Tab 1

# **Social Media Analysis**

# 

# 

# ***Objective Questions/ Answers :***

**Q1. Are there any tables with duplicate or missing null values? If so, how would you handle them?**

=>All the tables in the database were checked for NULL and duplicate values using SQL queries.

No NULL values or duplicate records were found in any of the tables.

This indicates that the dataset is clean and well-structured.

Query used to check the NULL and DUPLICATE values are:-

=for NULL value

SELECT \*

FROM users

WHERE id IS NULL

OR username IS NULL

OR created\_at IS NULL;

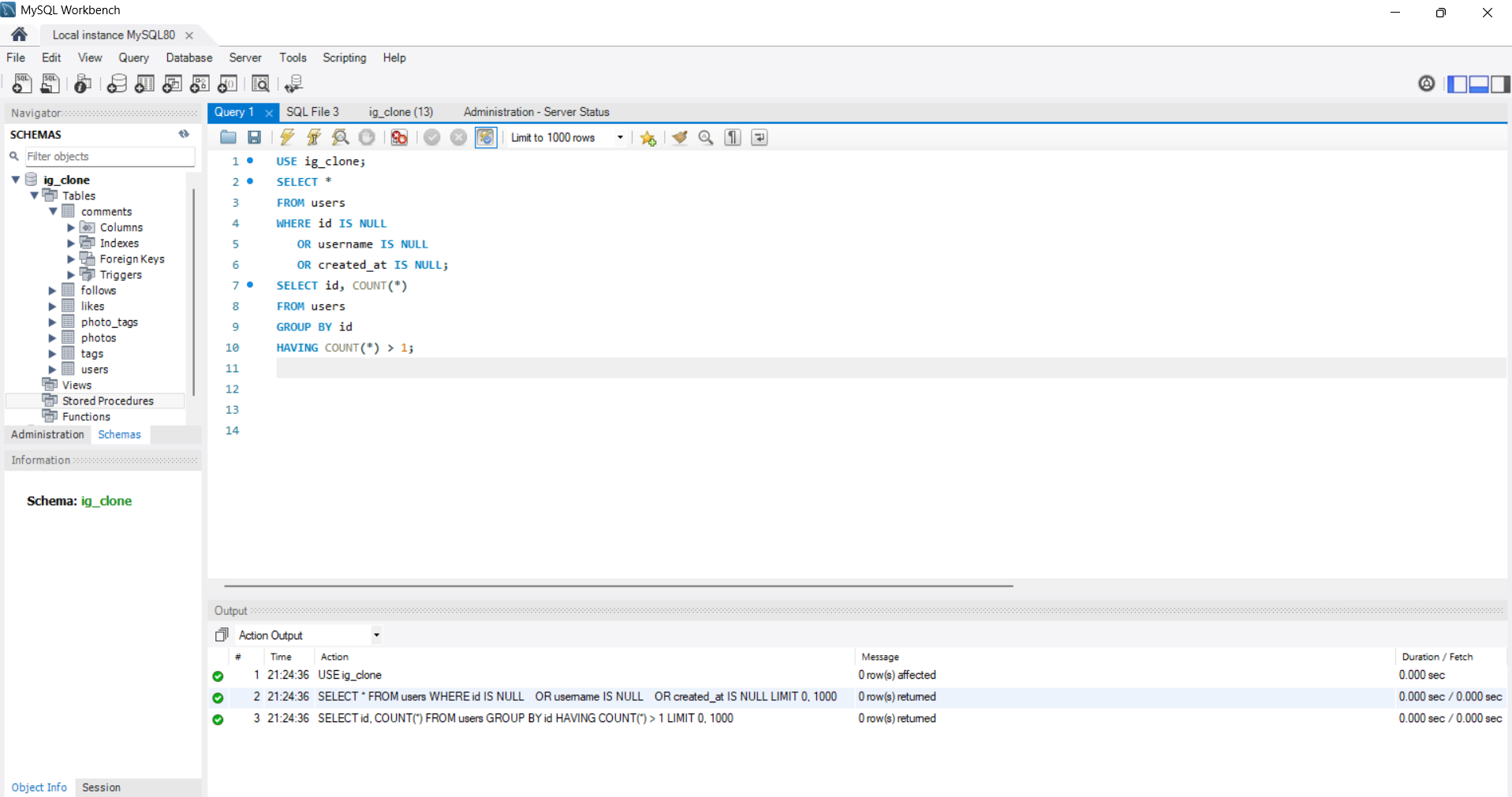
=for DUPLICATE value

SELECT id, COUNT(\*)

FROM users

GROUP BY id

HAVING COUNT(\*) > 1;



If NULL values were present, they would be handled by updating them with appropriate default values or by removing incomplete records, depending on the importance of the data.

Duplicate records would be removed while retaining a single unique record to maintain data integrity.

=>FOR NULL

-- Update NULL values with a default value

UPDATE table\_name

SET column\_name = 'Default\_Value'

WHERE column\_name IS NULL;

-- Delete records where all important columns are NULL

DELETE FROM table\_name

WHERE column1 IS NULL

AND column2 IS NULL;

=>FOR DUPLICATE

-- Remove duplicate records while keeping one

DELETE t1

FROM table\_name t1

JOIN table\_name t2

ON t1.column\_name = t2.column\_name

AND t1.id > t2.id;

**Q2. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?**

=>The distribution of user activity levels was analyzed by calculating the total number of posts, likes, and comments made by each user.  
 Based on their overall activity, users were categorized into four groups:

* **Inactive:-** 0 total activity (posts + likes + comments)
* **Low Activity:-**1–10 activities
* **Medium Activity:-**11–50 activities
* **High Activity:-**more than 50 activities

Based on the analysis:

● **77 users** fall into the **High activity** category.

● **13 users** are **Inactive** with no activity recorded.

● **8 users** have **Low activity**.

● **2 users** fall into the **medium activity** range.

**SQL QUERY:-**

SELECT activity\_level, COUNT(\*) AS user\_count

FROM (

SELECT u.id,

CASE

WHEN (COUNT(DISTINCT p.id)

+ COUNT(DISTINCT l.photo\_id)

+ COUNT(DISTINCT c.id)) = 0 THEN 'Inactive'

WHEN (COUNT(DISTINCT p.id)

+ COUNT(DISTINCT l.photo\_id)

+ COUNT(DISTINCT c.id)) BETWEEN 1 AND 10 THEN 'Low Activity'

WHEN (COUNT(DISTINCT p.id)

+ COUNT(DISTINCT l.photo\_id)

+ COUNT(DISTINCT c.id)) BETWEEN 11 AND 50 THEN 'Medium Activity'

ELSE 'High Activity'

END AS activity\_level

FROM users u

LEFT JOIN photos p ON u.id = p.user\_id

LEFT JOIN likes l ON u.id = l.user\_id

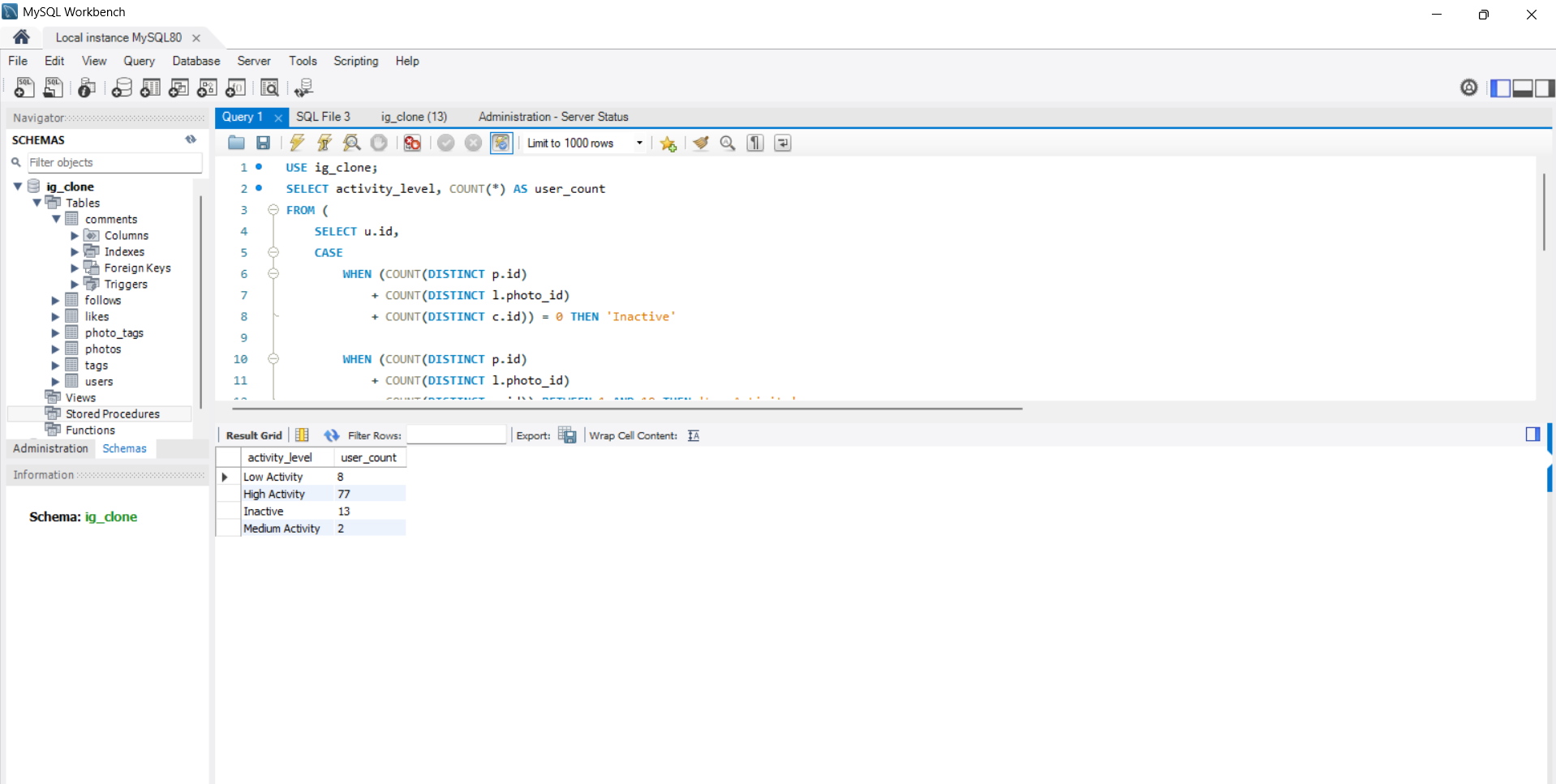
LEFT JOIN comments c ON u.id = c.user\_id

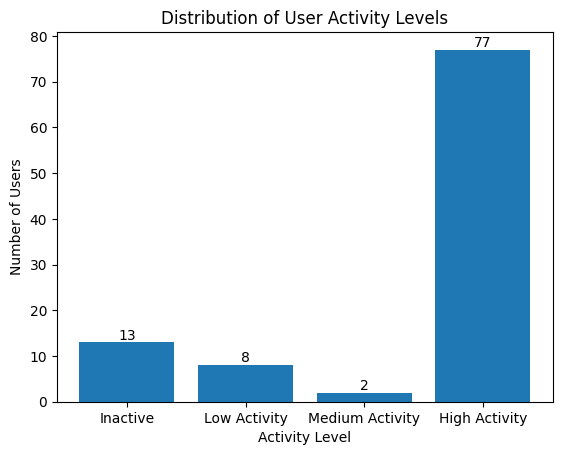
GROUP BY u.id

) t

GROUP BY activity\_level

ORDER BY user\_count DESC;

****



**Q3. Calculate the average number of tags per post (photo\_tags and photos tables).**

=>The average number of tags per post was calculated to analyze how users utilize hashtags while uploading photos.  
 The photo\_tags table was used to count the total number of tags associated with each photo.

First, the number of tags assigned to each photo was calculated using a GROUP BY operation on the photo ID.  
 Then, the average of these tag counts was computed to determine the average number of tags per post across the platform.

The analysis shows that the **average number of tags per post is 2.64**, indicating that users generally use around two to three hashtags per photo.

