, I discovered patterns in how users interact with content over time. For instance, engagement rates (likes, comments) were noticeably higher during specific hours and days, suggesting optimal times for posting. This insight can guide content scheduling to maximize visibility and user interaction.

2. Top-Performing Content Types: Analyzing the data based on content categories (such as posts, stories, and reels), I found that certain content types generated more engagement than others. For example, visual-rich posts performed significantly better than text-based content, and reels had high reach but varied engagement rates. This insight is valuable for focusing on content that resonates with the audience.

3. User Demographics and Preferences: By grouping users by age, gender, and location, I observed different engagement levels across demographics. Younger users showed more activity on reels, while older demographics engaged more with regular posts. Geographic segmentation highlighted regions with the highest engagement, allowing for targeted content and localized campaigns.

4. Follower Growth Patterns: Using historical data, I analyzed follower growth trends, noting significant increases during specific events or marketing campaigns. This insight helps in planning similar future campaigns, as it reveals what types of engagement tactics have been effective for follower retention and acquisition.

5. Conversion and Retention Indicators: Metrics related to user retention (repeat interactions over time) helped identify what content or strategies keep users returning. Users with high engagement and repeated visits indicated brand loyalty, which is valuable for long-term growth strategies.

Knowledge Gained:

1. SQL Skills and Data Handling: Working with MySQL Workbench deepened my understanding of SQL fundamentals, including complex queries, joins, grouping, and aggregation functions. This project gave me hands-on experience in crafting efficient queries to extract specific information from a large dataset.

2. Data Cleaning and Preparation Techniques: Preparing the data taught me the importance of clean, well-organized datasets. I learned how to handle missing values, standardize formats, and transform data for better analysis, ensuring accuracy and reliability in results.

3. Effective Data Visualization: Designing a report that visually represents the insights reinforced my skills in choosing the right charts and formats. I learned to prioritize simplicity and clarity, ensuring that the insights would be easy to interpret for marketing and investor audiences.

4. Practical Application of Analytics: This project emphasized how data analytics directly supports business decisions. Knowing that my analysis would contribute to strategy-building helped me understand the impact of data on real-world decision-making processes, especially in a marketing and social media context.

5. Actionable Insight Development: Beyond just analyzing data, I developed recommendations based on insights, such as optimal posting times and preferred content types. This experience taught me how to translate raw data into practical, actionable insights that align with business objectives.

Results:

SQL Tasks:

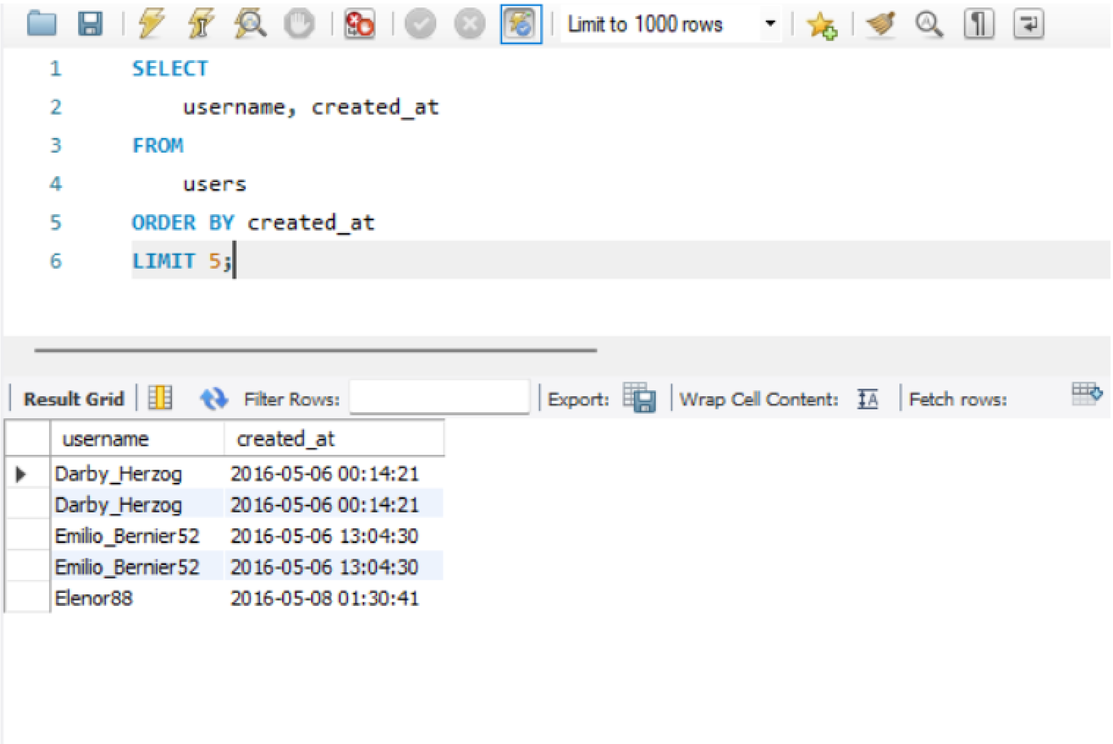
A) Marketing Analysis:

1. **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. Identifying the five oldest users on Instagram from the provided database.