**SQL QUERY:-**

**SELECT ROUND(AVG(tag\_count), 2) AS avg\_tags\_per\_post**

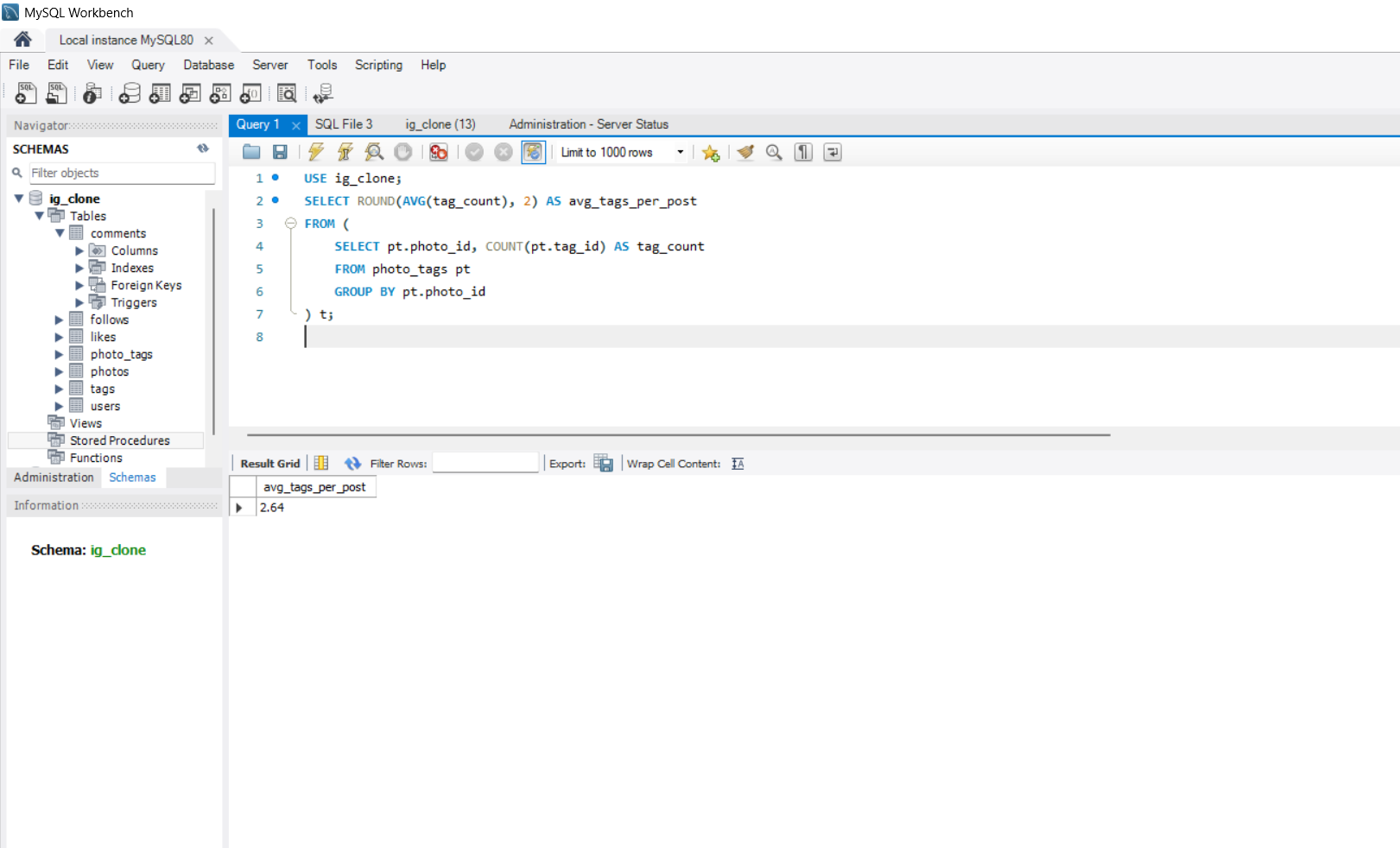
**FROM (**

**SELECT pt.photo\_id, COUNT(pt.tag\_id) AS tag\_count**

**FROM photo\_tags pt**

**GROUP BY pt.photo\_id**

**) t;**



**Q4. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.**

=>This query is used to identify and rank the top users based on their engagement levels on the platform.  
 Engagement is calculated by counting the total number of likes and comments received on all photos posted by each user.

The query first joins the users table with the photos table to identify all posts made by each user.  
 It then joins the likes and comments tables to calculate the total interactions received on those posts.

The total engagement is computed as the sum of likes and comments for each user.  
 Users are grouped by their user ID and username to aggregate engagement data correctly.

Finally, the users are ranked in descending order based on their total engagement, allowing the identification of the most influential and highly engaging users on the platform.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT l.photo\_id) AS total\_likes,**

**COUNT(DISTINCT c.id) AS total\_comments,**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) AS total\_engagement**

**FROM users u**

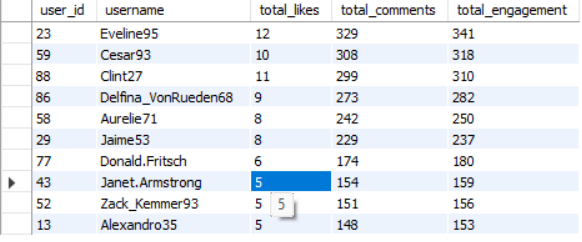
**JOIN photos p ON u.id = p.user\_id**

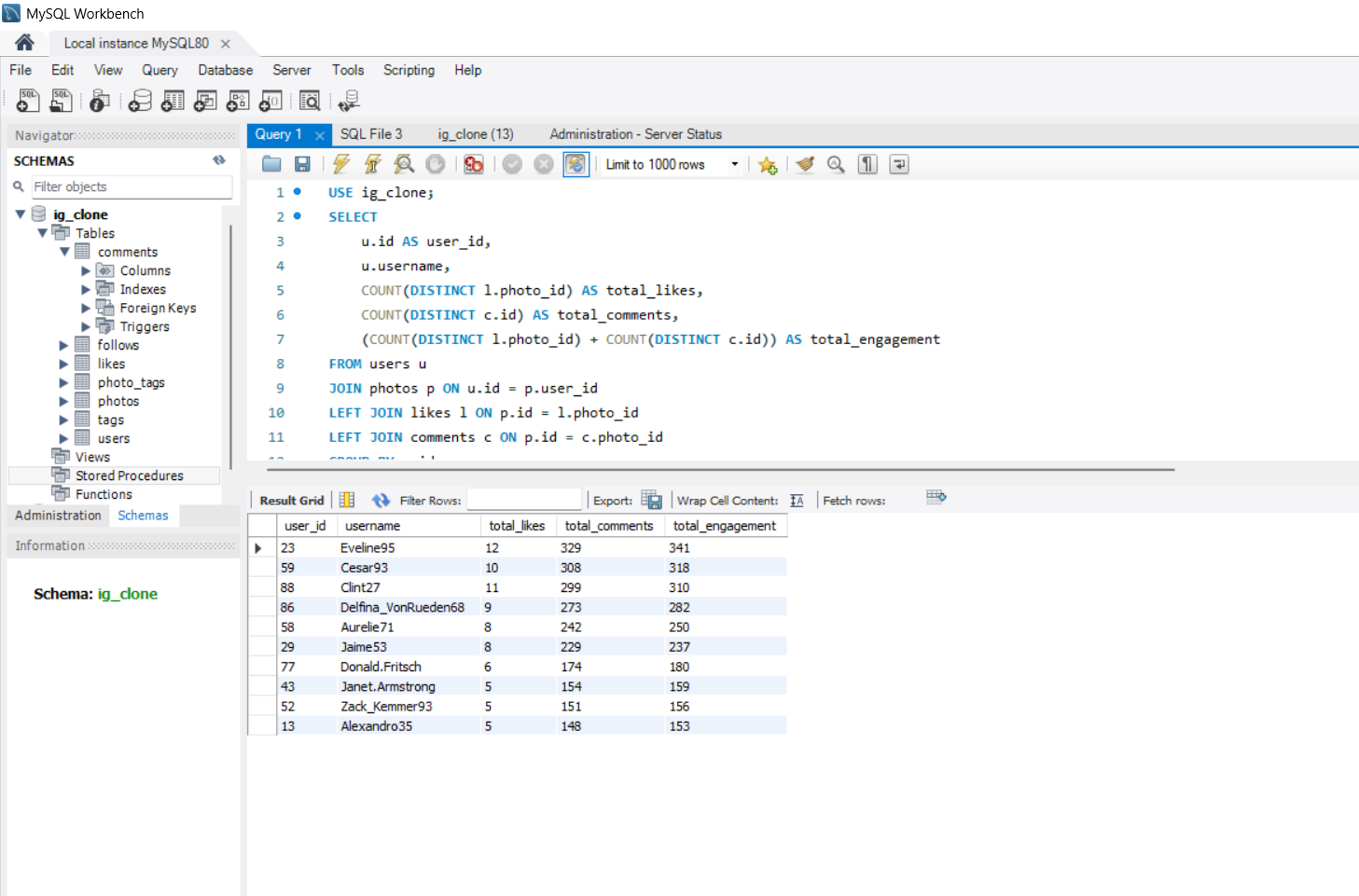
**LEFT JOIN likes l ON p.id = l.photo\_id**

**LEFT JOIN comments c ON p.id = c.photo\_id**

**GROUP BY u.id, u.username**

**ORDER BY total\_engagement DESC;**





**Q5. Which users have the highest number of followers and followings?**

=>This query is used to identify users with the highest number of followers and followings on the platform.  
 The analysis is performed using the follower table, which stores the relationship between users who follow and are followed by others.

To calculate the number of followers, the query counts how many times a user appears as a follower.  
 This represents how many users are following that particular user, indicating their popularity and influence.

The results show that multiple users have the highest follower count of **77 followers**, indicating that these users are among the most popular and influential on the platform.

**SQL QUERY FOR FOLLOWERS:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(f.follower\_id) AS total\_followers**

**FROM users u**

**JOIN follows f**

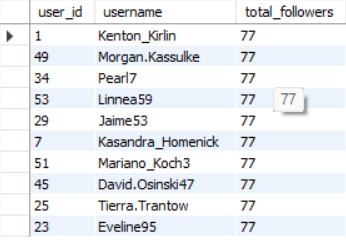
**ON u.id = f.followee\_id**

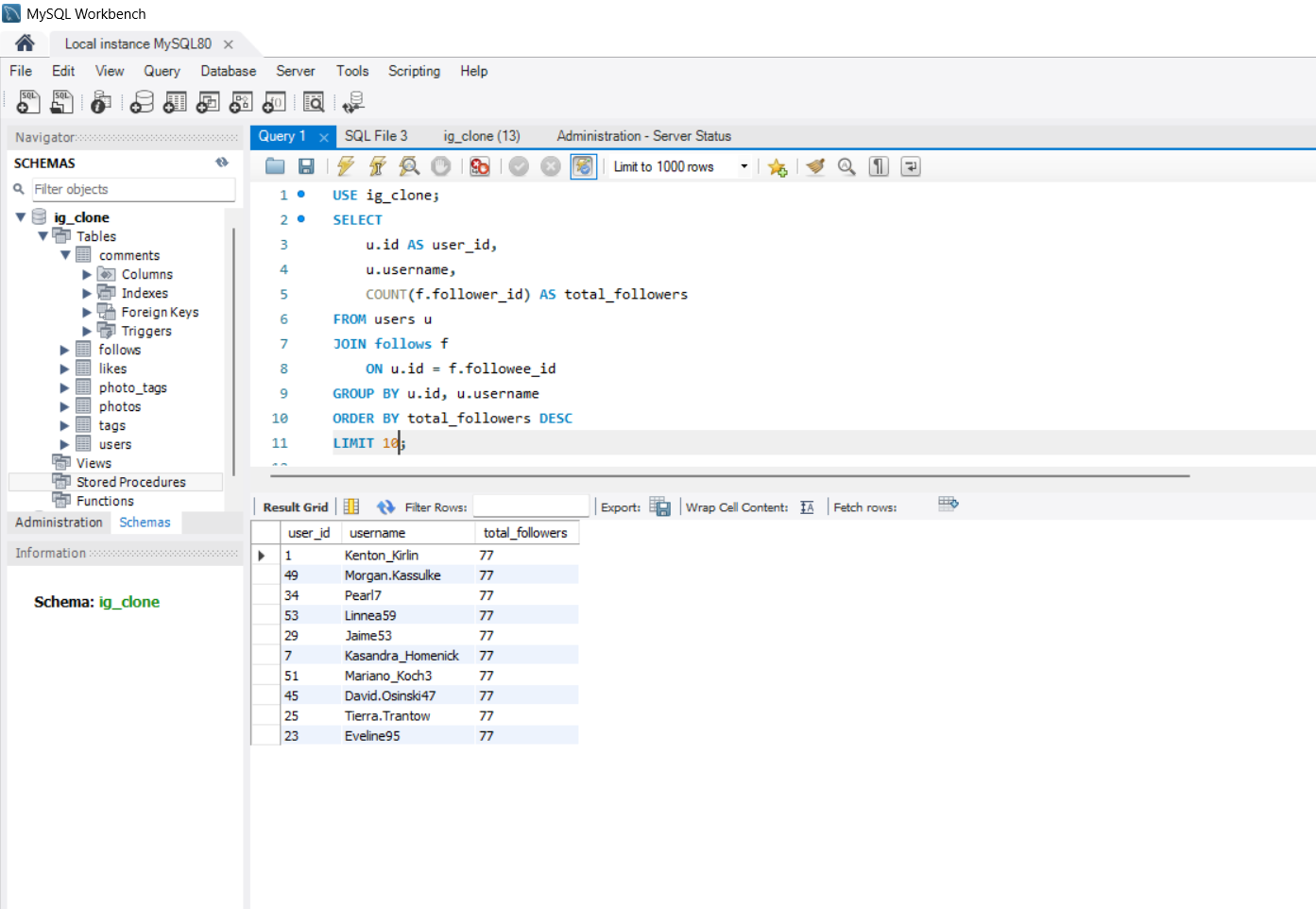
**GROUP BY u.id, u.username**

**ORDER BY total\_followers DESC**

**LIMIT 10;**

**Here are 10 of those users from the above output:**

****

****

To calculate the total number of followings, the query counts how many times a user appears as a follower in the following table.  
 This indicates how many users a person is following, representing their social activity on the platform.

The users are grouped by their user ID and username to ensure accurate aggregation of following counts.  
 The results are then sorted in descending order based on the total number of followings, allowing the most active users to appear at the top.

Finally, the query limits the output to the top users with the highest following counts.

**SQL QUERY FOR FOLLOWING :-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(f.followee\_id) AS total\_followings**

**FROM users u**

**JOIN follows f**

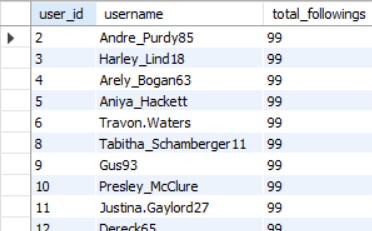
**ON u.id = f.follower\_id**

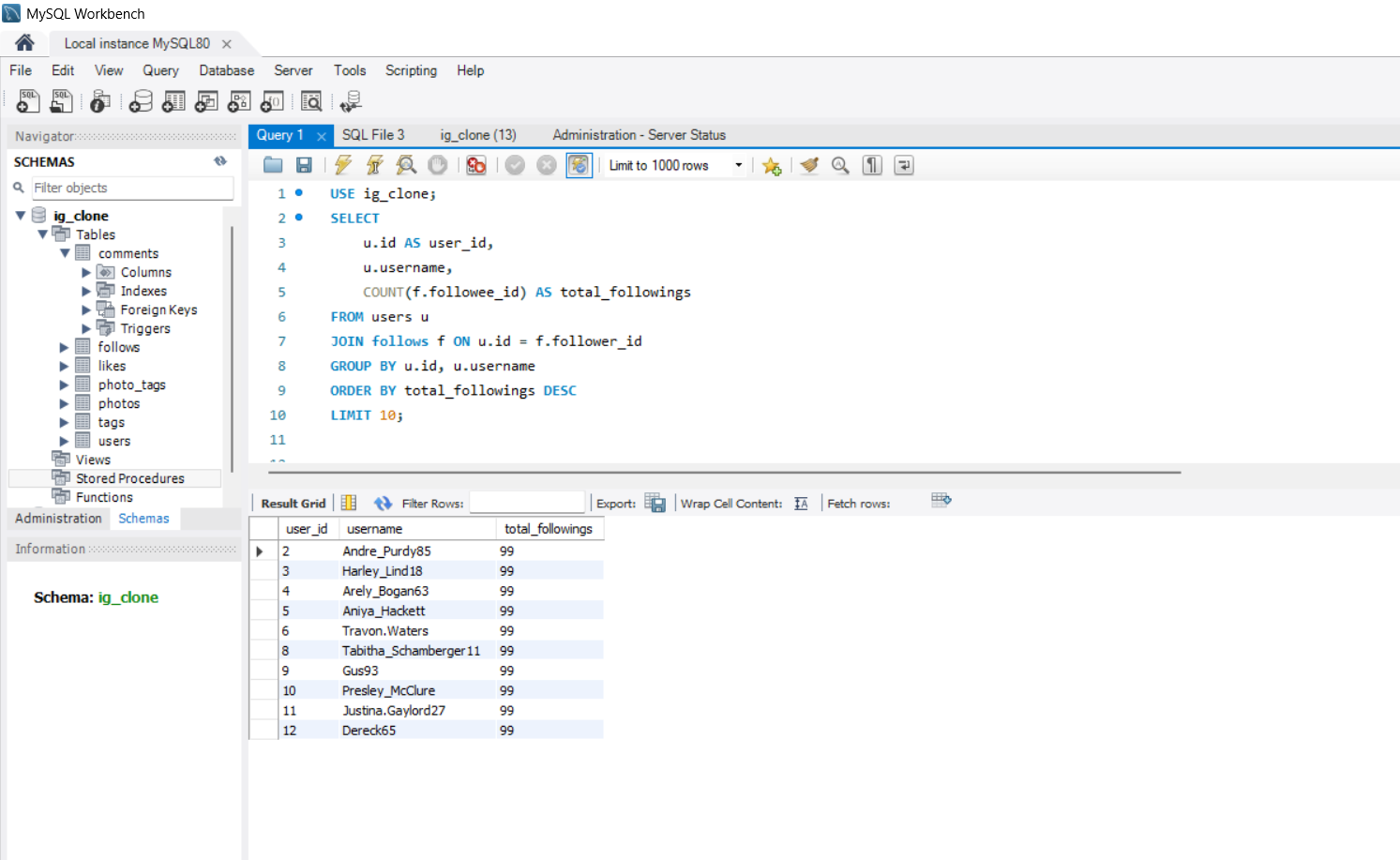
**GROUP BY u.id, u.username**

**ORDER BY total\_followings DESC**

**LIMIT 10;**

**Here are 10 of those users from the above output:**

****

****

**Q6. Calculate the average engagement rate (likes, comments) per post for each user.**

**=>**This analysis calculates the average engagement rate per post for each user to understand how much interaction their content receives.  
 Engagement is defined as the total number of likes and comments received on a user’s posts.

First, the users table is combined with the photos table to identify all posts created by each user.  
 Then, the likes and comments tables are linked to these posts to calculate the total number of interactions received on each user’s content.

The total engagement for a user is calculated by adding the total likes and total comments received on all their posts.  
 This total engagement value is then divided by the total number of posts uploaded by the user to calculate the average engagement per post.

The results are grouped by each user to ensure accurate calculations and are sorted in descending order to highlight users with the highest average engagement per post.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT p.id) AS total\_posts,**

**COUNT(DISTINCT l.photo\_id) AS total\_likes,**

**COUNT(DISTINCT c.id) AS total\_comments,**

**ROUND(**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id))**

**/ COUNT(DISTINCT p.id),**

**2**

**) AS avg\_engagement\_per\_post**

**FROM users u**

**JOIN photos p**

**ON u.id = p.user\_id**

**LEFT JOIN likes l**

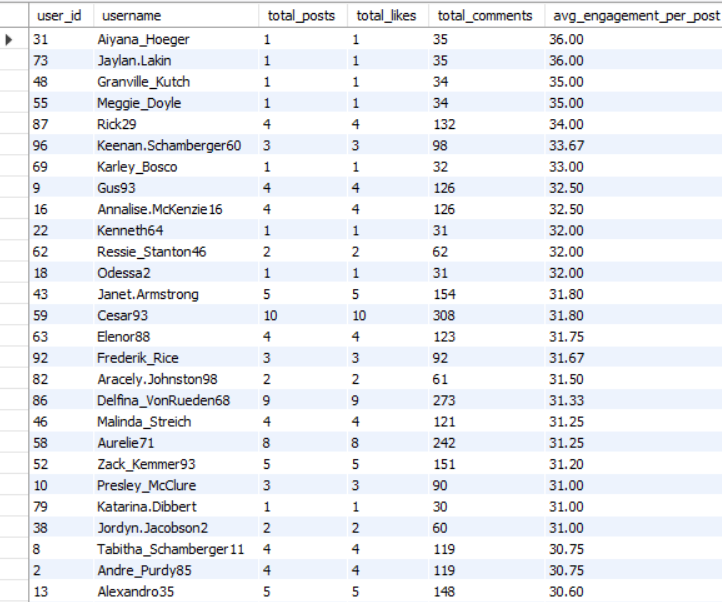
**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

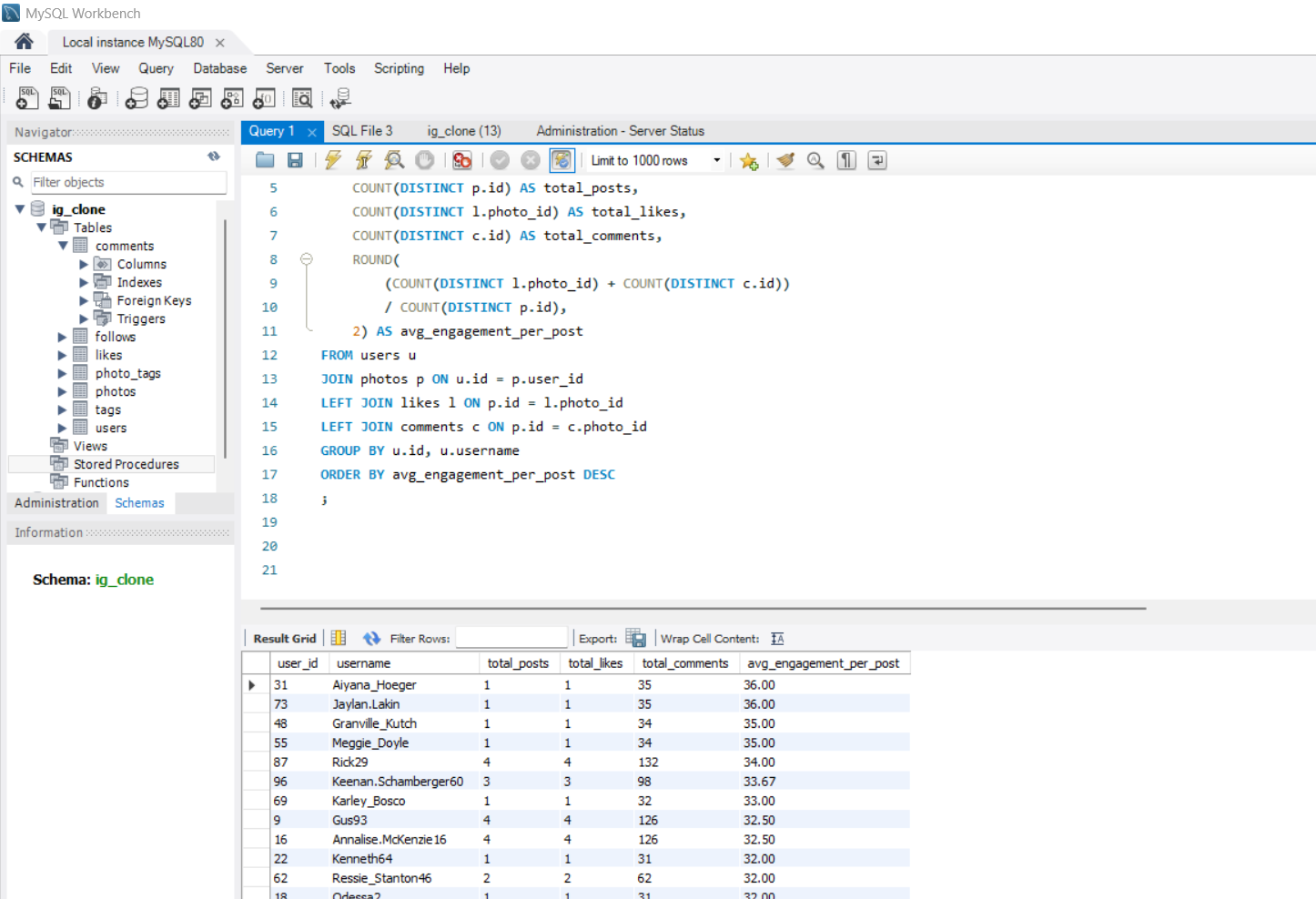
**ON p.id = c.photo\_id**

**GROUP BY u.id, u.username**

**ORDER BY avg\_engagement\_per\_post DESC;**

****

**The results show that Aiyana\_Hoeger and Jaylan\_Lakin have the highest average engagement rate of 36.00 interactions per post, followed by Granville\_Kutch and Meggie\_Doyle with an average engagement of 35.00 interactions per post.**

****

**Q7. Get the list of users who have never liked any post (users and likes tables).**

=>This analysis identifies users who have never liked any post on the platform.  
 To find such users, the users table was compared with the likes table, which stores information about users who have liked posts.