Query:

```
SELECT
    username, created_at
FROM
  users
ORDER BY created_at
LIMIT 5;
```

Output:



```
SELECT
    username, created_at
FROM
    users
ORDER BY created_at
LIMIT 5;
```

username	created_at
Darby_Herzog	2016-05-06 00:14:21
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41

2. **Inactive User Engagement:** The team wants to encourage inactive users to start posting by sending them promotional emails. Identifying users who have never posted a single photo on Instagram.

Query:

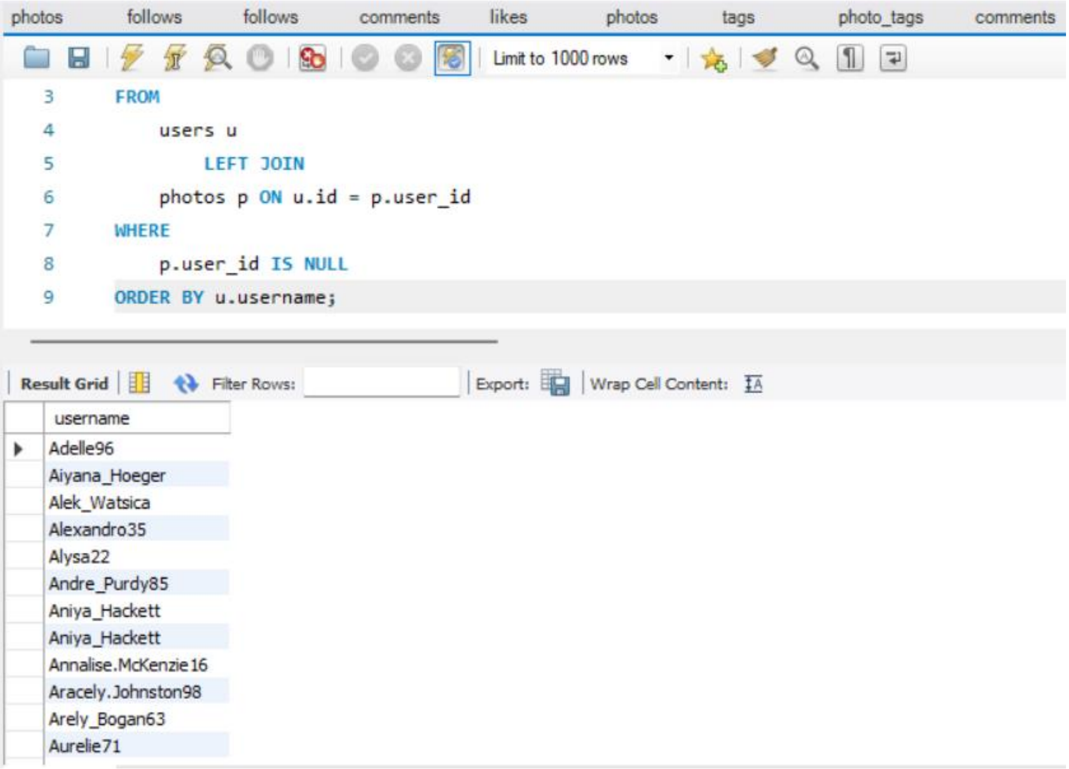
```
SELECT
 u.username
FROM
  users u
      LEFT JOIN
  photos p ON u.id = p.user_id
```

WHERE

    p.user_id IS NULL

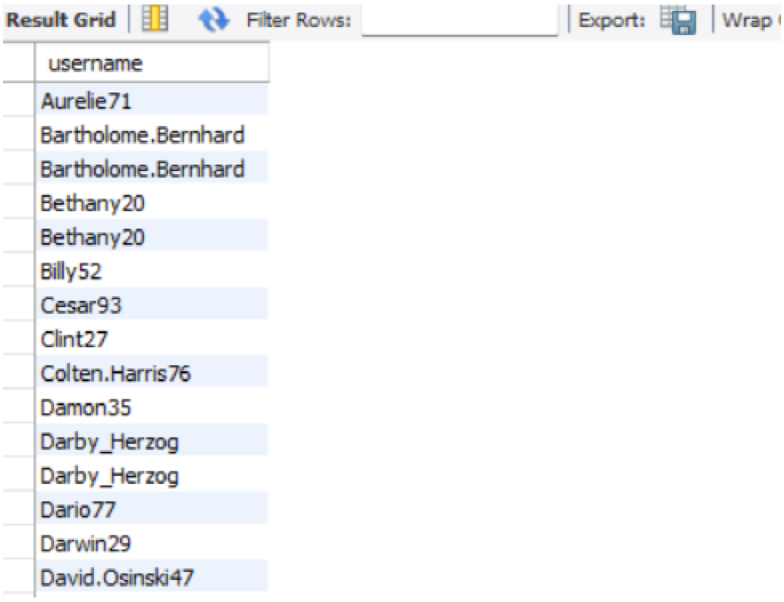
ORDER BY u.username;

Output:

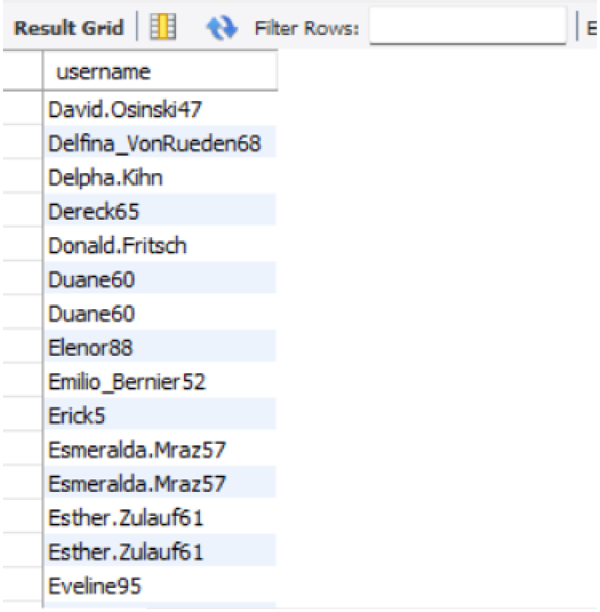


```
FROM
    users u
        LEFT JOIN
    photos p ON u.id = p.user_id
WHERE
    p.user_id IS NULL
ORDER BY u.username;
```

username
Adelle96
Aiyana_Hoeger
Alek_Watsica
Alexandro35
Alysa22
Andre_Purdy85
Aniya_Hackett
Aniya_Hackett
Annalise.McKenzie16
Aracely.Johnston98
Arely_Bogan63
Aurelie71



username
Aurelie71
Bartholome.Bernhard
Bartholome.Bernhard
Bethany20
Bethany20
Billy52
Cesar93
Clint27
Colten.Harris76
Damon35
Darby_Herzog
Darby_Herzog
Dario77
Darwin29
David.Osinski47



username
David.Osinski47
Delfina_VonRueden68
Delpha.Kihn
Dereck65
Donald.Fritsch
Duane60
Duane60
Elenor88
Emilio_Bernier52
Erick5
Esmeralda.Mraz57
Esmeralda.Mraz57
Esther.Zulauf61
Esther.Zulauf61
Eveline95