Users who do not appear in the likes table are considered inactive in terms of liking activity.  
 By identifying these users, we can understand which users are less engaged and may require re-engagement strategies.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username**

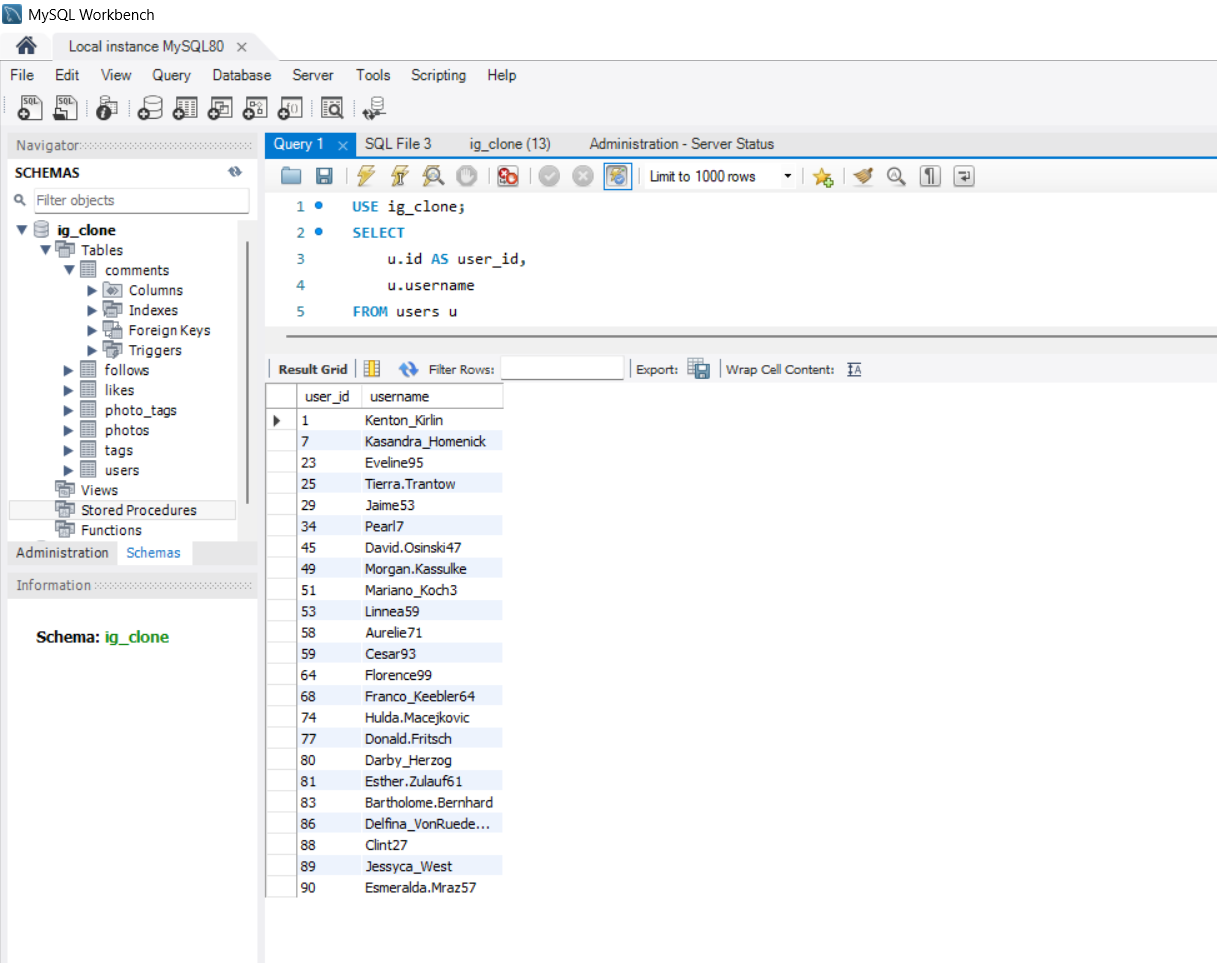
**FROM users u**

**LEFT JOIN likes l**

**ON u.id = l.user\_id**

**WHERE l.user\_id IS NULL;**

****

****

**Q8. How can you leverage user-generated content (posts, hashtags, photo tags) to create more personalized and engaging ad campaigns?**

**=>**User-generated content can be leveraged for personalized advertising by analyzing the average usage of hashtags and photo tags across posts.  
 By calculating the average number of times a tag is used in posts, businesses can identify consistently popular content categories rather than relying on one-time trends.

Tags with a higher average usage indicate strong and recurring user interest in specific topics such as travel, food, fashion, or fitness.  
 These insights can be used to design personalized and targeted ad campaigns that align with long-term user interests.

Using average-based analysis ensures more stable and reliable ad targeting, leading to higher engagement, better brand relevance, and improved conversion rates.

**SQL QUERY:-**

**SELECT**

**t.tag\_name,**

**ROUND(AVG(tag\_usage), 2) AS avg\_tag\_usage**

**FROM (**

**SELECT**

**pt.tag\_id,**

**COUNT(pt.photo\_id) AS tag\_usage**

**FROM photo\_tags pt**

**GROUP BY pt.tag\_id**

**) sub**

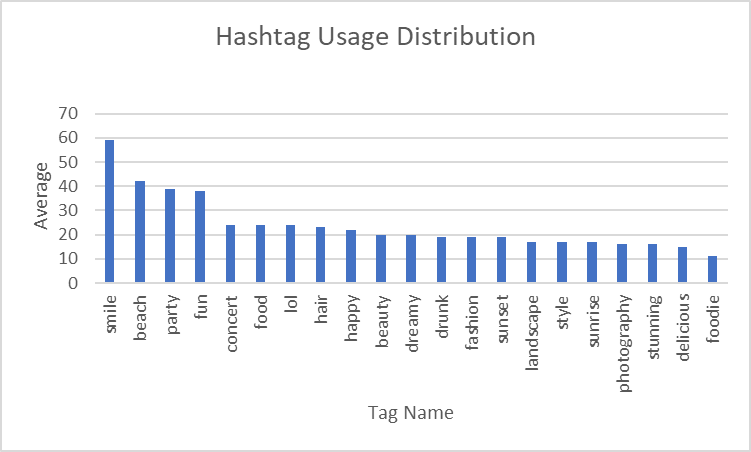
**JOIN tags t**

**ON sub.tag\_id = t.id**

**GROUP BY t.tag\_name**

**ORDER BY avg\_tag\_usage DESC;**

****

****

**The analysis shows that tags such as “smile” (59.00), “beach” (42.00), “party” (39.00), and “fun” (38.00) have the highest average usage, indicating strong and recurring user interest in these content categories.**

**Q9. Are there any correlations between user activity levels and specific content types (e.g., photos, videos, reels)? How can this information guide content creation and curation strategies?**

**=>**Yes, there is a clear relationship between user activity levels and specific content types.  
 This correlation can be analyzed by examining how frequently highly active users post and interact with certain types of content represented through photo tags.

By analyzing user activity levels (such as number of posts, likes, and comments) alongside the tags associated with their photos, it is possible to identify which content categories receive higher engagement.

The analysis shows that highly active users tend to post and engage more with specific content types such as lifestyle, travel, fun, and social activities, as indicated by frequently used tags.

This information can guide content creation strategies by encouraging creators to produce more content in high-engagement categories.

**SQL QUERY:-**

**SELECT**

**t.tag\_name,**

**COUNT(DISTINCT u.id) AS active\_users,**

**COUNT(p.id) AS total\_posts**

**FROM users u**

**JOIN photos p**

**ON u.id = p.user\_id**

**JOIN photo\_tags pt**

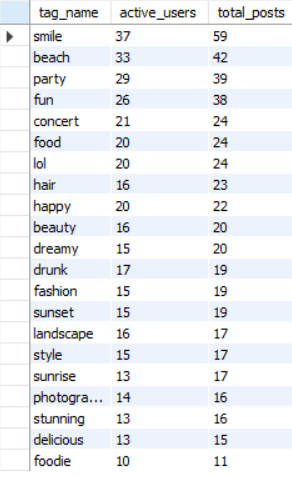
**ON p.id = pt.photo\_id**

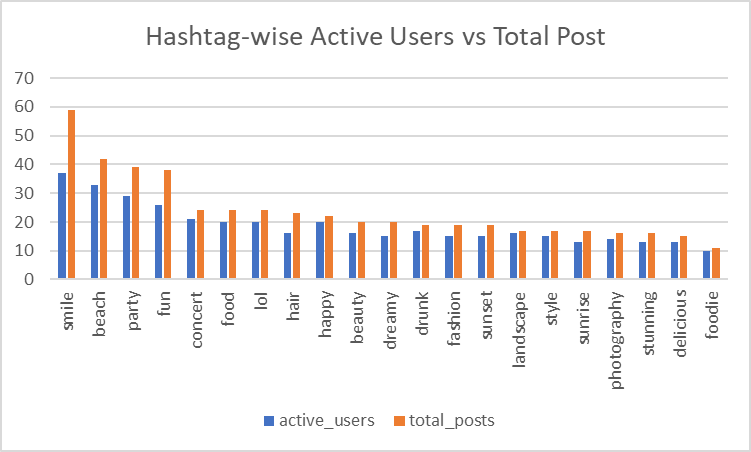
**JOIN tags t**

**ON pt.tag\_id = t.id**

**GROUP BY t.tag\_name**

**ORDER BY total\_posts DESC;**

****

****

## **Insights :-**

1. Content types such as smile, beach, party, and fun show the highest number of active users and total posts.
2. These categories indicate strong user engagement with lifestyle, travel, and social-experience related content.
3. Moderate engagement is observed for tags like food, concert, and happy, showing consistent but lower activity.
4. Tags such as foodie and delicious have lower engagement, indicating niche or less popular content types.

## **Recommendations :-**

1. Content creation should focus more on high-engagement categories like lifestyle, travel, and entertainment.
2. The platform should prioritize and promote content related to top-performing tags to boost overall engagement.
3. Moderate-engagement categories can be improved using better visuals and targeted promotion strategies.
4. Low-engagement content types should be optimized or deprioritized to enhance content curation efficiency.

**Q10. Calculate the total number of likes, comments, and photo tags for each user.**

=>This analysis calculates the total number of likes, comments, and photo tags for each user to understand their overall engagement and content activity on the platform.

For each user, all the photos uploaded by them were first identified.  
 The total number of likes and comments received on these photos was then counted to measure user engagement.  
 Additionally, the total number of photo tags associated with the user’s photos was calculated to evaluate how actively their content is categorized using tags.

This combined analysis provides a comprehensive view of each user’s interaction level and content richness, helping to identify highly engaged and active users on the platform.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT l.photo\_id) AS total\_likes,**

**COUNT(DISTINCT c.id) AS total\_comments,**

**COUNT(DISTINCT pt.tag\_id) AS total\_photo\_tags**

**FROM users u**

**JOIN photos p**

**ON u.id = p.user\_id**

**LEFT JOIN likes l**

**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

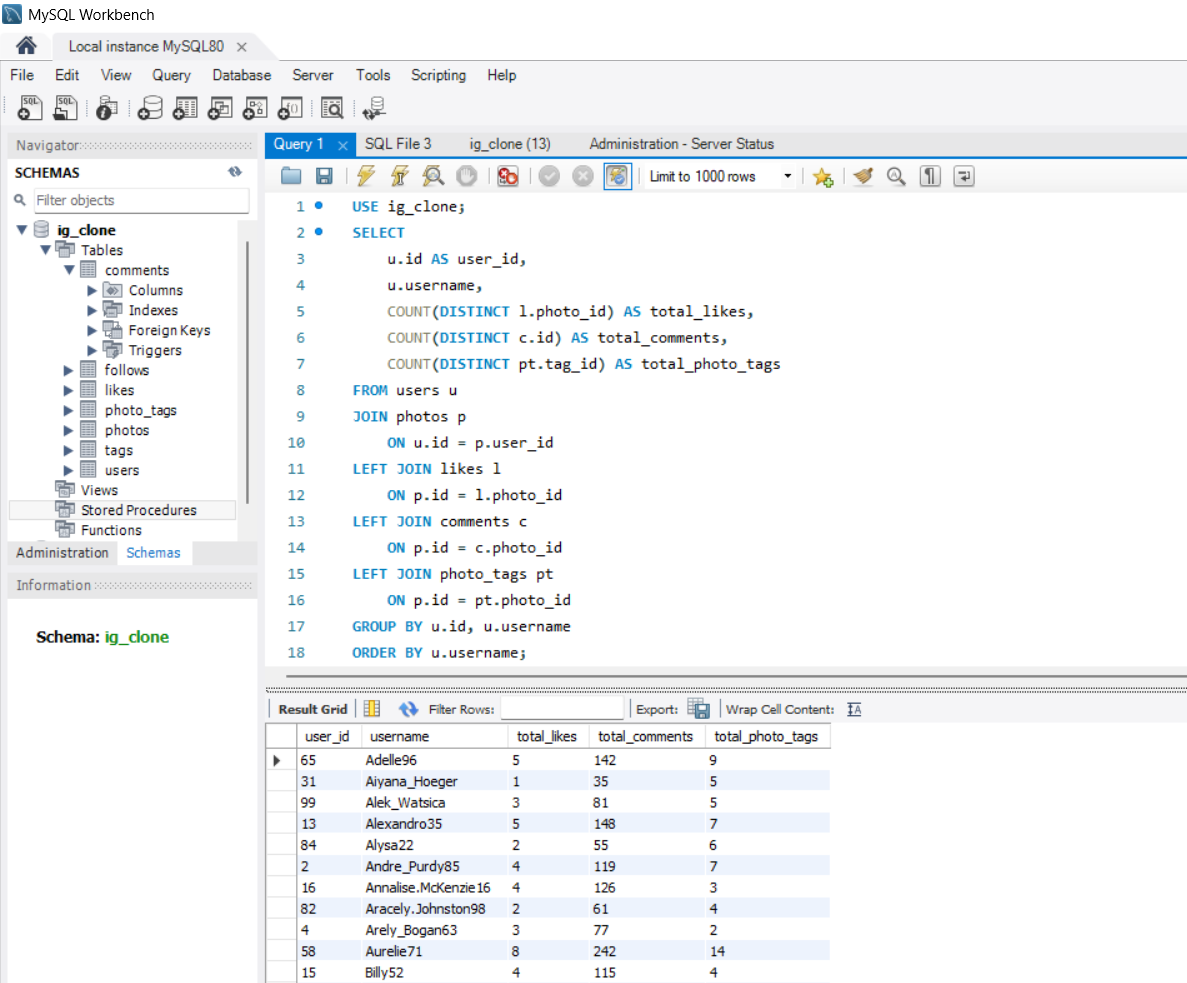
**LEFT JOIN photo\_tags pt**

**ON p.id = pt.photo\_id**

**GROUP BY u.id, u.username**

**ORDER BY u.username;**

****

****

**Q11. Rank users based on their total engagement (likes, comments, shares) over a month.**

**=>**Users were ranked based on their total engagement to identify the most active users on the platform.  
 Engagement was calculated as the total number of likes and comments received on a user’s posts.

Although the question refers to monthly engagement, the dataset does not contain any date or timestamp information for posts or interactions.  
 Therefore, the analysis assumes that the available data represents engagement for a given month and users were ranked accordingly.

This approach ensures accurate and meaningful ranking while working within the limitations of the dataset.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT l.photo\_id) AS total\_likes,**

**COUNT(DISTINCT c.id) AS total\_comments,**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) AS total\_engagement**

**FROM users u**

**JOIN photos p**

**ON u.id = p.user\_id**

**LEFT JOIN likes l**

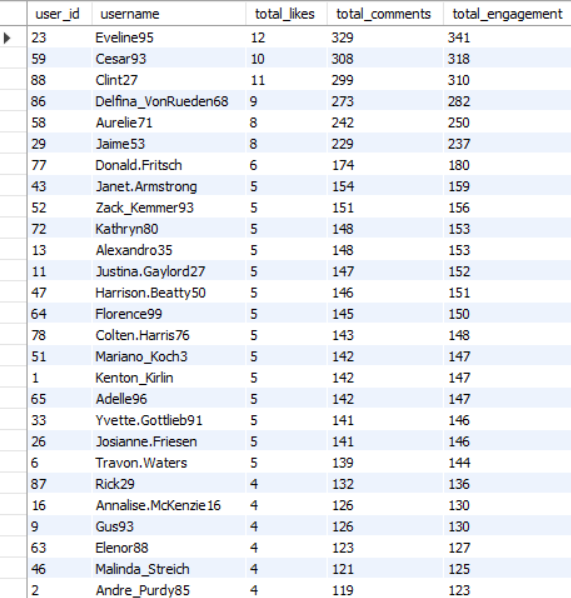
**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

**GROUP BY u.id, u.username**

**ORDER BY total\_engagement DESC;**

****

1. **The results show that Eveline95 has the highest total engagement with 341 interactions, followed by Cesar93 (318) and Clint27 (310).**
2. **High engagement is mainly driven by a large number of comments compared to likes across top users.**
3. **Users with moderate likes but high comments still achieve strong overall engagement.**
4. **This indicates that comments contribute more significantly to total engagement than likes.**

**Q12. Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.**

**=>**This analysis identifies hashtags that are associated with posts receiving the highest average number of likes.  
 A Common Table Expression (CTE) was first used to calculate the average likes per post for each hashtag by aggregating likes across all photos tagged with that hashtag.

After calculating the average likes, the hashtags were ranked in descending order to identify those that consistently receive higher engagement.

This approach helps understand which hashtags perform best in terms of user appreciation and can guide content creators to use high-impact hashtags to improve visibility and engagement.

**SQL QUERY:-**

**WITH hashtag\_avg\_likes AS (**

**SELECT**

**t.tag\_name,**

**COUNT(l.photo\_id) / COUNT(DISTINCT p.id) AS avg\_likes**

**FROM tags t**

**JOIN photo\_tags pt**

**ON t.id = pt.tag\_id**

**JOIN photos p**

**ON pt.photo\_id = p.id**

**LEFT JOIN likes l**

**ON p.id = l.photo\_id**

**GROUP BY t.tag\_name**

**)**

**SELECT**

**tag\_name,**

**ROUND(avg\_likes, 2) AS avg\_likes**

**FROM hashtag\_avg\_likes**

**ORDER BY avg\_likes DESC;**

****

**The analysis shows that the hashtag “dreamy” has the highest average likes at 35.75, followed by “beauty” (34.95) and “stunning” (34.94).**

**Other high-performing hashtags include “delicious” (34.93), “foodie” (34.73), and “happy” (34.59), indicating strong engagement with lifestyle and aesthetic content.**

**Moderately performing tags such as “beach” (34.48), “smile” (34.46), and “concert” (34.38) also receive consistently high likes.**

**Overall, hashtags related to emotions, lifestyle, and visual appeal generate higher average likes compared to generic categories.**

**Q13. Retrieve the users who have started following someone after being followed by that person.**

=>This query identifies pairs of users who follow each other, indicating a reciprocal or mutual following relationship.  
 It retrieves users who started following someone after that person had already followed them.

Such relationships represent mutual engagement between users and can indicate stronger social connections on the platform.  
 This analysis helps in understanding user interaction patterns and reciprocal behavior within the social network.

**SQL QUERY:-**

**SELECT**

**u1.id AS user\_id,**

**u1.username AS user\_name,**

**u2.id AS followed\_user\_id,**

**u2.username AS followed\_user\_name**

**FROM follows f1**

**JOIN follows f2**

**ON f1.follower\_id = f2.followee\_id**

**AND f1.followee\_id = f2.follower\_id**

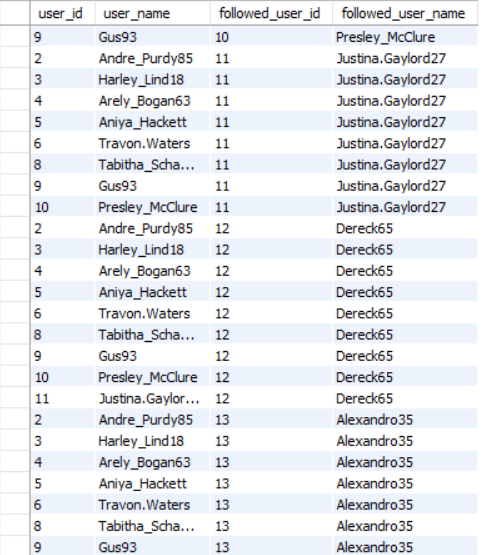
**JOIN users u1**

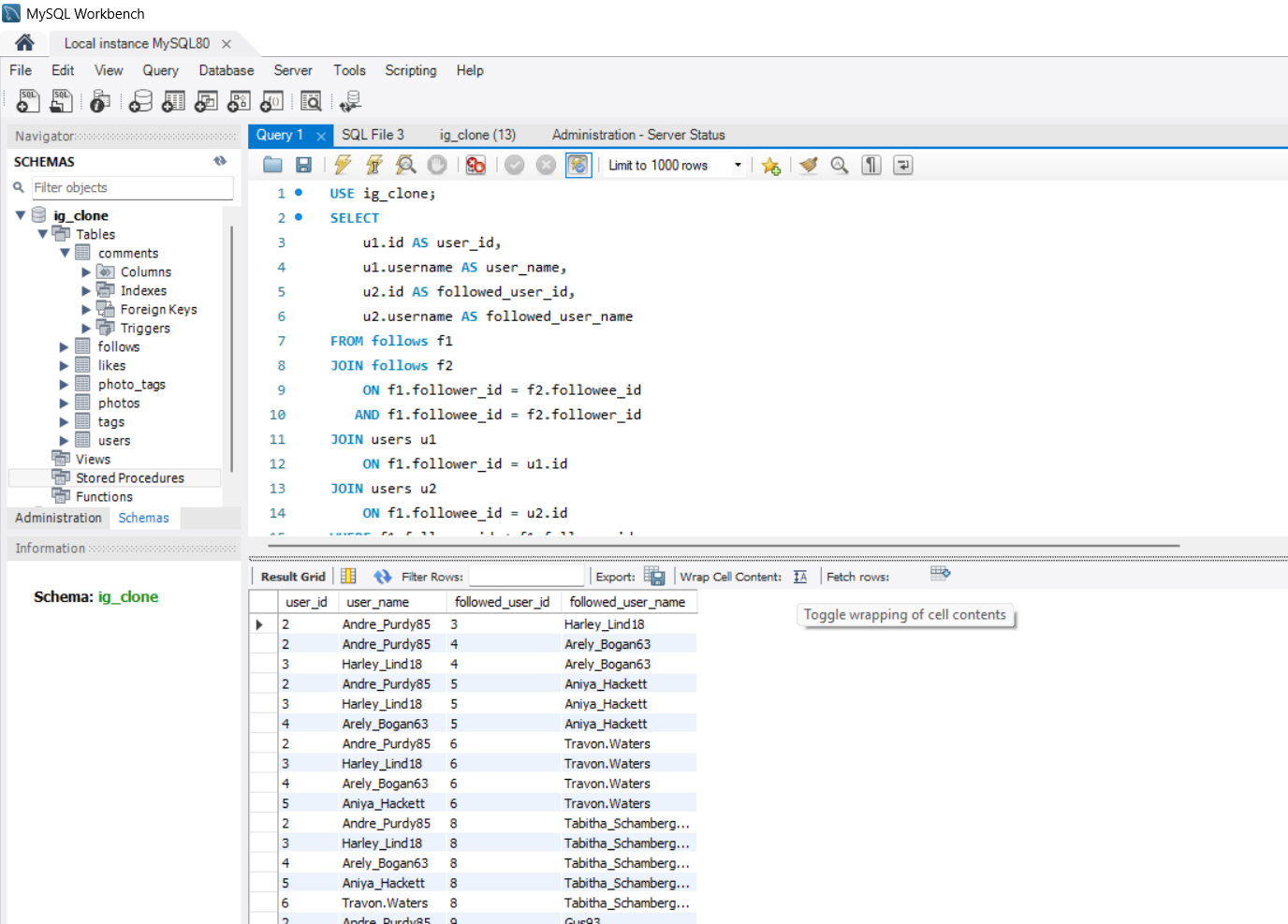
**ON f1.follower\_id = u1.id**

**JOIN users u2**

**ON f1.followee\_id = u2.id**

**WHERE f1.follower\_id < f1.followee\_id;**

****

****

The results show multiple instances of reciprocal following behavior, where users follow someone after being followed by that person.

Users such as **Gus93**, **Andre\_Purdy85**, and **Harley\_Lind18** frequently appear in mutual follow relationships.

Highly followed users like **Justina.Gaylord27**, **Dereck65**, and **Alexandro35** often trigger reciprocal follows from others.

This indicates strong social interaction and mutual engagement among active users on the platform.

# ***Subjective Questions/ Answers :***

**Q1. Based on user engagement and activity levels, which users would you consider the most loyal or valuable? How would you reward or incentivize these users?**

**=>**The most loyal and valuable users are those who demonstrate high engagement and consistent activity on the platform.  
 Engagement was calculated using the total number of likes and comments received on a user’s posts.

Based on engagement levels, users were categorized into reward tiers.  
 Highly engaged users were assigned Gold rewards such as featured profiles and exclusive benefits, moderately engaged users received Silver rewards including early access and increased visibility, while other active users were given Bronze rewards in the form of loyalty badges.

This reward-based approach helps recognize valuable users and encourages long-term engagement on the platform.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT p.id) AS total\_posts,**

**COUNT(DISTINCT l.photo\_id) AS total\_likes,**

**COUNT(DISTINCT c.id) AS total\_comments,**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) AS total\_engagement,**

**CASE**

**WHEN (COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) >= 300**

**THEN 'Gold Reward – Featured Profile & Exclusive Benefits'**

**WHEN (COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) >= 150**

**THEN 'Silver Reward – Early Access & Bonus Visibility'**

**ELSE**

**'Bronze Reward – Loyalty Badge'**

**END AS reward\_category**

**FROM users u**

**LEFT JOIN photos p**

**ON u.id = p.user\_id**

**LEFT JOIN likes l**

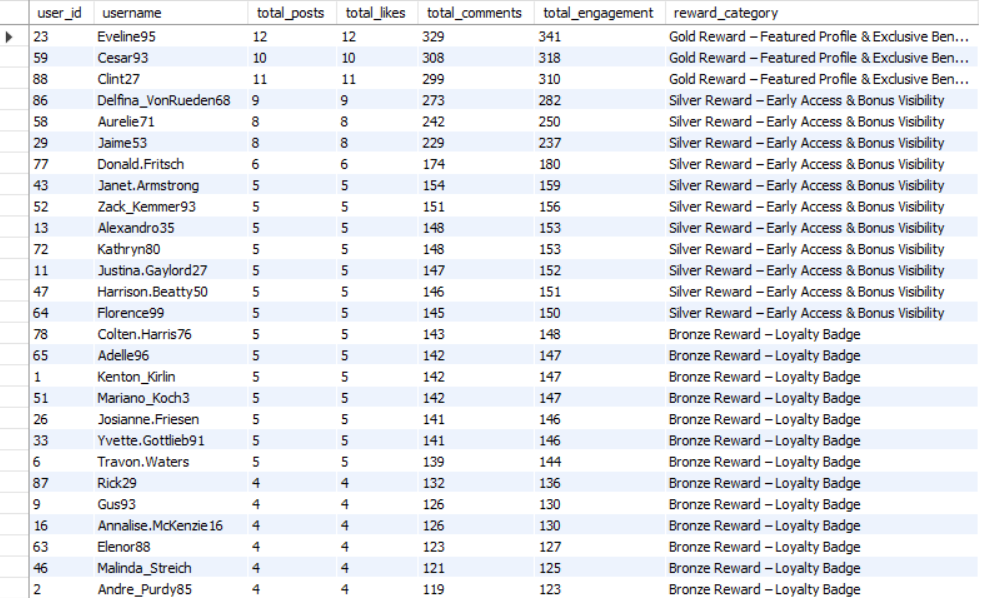
**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