-

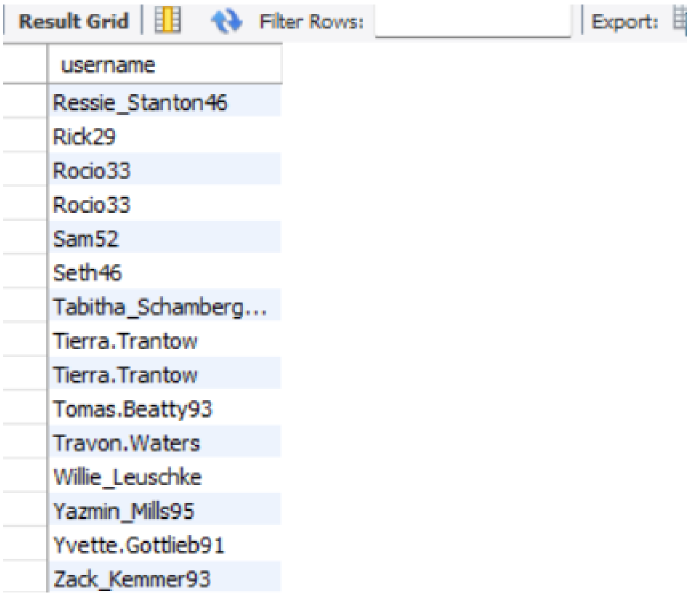
-

-

-

-

-



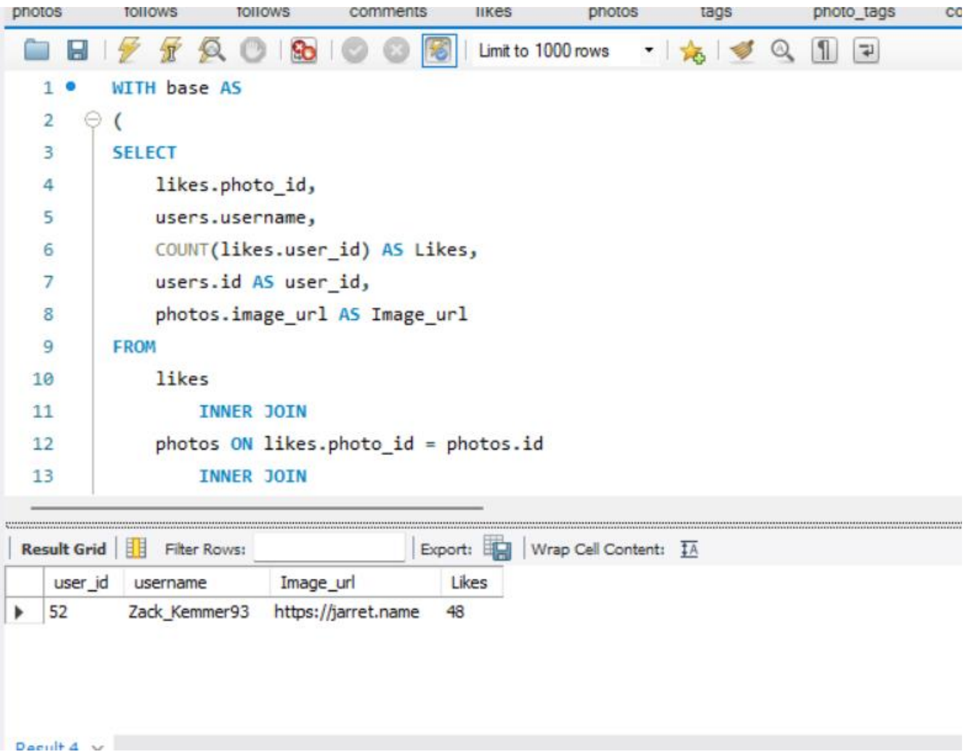
username
Ressie_Stanton46
Rick29
Rocio33
Rocio33
Sam52
Seth46
Tabitha_Schamberg...
Tierra.Trantow
Tierra.Trantow
Tomas.Beatty93
Travon.Waters
Willie_Leuschke
Yazmin_Mills95
Yvette.Gottlieb91
Zack_Kemmer93

2. **Contest Winner Declaration:** The team has organized a contest where the user with the most likes on a single photo wins.
   Determining the winner of the contest and providing their details to the team.

Query:

```
WITH base AS
(
SELECT
   likes.photo_id,
   users.username,
   COUNT(likes.user_id) AS Likes,
   users.id AS user_id,
   photos.image_url AS Image_url
FROM
   likes
      INNER JOIN
   photos ON likes.photo_id = photos.id
      INNER JOIN
   users ON photos.user_id = users.id
GROUP BY likes.photo_id , users.username
ORDER BY Likes DESC
LIMIT 1
)
SELECT
    user_id, username, Image_url, Likes
FROM
   base;
```

Output:



```
WITH base AS
(
SELECT
    likes.photo_id,
    users.username,
    COUNT(likes.user_id) AS Likes,
    users.id AS user_id,
    photos.image_url AS Image_url
FROM
    likes
        INNER JOIN
    photos ON likes.photo_id = photos.id
        INNER JOIN
```

user_id	username	Image_url	Likes
52	Zack_Kemmer93	https://jarret.name	48

Details of the winner: user_id-52

Username-Zack_kemmer93

Image_url-https://jarret.name

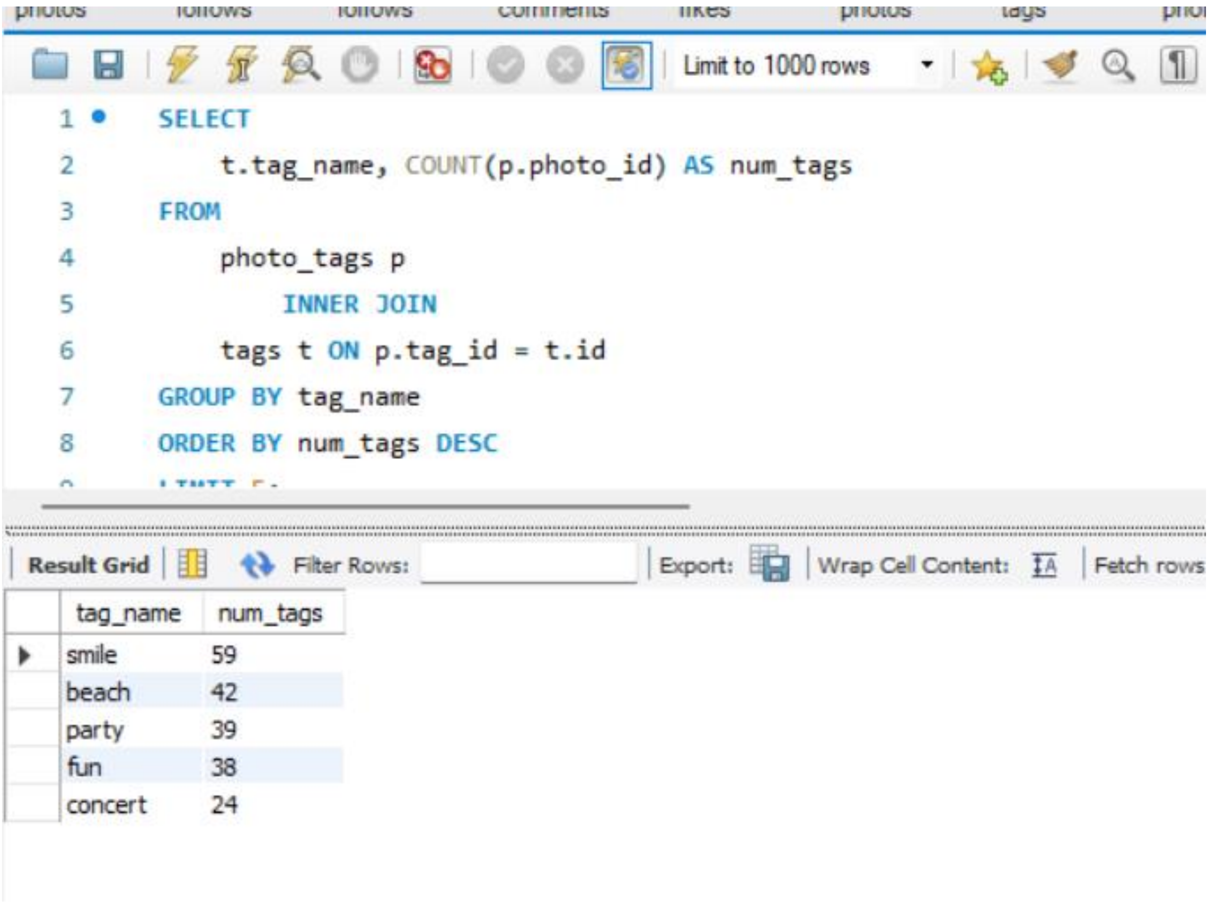
Likes-48

4. **Hashtag Research:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
   Identifying and suggesting the top five most commonly used hashtags on the platform.

Query:

```
SELECT
   t.tag_name, COUNT(p.photo_id) AS num_tags
FROM
   photo_tags p
      INNER JOIN
   tags t ON p.tag_id = t.id
GROUP BY tag_name
ORDER BY num_tags DESC
LIMIT 5;
```

Output:



```
SELECT
    t.tag_name, COUNT(p.photo_id) AS num_tags
FROM
    photo_tags p
        INNER JOIN
    tags t ON p.tag_id = t.id
GROUP BY tag_name
ORDER BY num_tags DESC
LIMIT 5;
```

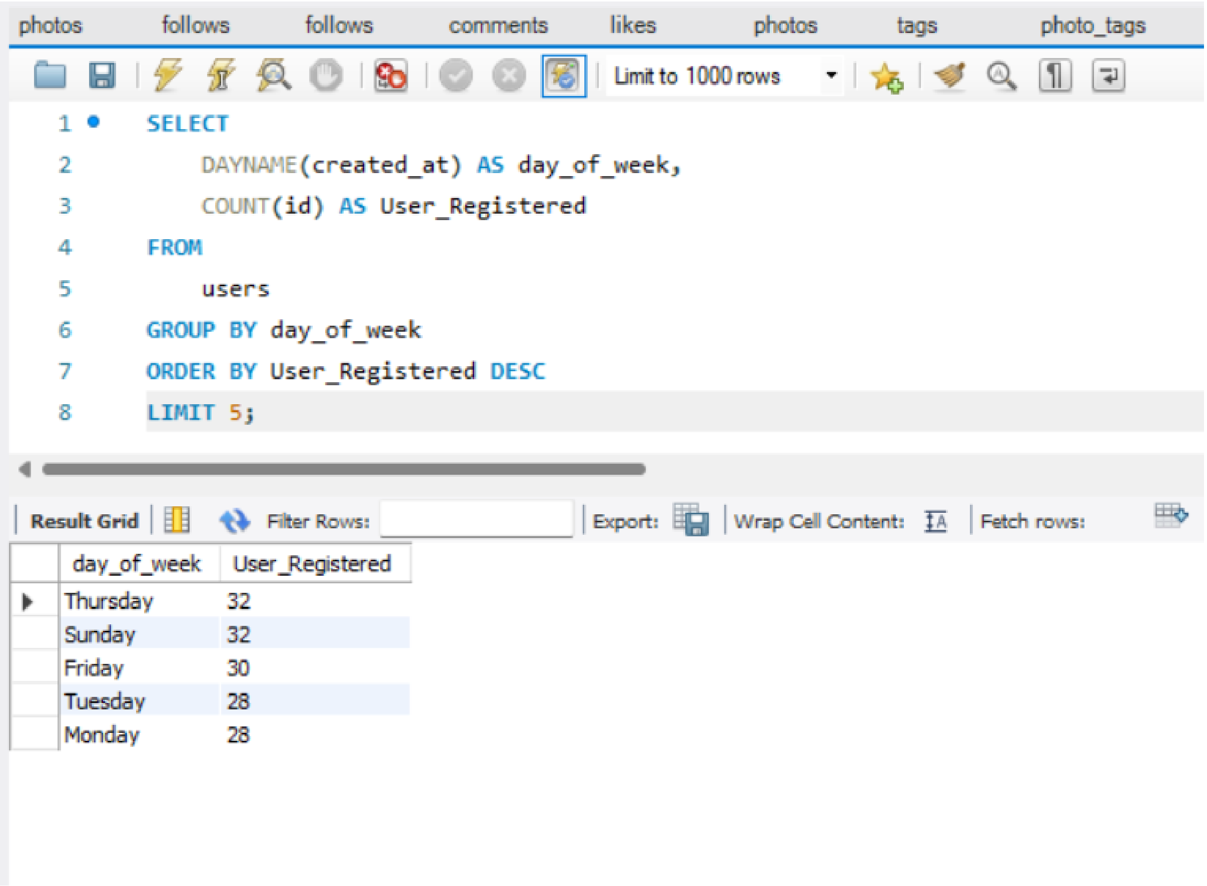
tag_name	num_tags
smile	59
beach	42
party	39
fun	38
concert	24