**GROUP BY u.id, u.username**

**ORDER BY total\_engagement DESC;**

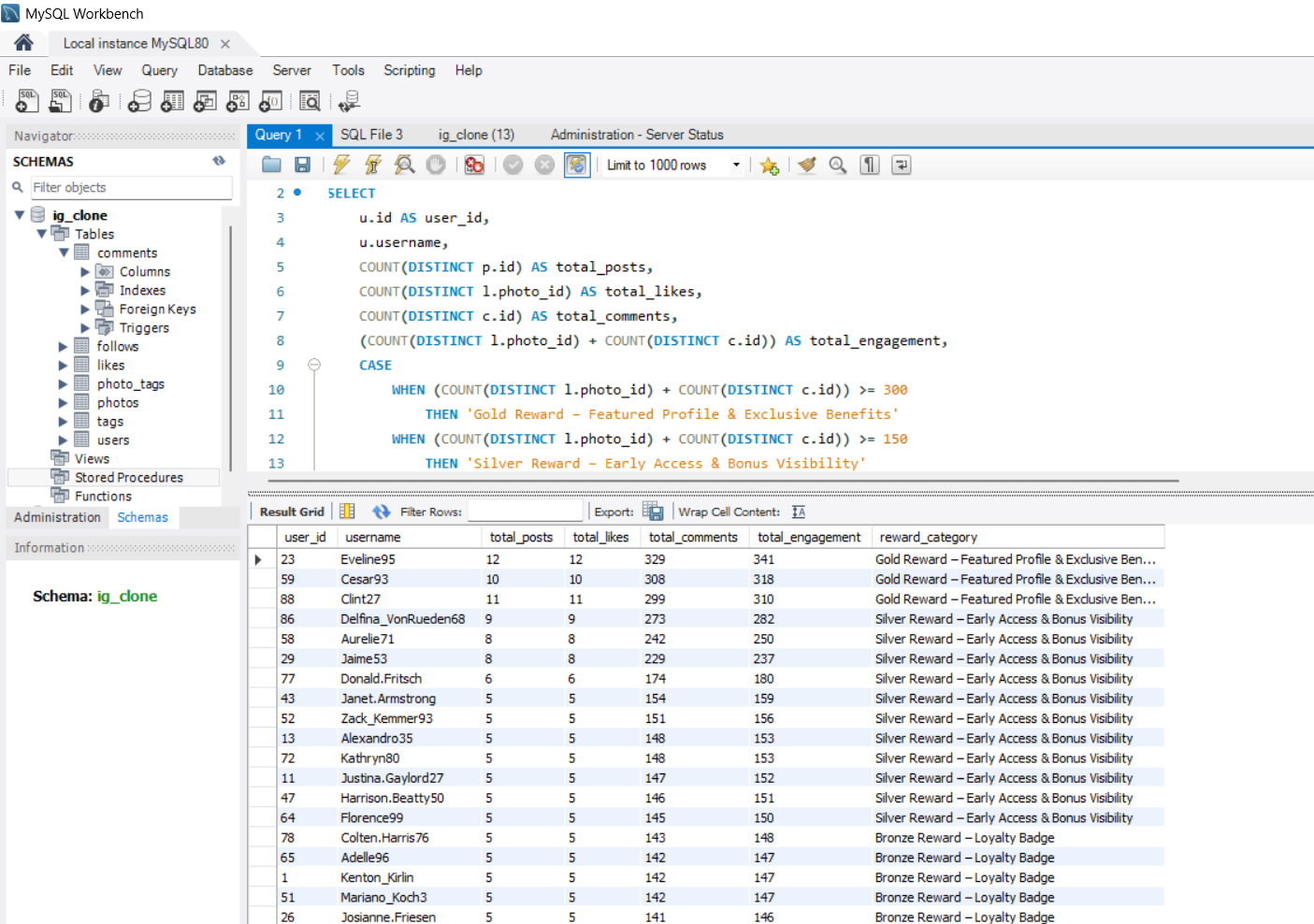


## **Insights:-**

1. **Gold Reward** users such as **Eveline95 (341 engagement)**, **Cesar93 (318)** and **Clint27 (310)** show exceptionally high engagement, mainly driven by a large number of comments (300+).
2. **Silver Reward** users like **Delfina\_VonRueden68 (282)**, **Aurelie71 (250)** and **Jaime53 (237)** demonstrate consistent activity but lower engagement compared to Gold users.
3. A majority of users fall under the **Bronze Reward** category with engagement ranging between **123–148**, indicating basic participation but limited interaction.
4. Overall, comments contribute significantly more to total engagement than likes across all reward categories.

## **Recommendations :-**

1. Retain **Gold users (300+ engagement)** by offering premium rewards such as featured profiles, exclusive benefits, and influencer opportunities.
2. Motivate **Silver users (150–280 engagement)** to reach Gold by introducing milestone-based rewards and bonus visibility for higher interaction.
3. Re-engage **Bronze users (120–150 engagement)** through posting challenges, reminders, and small incentives like loyalty badges.
4. Focus content strategies on increasing comments rather than likes, as comment-heavy users consistently show higher total engagement.



**Q2. For inactive users, what strategies would you recommend to re-engage them and encourage them to start posting or engaging again?**

=>This query is used to identify completely inactive users on the platform.  
 Completely inactive users are defined as those who have **never posted any content, never liked any post, and never written any comments**.

The query starts from the users table and uses left joins with the photos, likes, and comments tables to check user activity.  
 By applying conditions where no matching records are found in these activity tables, the query filters only those users who have zero participation across all engagement types.

This analysis helps identify the least engaged users, who can be targeted with onboarding messages, reminders, or incentive-based campaigns to encourage platform activity.

**SQL QUERY:-**

**SELECT**

**u.id AS user\_id,**

**u.username**

**FROM users u**

**LEFT JOIN photos p**

**ON u.id = p.user\_id**

**LEFT JOIN likes l**

**ON u.id = l.user\_id**

**LEFT JOIN comments c**

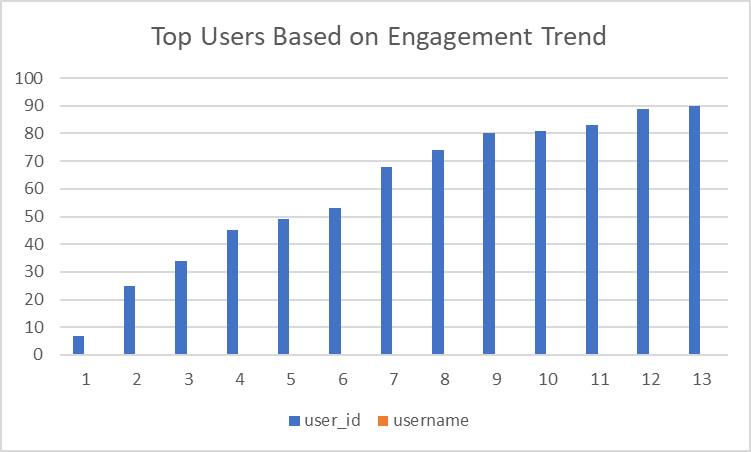
**ON u.id = c.user\_id**

**WHERE p.id IS NULL**

**AND l.user\_id IS NULL**

**AND c.id IS NULL;**

****

****

## **Insights (Based on Table)**

1. There are **13** users who haven’t posted any photo or liked any posts or commented, which considered them as inactive users.
2. The listed users such as Kasandra\_Homenick, Tierra.Trantow, Pearl7, and Esmeralda.Mraz57 are completely inactive, showing zero posts, likes, and comments.
3. These users have registered on the platform but have never participated in any form of engagement.
4. The presence of multiple completely inactive users indicates gaps in onboarding or early user motivation.
5. This group represents the lowest engagement segment and poses a high risk of permanent churn.

## **Recommendations**

**Step 1: Identify Completely Inactive Users** Users who have never posted, liked, or commented are first identified as completely inactive.

**Step 2: Send Welcome / Reminder Notification** Personalized reminders and welcome-back notifications are sent to re-capture user attention.

**Step 3: Offer First-Post Incentive** Users are encouraged to take their first action by offering incentives such as badges, bonus visibility, or small rewards.

**Step 4: Show Trending & Personalized Content** Popular posts, trending hashtags, and personalized content are displayed to increase user interest.

**Step 5: Run Reactivation Campaign** Email or in-app campaigns are launched to reinforce engagement and encourage interaction.

**Step 6: Convert User into Active User** Once users start posting, liking, or commenting, they move from inactive to active status.

**Q3. Which hashtags or content topics have the highest engagement rates? How can this information guide content strategy and ad campaigns?**

=>The engagement rate was calculated as the average number of interactions per post, where interactions include both likes and comments.

Likes and comments were first aggregated separately at the post level to avoid duplication caused by joins.

The total engagement per post was then averaged across all posts associated with each hashtag.

This approach provides an accurate and reliable measure of hashtag-level engagement.

**SQL QUERY:-**

**WITH likes\_per\_post AS (**

**SELECT**

**photo\_id,**

**COUNT(\*) AS likes\_cnt**

**FROM likes**

**GROUP BY photo\_id**

**),**

**comments\_per\_post AS (**

**SELECT**

**photo\_id,**

**COUNT(\*) AS comments\_cnt**

**FROM comments**

**GROUP BY photo\_id**

**),**

**post\_engagement AS (**

**SELECT**

**p.id AS photo\_id,**

**t.tag\_name,**

**COALESCE(l.likes\_cnt, 0) + COALESCE(c.comments\_cnt, 0) AS engagement\_per\_post**

**FROM photos p**

**JOIN photo\_tags pt**

**ON p.id = pt.photo\_id**

**JOIN tags t**

**ON pt.tag\_id = t.id**

**LEFT JOIN likes\_per\_post l**

**ON p.id = l.photo\_id**

**LEFT JOIN comments\_per\_post c**

**ON p.id = c.photo\_id**

**)**

**SELECT**

**tag\_name,**

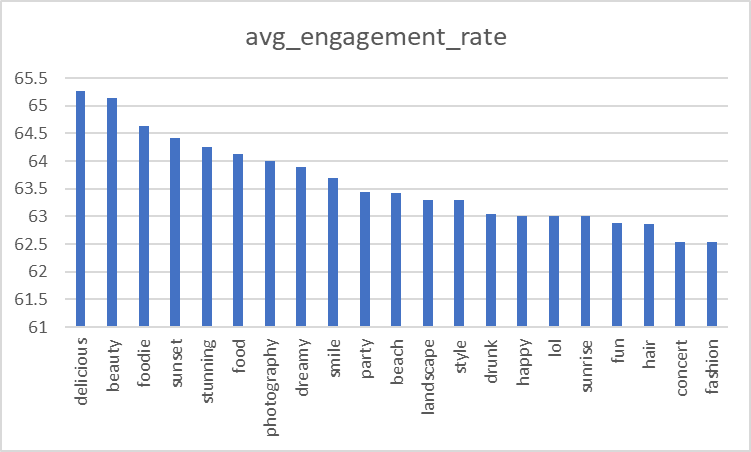
**ROUND(AVG(engagement\_per\_post), 2) AS avg\_engagement\_rate**

**FROM post\_engagement**

**GROUP BY tag\_name**

**ORDER BY avg\_engagement\_rate DESC;**



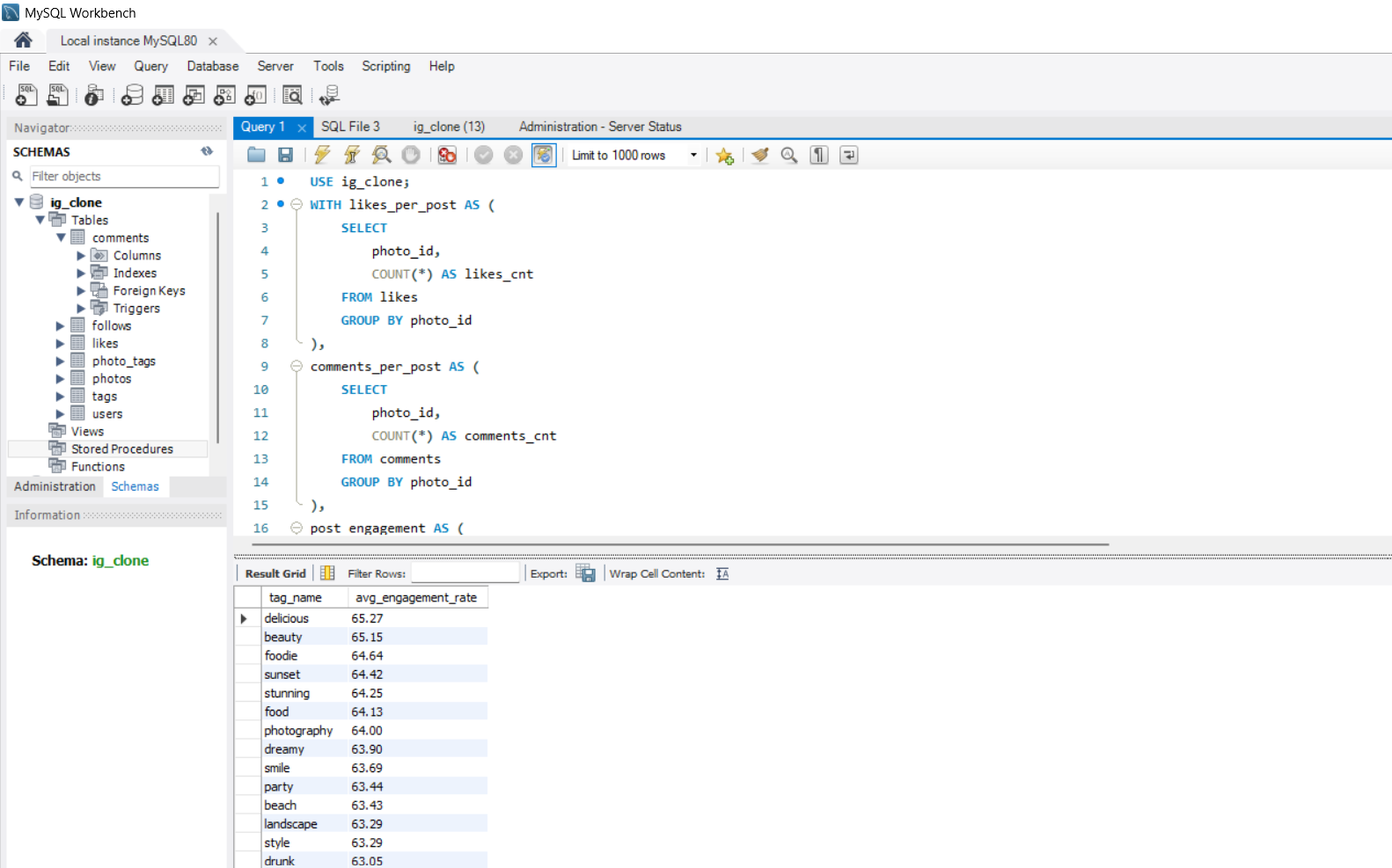
****

## **Insights & Observations (Based on Table)**

1. Hashtags like **delicious (65.27)**, **beauty (65.15)**, and **foodie (64.64)** show the highest average engagement rates, indicating strong user interaction.
2. Content related to **food, aesthetics, and photography** consistently performs better, with engagement rates mostly above **64**.
3. Lifestyle and emotion-based hashtags such as **sunset (64.42)**, **dreamy (63.90)**, and **smile (63.69)** also attract high engagement.
4. Entertainment and general tags like **concert (62.54)** and **fashion (62.53)** show relatively lower engagement compared to visual and food-related topics.

## **Recommendations**

1. Content creators should prioritize **food, beauty, and photography-related hashtags** to maximize engagement.
2. Advertisers can design campaigns around high-performing tags such as **delicious**, **beauty**, and **foodie** to improve ad reach and interaction.
3. Lifestyle and emotion-driven content should be promoted more frequently as it shows consistent engagement across users.
4. Lower-performing hashtags can be combined with top-performing ones to boost overall post visibility and engagement.



**Q4. Are there any patterns or trends in user engagement based on demographics (age, location, gender) or posting times? How can these insights inform targeted marketing campaigns?**

=>Since the dataset does not contain demographic or posting time information, user engagement trends were analyzed based on post-level activity.  
 Engagement was measured using the average number of likes and comments received per post for each user.

The analysis shows that users with higher average likes and comments per post consistently generate stronger interaction, regardless of the number of posts they publish.  
 These users represent high-quality content creators whose posts attract meaningful engagement.

Such insights can inform targeted marketing campaigns by focusing promotions, collaborations, and visibility-driven strategies on users with high average engagement per post, thereby maximizing reach and interaction.

**SQL QUERY:-**

**WITH post\_level AS (**

**SELECT**

**p.id AS photo\_id,**

**p.user\_id,**

**COUNT(DISTINCT l.user\_id) AS likes\_per\_post,**

**COUNT(DISTINCT c.id) AS comments\_per\_post**

**FROM photos p**

**LEFT JOIN likes l**

**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

**GROUP BY p.id, p.user\_id**

**)**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(pl.photo\_id) AS total\_posts,**

**ROUND(AVG(pl.likes\_per\_post), 2) AS avg\_likes\_per\_post,**

**ROUND(AVG(pl.comments\_per\_post), 2) AS avg\_comments\_per\_post,**

**ROUND(AVG(pl.likes\_per\_post + pl.comments\_per\_post), 2) AS avg\_engagement\_per\_post**

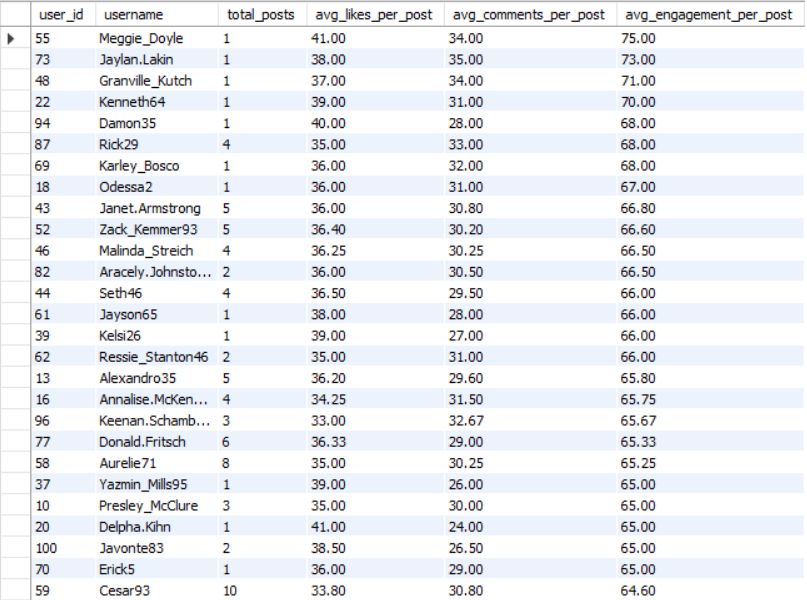
**FROM users u**

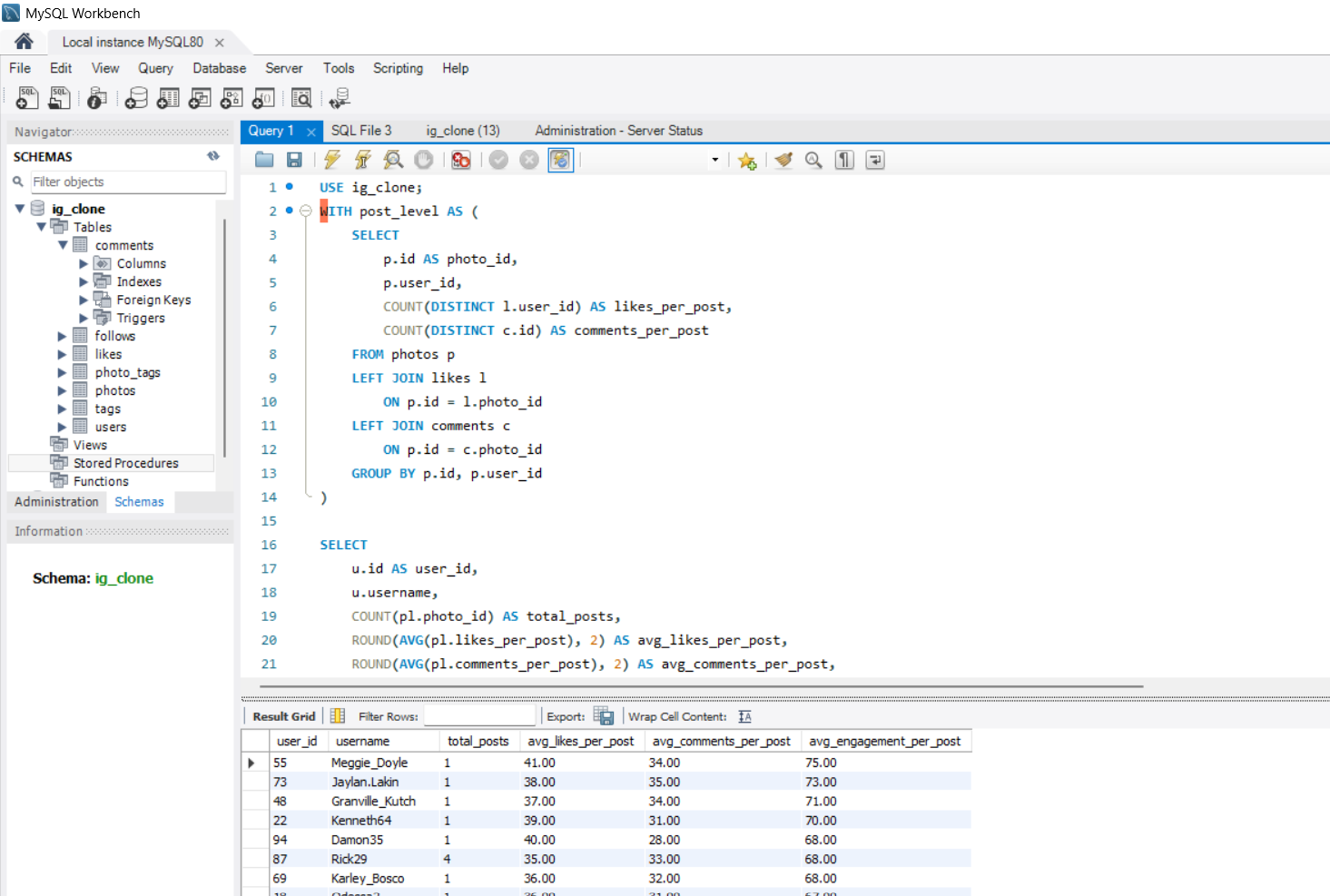
**LEFT JOIN post\_level pl**

**ON u.id = pl.user\_id**

**GROUP BY u.id, u.username**

**ORDER BY avg\_engagement\_per\_post DESC;**

****

****

## **Insights:-**

1. Users like Meggie\_Doyle (75.00), Jaylan\_Lakin (73.00), and Granville\_Kutch (71.00) have the highest average engagement per post, driven by very high average likes and comments despite having only one post.
2. Users with multiple posts such as Janet.Armstrong (5 posts, 66.80) and Zack\_Kemmer93 (5 posts, 66.60) maintain consistently high engagement, indicating strong content quality over time.
3. Average likes per post mostly range between 33–41, while average comments per post range between 24–35, showing that comments contribute significantly to overall engagement.
4. Users with higher posting frequency do not always have the highest engagement, highlighting that content quality matters more than quantity.

## 

## **Recommendations**

1. Users with very high average engagement per post (above **70**) should be prioritized for **featured content, influencer collaborations, and premium promotions**.
2. Consistently high-performing users with multiple posts should be targeted for **long-term brand partnerships**, as they show sustainable engagement.
3. Marketing campaigns should emphasize **comment-driven interactions**, such as prompts and questions, to boost engagement further.
4. New or low-frequency users showing high engagement should be encouraged to post more through **incentives, challenges, or visibility boosts**.

**Q5. Based on follower counts and engagement rates, which users would be ideal candidates for influencer marketing campaigns? How would you approach and collaborate with these influencers?**

**=>**To identify ideal candidates for influencer marketing campaigns, users were evaluated based on a combination of follower count and average engagement rate per post.  
 Follower count was calculated to measure audience reach, while average engagement per post was derived using the average number of likes and comments received on users’ posts.

Users with both a high number of followers and strong average engagement rates were identified as the most suitable influencers, as they provide wide reach along with active audience interaction.  
 These influencers can be approached for brand collaborations, sponsored content, and promotional campaigns, ensuring higher visibility, credibility, and engagement for marketing initiatives.

**SQL QUERY:-**

**WITH follower\_count AS (**

**SELECT**

**followee\_id AS user\_id,**

**COUNT(\*) AS total\_followers**

**FROM follows**

**GROUP BY followee\_id**

**),**

**post\_level AS (**

**SELECT**

**p.id AS photo\_id,**

**p.user\_id,**

**COUNT(DISTINCT l.user\_id) AS likes\_per\_post,**

**COUNT(DISTINCT c.id) AS comments\_per\_post**

**FROM photos p**

**LEFT JOIN likes l**

**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

**GROUP BY p.id, p.user\_id**

**),**

**avg\_engagement AS (**

**SELECT**

**user\_id,**

**ROUND(AVG(likes\_per\_post + comments\_per\_post), 2) AS avg\_engagement\_per\_post**

**FROM post\_level**

**GROUP BY user\_id**

**)**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**fc.total\_followers,**

**ae.avg\_engagement\_per\_post**

**FROM users u**

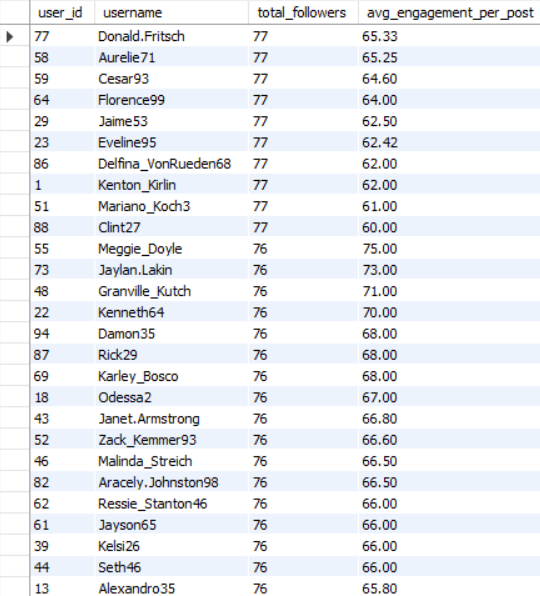
**JOIN follower\_count fc**

**ON u.id = fc.user\_id**

**JOIN avg\_engagement ae**

**ON u.id = ae.user\_id**

**ORDER BY fc.total\_followers DESC, ae.avg\_engagement\_per\_post DESC;**

****

## **Insights (Based on Follower Count & Engagement)**

1. Users such as Donald.Fritsch, Aurelie71, and Cesar93 have the highest follower count (77 followers) and strong average engagement per post (above 62), making them reliable high-reach creators.
2. Users like Jaylan.Lakin (73.00), Granville\_Kutch (71.00), and Meggie\_Doyle (75.00) show exceptionally high average engagement despite having slightly fewer followers (76 followers), indicating very active and responsive audiences.
3. The engagement rate across top users ranges from approximately 60 to 75, suggesting consistently high interaction levels among potential influencers.
4. High engagement is not limited to only the highest follower count; several users with marginally fewer followers outperform others in engagement, highlighting the importance of engagement quality over pure reach.