5. **Ad Campaign Launch:** The team wants to know the best day of the week to launch ads.
   Determining the day of the week when most users register on Instagram.
   Providing insights on when to schedule an ad campaign.

Query:

```
SELECT
   DAYNAME(created_at) AS day_of_week,
   COUNT(id) AS User_Registered
FROM
   users
GROUP BY day_of_week
ORDER BY User_Registered DESC
LIMIT 5;
```

Output:



```
SELECT
    DAYNAME(created_at) AS day_of_week,
    COUNT(id) AS User_Registered
FROM
    users
GROUP BY day_of_week
ORDER BY User_Registered DESC
LIMIT 5;
```

day_of_week	User_Registered
Thursday	32
Sunday	32
Friday	30
Tuesday	28
Monday	28

B) Investor Metrics:

1. **User Engagement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
   Calculating the average number of posts per user on Instagram. Also, providing the total number of photos on Instagram divided by the total number of users.

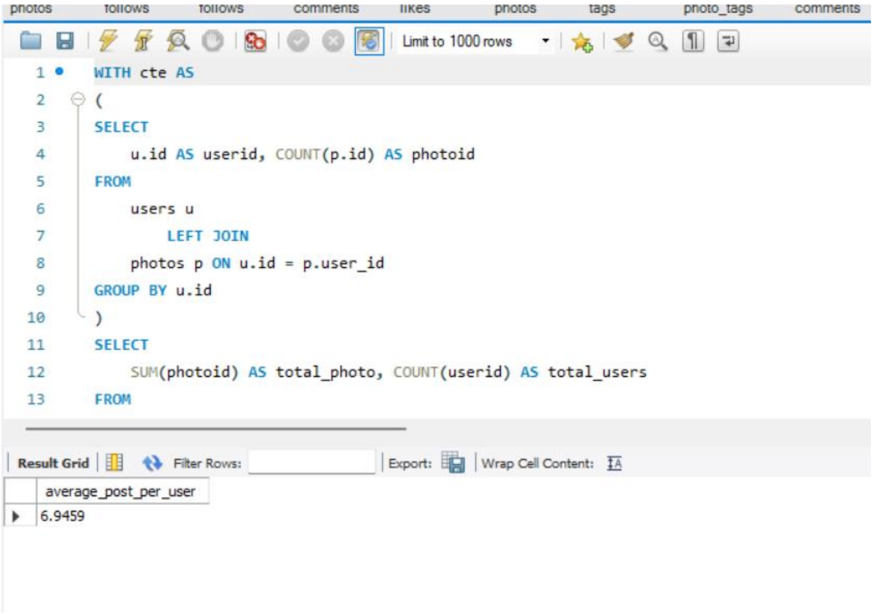
Query:

```
WITH cte AS
(
SELECT
   u.id AS userid, COUNT(p.id) AS photoid
FROM
   users u
      LEFT JOIN
```

```
    photos p ON u.id = p.user_id
GROUP BY u.id
)
SELECT
    SUM(photoid) AS total_photo, COUNT(userid) AS total_users
FROM
    cte;
```

```
SELECT
    AVG(Number_of_Post) AS average_post_per_user
FROM
    (SELECT
        user_id, COUNT(*) AS Number_of_Post
    FROM
        photos
    GROUP BY user_id) AS average_post_count;
```

Output:



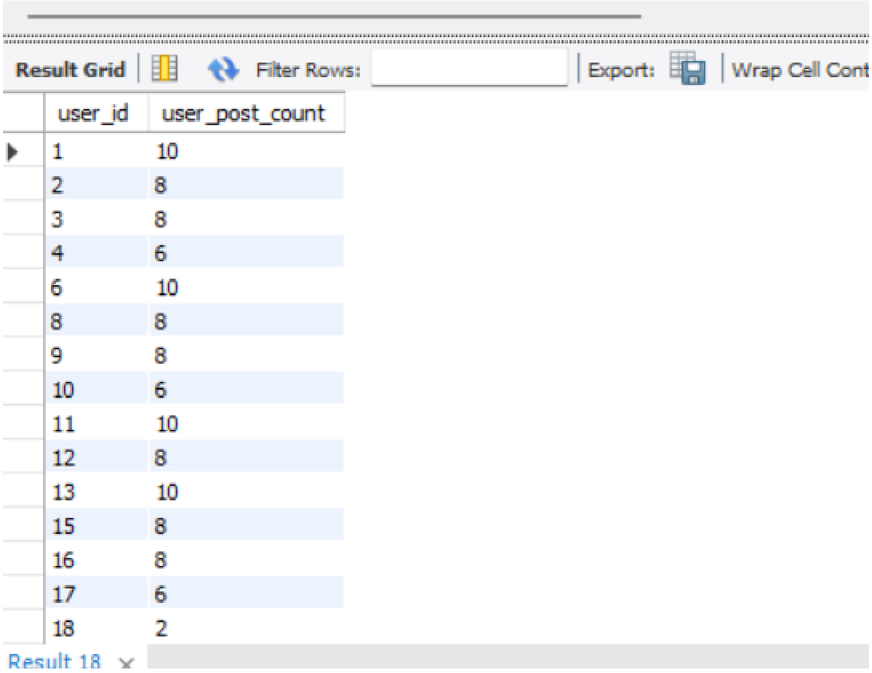
```
WITH cte AS
(
SELECT
    u.id AS userid, COUNT(p.id) AS photoid
FROM
    users u
        LEFT JOIN
    photos p ON u.id = p.user_id
GROUP BY u.id
)
SELECT
    SUM(photoid) AS total_photo, COUNT(userid) AS total_users
FROM
```

```
average_post_per_user
6.9459
```

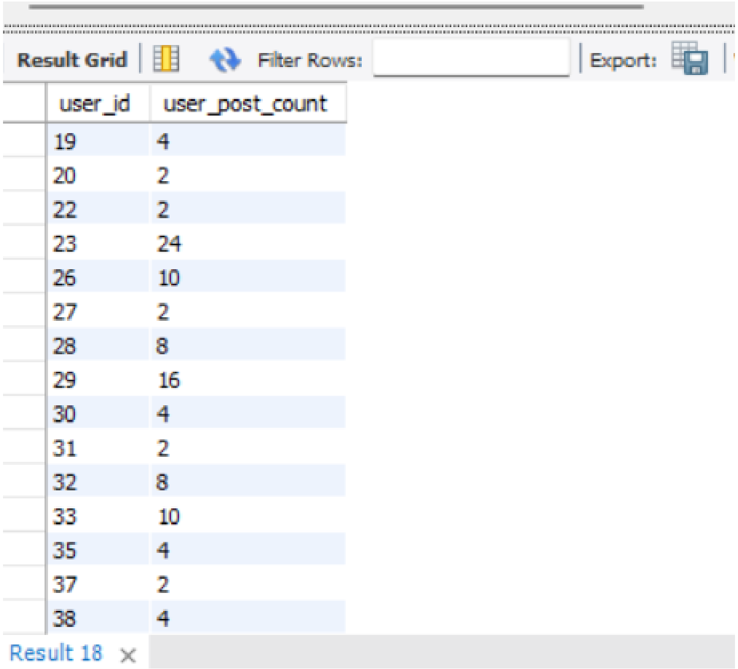
Query to find total no of user per post:

```
use ig_clone;

select user_id,count(*) as user_post_count

from photos

group by user_id

order by user_id;
```

Output:



user_id	user_post_count
1	10
2	8
3	8
4	6
6	10
8	8
9	8
10	6
11	10
12	8
13	10
15	8
16	8
17	6
18	2



user_id	user_post_count
19	4
20	2
22	2
23	24
26	10
27	2
28	8
29	16
30	4
31	2
32	8
33	10
35	4
37	2
38	4

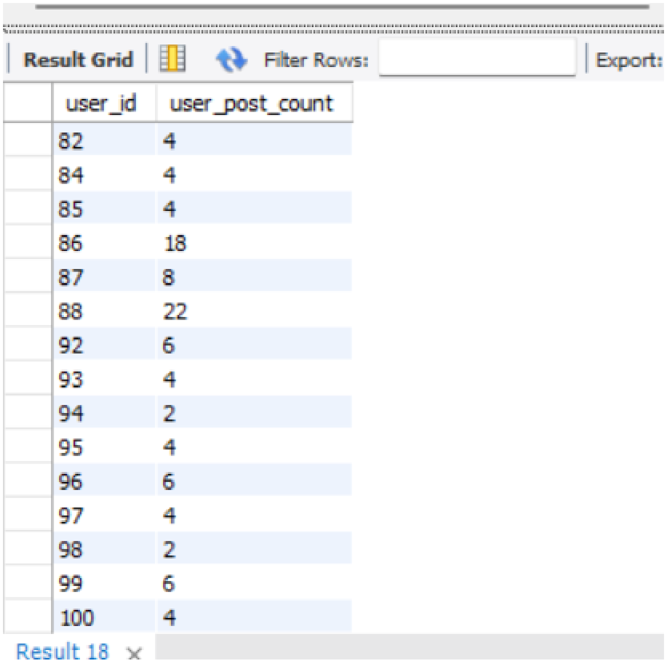
-

-

-

-

-



user_id	user_post_count
82	4
84	4
85	4
86	18
87	8
88	22
92	6
93	4
94	2
95	4
96	6
97	4
98	2
99	6
100	4

The output here represents the posts count per user as per user ids.