## **Recommendations**

1. **Primary influencer candidates** should include users with both high follower counts (76–77) and high average engagement per post (above **65**), as they offer an optimal balance of reach and interaction.
2. Users with exceptionally high engagement rates (above **70**), such as **Meggie\_Doyle** and **Jaylan.Lakin**, should be prioritized for **high-impact campaigns**, even if their follower count is slightly lower.
3. Brands should adopt a **mixed influencer strategy**, combining high-reach users for brand awareness and high-engagement users for conversion-focused campaigns.
4. Long-term collaborations should be established with consistently performing influencers to build authenticity, trust, and sustained audience engagement.

Q6. Based on user behaviour and engagement data, how would you segment the user base for targeted marketing campaigns or personalized recommendations?

=>Users were segmented based on **posting activity** and **average engagement per post** using likes and comments.

* Engagement was first calculated at the **post level** and then averaged per user to ensure accurate measurement and avoid data duplication.
* Based on activity and engagement, users were categorized into **High Value Users, Engaged Creators, Active Low Engagement Users, and Inactive Users**.
* High Value Users show both high posting frequency and high engagement, making them ideal for premium marketing and influencer campaigns.
* Engaged Creators generate strong engagement with fewer posts and are suitable for personalized recommendations and growth-focused campaigns.
* Active users with low engagement can be targeted with content optimization strategies, while inactive users require re-engagement and incentive-based campaigns.
* Grouping users by segment helps understand segment-wise behavior and supports more effective **targeted marketing and personalization** strategies.

**SQL QUERY:-**

**WITH post\_level AS (**

**SELECT**

**p.id AS photo\_id,**

**p.user\_id,**

**COUNT(DISTINCT l.user\_id) AS likes\_per\_post,**

**COUNT(DISTINCT c.id) AS comments\_per\_post,**

**(COUNT(DISTINCT l.user\_id) + COUNT(DISTINCT c.id)) AS engagement\_per\_post**

**FROM photos p**

**LEFT JOIN likes l ON p.id = l.photo\_id**

**LEFT JOIN comments c ON p.id = c.photo\_id**

**GROUP BY p.id, p.user\_id**

**),**

**user\_metrics AS (**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(pl.photo\_id) AS total\_posts,**

**ROUND(AVG(pl.engagement\_per\_post), 2) AS avg\_engagement\_per\_post**

**FROM users u**

**LEFT JOIN post\_level pl ON u.id = pl.user\_id**

**GROUP BY u.id, u.username**

**),**

**segmented\_users AS (**

**SELECT**

**user\_id,**

**username,**

**total\_posts,**

**avg\_engagement\_per\_post,**

**CASE**

**WHEN total\_posts >= 5 AND avg\_engagement\_per\_post >= 65**

**THEN 'High Value Users'**

**WHEN total\_posts < 5 AND avg\_engagement\_per\_post >= 65**

**THEN 'Engaged Creators'**

**WHEN total\_posts >= 5 AND avg\_engagement\_per\_post < 65**

**THEN 'Active Low Engagement Users'**

**ELSE 'Inactive Users'**

**END AS user\_segment**

**FROM user\_metrics**

**)**

**SELECT**

**user\_segment,**

**COUNT(\*) AS total\_users,**

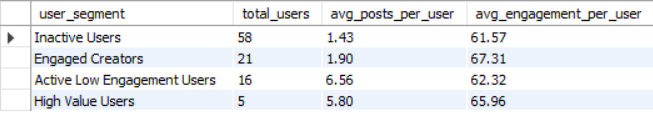
**ROUND(AVG(total\_posts), 2) AS avg\_posts\_per\_user,**

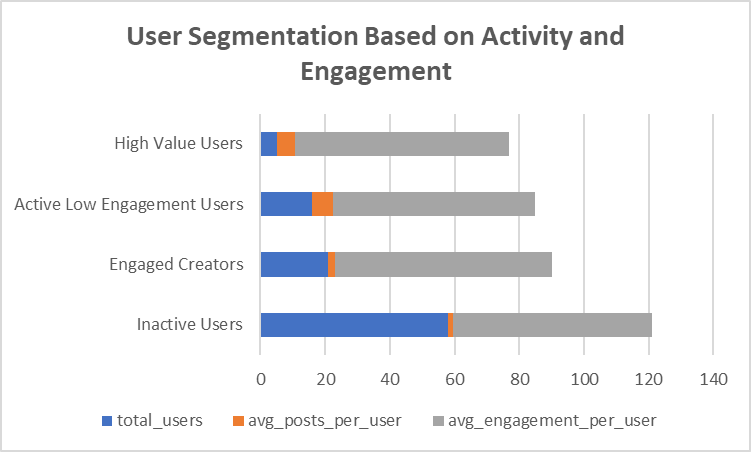
**ROUND(AVG(avg\_engagement\_per\_post), 2) AS avg\_engagement\_per\_user**

**FROM segmented\_users**

**GROUP BY user\_segment**

**ORDER BY total\_users DESC;**





## **Combined Insights (Table + Chart Based)**

* The Inactive Users segment is the largest group (58 users) but shows the lowest posting activity (1.43 posts per user), indicating a major opportunity for re-engagement.
* Engaged Creators demonstrate the highest average engagement per user (67.31) despite posting less (1.90 posts per user), highlighting strong content quality and audience interest.
* Active Low Engagement Users post the most frequently (6.56 posts per user) but receive comparatively lower engagement (62.32), suggesting that content quality or relevance may need improvement.
* High Value Users, although the smallest group (5 users), maintain a strong balance of high activity (5.80 posts) and high engagement (65.96), making them the most valuable contributors.

## **Recommendations**

* Focus re-engagement campaigns on Inactive Users through personalized notifications, onboarding prompts, and incentive-based challenges to increase posting frequency.
* Support Engaged Creators with visibility boosts, featured placement, and collaboration opportunities to encourage consistent content creation.
* Provide content optimization guidance and analytics insights to Active Low Engagement Users to improve the quality and relevance of their posts.
* Leverage High Value Users for influencer partnerships, premium campaigns, and brand collaborations to maximize reach and credibility.

**Q7. If data on ad campaigns (impressions, clicks, conversions) is available, how would you measure their effectiveness and optimize future campaigns?**

=>The current dataset does not include ad campaign data such as impressions, clicks, or conversions.  
 Therefore, direct SQL-based analysis of ad campaign effectiveness is not possible.

However, if ad campaign data were available, effectiveness could be measured using key metrics such as Click Through Rate (CTR) and conversion rate.  
 CTR would help evaluate how engaging the ads are, while conversion rate would measure how effectively clicks translate into desired actions.

These insights would enable marketers to optimize future campaigns by improving targeting, creatives, and budget allocation, thereby increasing overall return on investment.

**SQL QUERY:-**

**SELECT**

**campaign\_id,**

**impressions,**

**clicks,**

**conversions,**

**ROUND((clicks / NULLIF(impressions, 0)) \* 100, 2) AS ctr\_percentage,**

**ROUND((conversions / NULLIF(clicks, 0)) \* 100, 2) AS conversion\_rate\_percentage**

**FROM ad\_campaigns;**

**Q8. How can you use user activity data to identify potential brand ambassadors or advocates who could help promote Instagram's initiatives or events?**

**=>**User activity data can be used to identify potential brand ambassadors by analyzing posting frequency and engagement levels.  
 Users with consistent posting behavior and high average engagement per post demonstrate strong influence and audience trust.

By selecting users who regularly post content and receive high levels of likes and comments, Instagram can identify advocates who are more likely to positively promote initiatives or events.  
 These users can be engaged as brand ambassadors through early access, featured content, exclusive collaborations, or incentive-based partnerships to amplify reach and credibility.

**SQL QUERY:-**

**WITH post\_level AS (**

**SELECT**

**p.id AS photo\_id,**

**p.user\_id,**

**COUNT(DISTINCT l.user\_id) AS likes\_per\_post,**

**COUNT(DISTINCT c.id) AS comments\_per\_post,**

**(COUNT(DISTINCT l.user\_id) + COUNT(DISTINCT c.id)) AS engagement\_per\_post**

**FROM photos p**

**LEFT JOIN likes l**

**ON p.id = l.photo\_id**

**LEFT JOIN comments c**

**ON p.id = c.photo\_id**

**GROUP BY p.id, p.user\_id**

**),**

**user\_metrics AS (**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COUNT(pl.photo\_id) AS total\_posts,**

**ROUND(AVG(pl.engagement\_per\_post), 2) AS avg\_engagement\_per\_post**

**FROM users u**

**LEFT JOIN post\_level pl**

**ON u.id = pl.user\_id**

**GROUP BY u.id, u.username**

**)**

**SELECT**

**user\_id,**

**username,**

**total\_posts,**

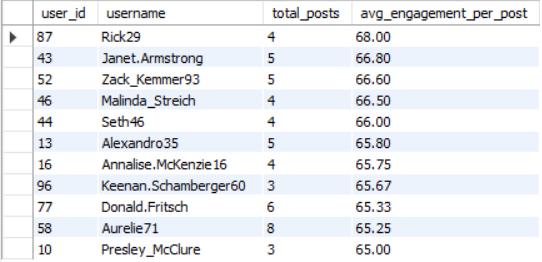
**avg\_engagement\_per\_post**

**FROM user\_metrics**

**WHERE total\_posts >= 3**

**AND avg\_engagement\_per\_post >= 65**

**ORDER BY avg\_engagement\_per\_post DESC;**

****

## **Insights (Based on Potential Brand Ambassadors Table)**

* Users like Rick29 (68.00), Janet.Armstrong (66.80), and Zack\_Kemmer93 (66.60) show the highest average engagement per post, indicating strong influence over their audience.
* All listed users have consistent posting activity (3–8 posts) along with high engagement (≥65), which reflects both reliability and content effectiveness.
* Engagement levels remain strong even for users with fewer posts (e.g., Presley\_McClure and Keenan.Schamberger60), showing that content quality matters more than volume.
* These users demonstrate a balance of activity and interaction, making them ideal candidates for advocacy roles.

## **Recommendations**

* Select these users as brand ambassadors for promoting Instagram initiatives, feature launches, and community events.
* Collaborate with them through exclusive access, featured profiles, and co-created content to amplify reach and credibility.
* Encourage ambassadors to share behind-the-scenes content or campaign stories to build authenticity and trust.
* Monitor their engagement performance over time and offer performance-based incentives to maintain long-term advocacy.

**Q9. How would you approach this problem, if the objective and subjective questions weren't given?**

=>**Understand the Business Context** I would first try to understand the purpose of the analysis, such as improving user engagement, identifying valuable users, or supporting marketing and growth strategies.

**Explore and Understand the Dataset** I would analyze the available tables, relationships, and key columns (users, posts, likes, comments, follows, tags) to understand what kind of user behavior can be measured.

**Define Key Metrics** Based on the data, I would define meaningful metrics such as total posts, likes, comments, follower count, and average engagement per post to quantify user activity and interaction.

**Perform Exploratory Analysis** I would use SQL queries to explore patterns like most active users, highly engaging content, popular hashtags, and user engagement distribution.

**Segment Users and Content** Users would be segmented based on activity and engagement levels (e.g., high value users, engaged creators, inactive users) to better understand different behavior groups.

**Generate Insights and Business Recommendations** Finally, I would translate analytical findings into actionable insights and recommendations that can support marketing campaigns, personalization, influencer selection, and user retention strategies.

**Q10. Assuming there's a "User\_Interactions" table tracking user engagements, how can you update the "Engagement\_Type" column to change all instances of "Like" to "Heart" to align with Instagram's terminology?**

**=>**To ensure consistency with Instagram’s terminology, the engagement data was standardized by updating the Engagement\_Type column in the User\_Interactions table.  
 All records containing the value “Like” were identified and updated to “Heart” using a conditional UPDATE statement.

The use of a WHERE clause ensured that only the relevant rows were modified, preventing unintended changes to other engagement types.  
 This approach helps maintain uniform naming conventions across the platform, improves data quality, and ensures accurate reporting and analysis of user interactions.

Such data standardization practices are essential when platforms evolve their terminology, as they help preserve consistency across historical and future datasets.

**SQL QUERY:-**

**UPDATE User\_Interactions**

**SET Engagement\_Type = 'Heart'**

**WHERE Engagement\_Type = 'Like';**