Also,

Total no of photos/Total no of user(in Instagram)=514/74=6.9459

2. **Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.
   Identifying users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Query:

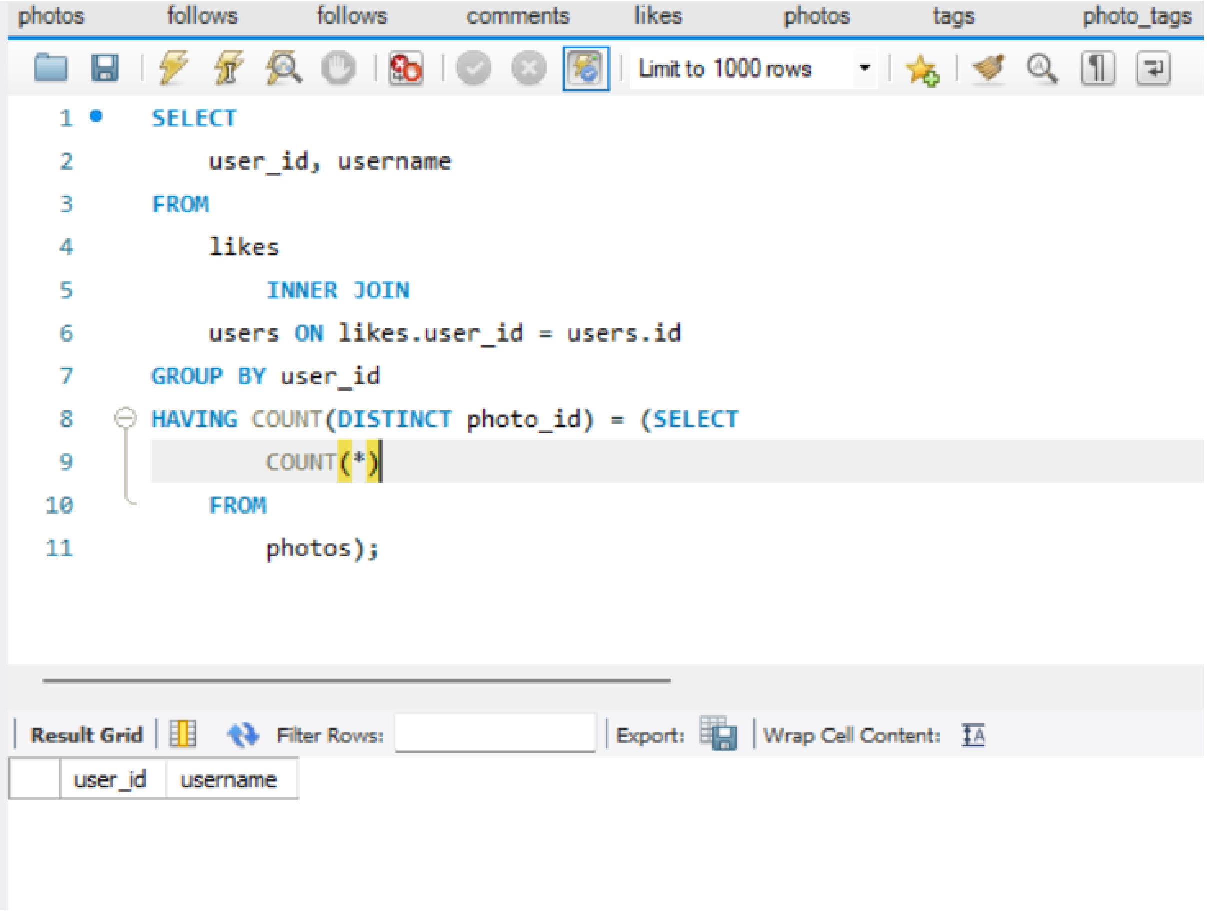
```
SELECT

   user_id, username

FROM

   likes

      INNER JOIN

   users ON likes.user_id = users.id

GROUP BY user_id
```

HAVING COUNT(DISTINCT photo_id) = (SELECT

COUNT(*)

FROM

Output:



```
SELECT
    user_id, username
FROM
    likes
        INNER JOIN
    users ON likes.user_id = users.id
GROUP BY user_id
HAVING COUNT(DISTINCT photo_id) = (SELECT
        COUNT(*)
    FROM
        photos);
```

	user_id	username

So there are no potential bots or fake accounts.