**SQL: Social Media Project**

**OBJECTIVE QUESTIONS**

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

**Answer:**

In the provided data there are no any duplicates or missing values.

* If there would have been missing values then I could have been identified it by :

SELECT COUNT(\*)

FROM users

WHERE username IS NULL;

* After identifying i would have used below method for deleting it:

DELETE FROM users

WHERE username IS NULL;

* ANd could have replaced or update it by:

UPDATE users

SET username = default\_value

WHERE id IS NULL;

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

Query for user activity across the user base:

SELECT u.id AS user\_id, u.username,

COALESCE(post\_counts.post\_count, 0) AS post\_count,

COALESCE(like\_counts.like\_count, 0) AS like\_count,

COALESCE(comment\_counts.comment\_count, 0) AS comment\_count

FROM users u

LEFT JOIN

(SELECT user\_id, COUNT(id) AS post\_count

FROM photos

GROUP BY user\_id) post\_counts

ON u.id = post\_counts.user\_id

LEFT JOIN

(SELECT user\_id, COUNT(photo\_id) AS like\_count

FROM likes

GROUP BY user\_id) like\_counts

ON u.id = like\_counts.user\_id

LEFT JOIN

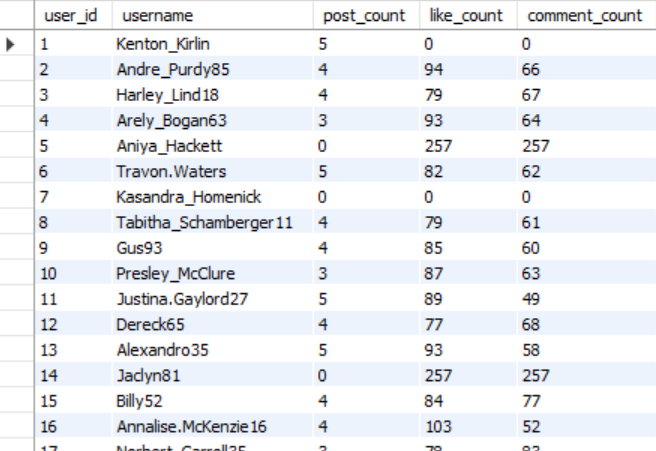
(SELECT user\_id, COUNT(id) AS comment\_count

FROM comments

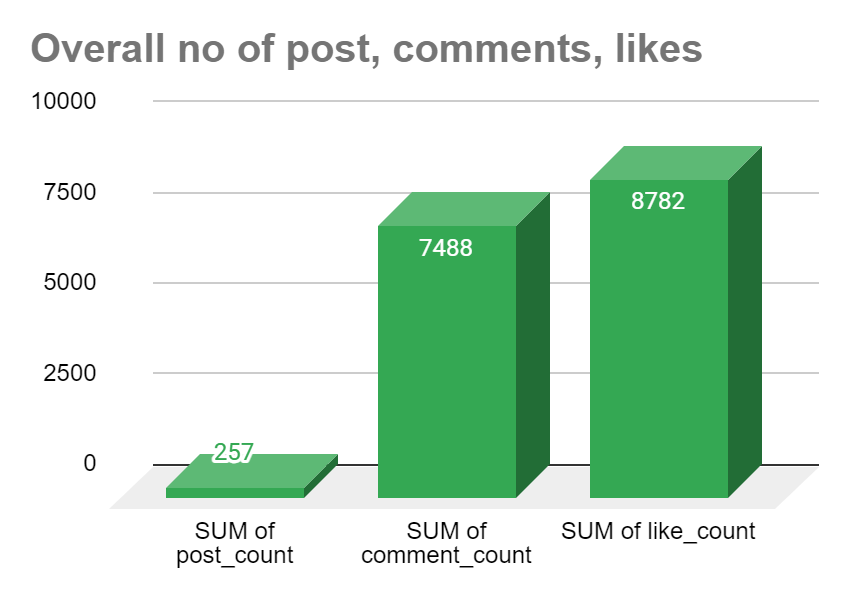
GROUP BY user\_id) comment\_counts

ON u.id = comment\_counts.user\_id;

Here is data of some of users:



Also made visualisation of overall posts, likes and comments:



**—---------------------------------------------------------------------------------------------**

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

**Answer:**

The average number of tags per post i found to be **1.95**

****

SELECT round(AVG(tag\_count),2) AS average\_tags\_per\_post

FROM (

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

FROM photos p

LEFT JOIN photo\_tags pt ON p.id = pt.photo\_id

GROUP BY p.id

) AS photo\_tag\_counts;

I have written a query in which first I calculated the average number of tags per post, rounded to two decimal places. It first counts the number of tags associated with each photo by joining the “photos” and “photo\_tags” tables, then computes the average tag count across all photos.

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

I have identified top users with the highest engagement rates on their posts.

WITH EngagementRate AS (

SELECT

u.id AS user\_id,

u.username,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

COALESCE(p.total\_posts, 0) AS total\_posts,

(COALESCE(l.total\_likes, 0) + COALESCE(c.total\_comments, 0)) / COALESCE(p.total\_posts, 1) AS engagement\_rate

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(photo\_id) AS total\_likes

FROM likes

GROUP BY user\_id

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

LEFT JOIN (

SELECT user\_id, COUNT(id) AS total\_comments

FROM comments

GROUP BY user\_id

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

LEFT JOIN (

SELECT user\_id, COUNT(id) AS total\_posts

FROM photos

GROUP BY user\_id

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

)

SELECT

user\_id,

username,

total\_likes,

total\_comments,

total\_posts,

ROUND(engagement\_rate, 2) AS engagement\_rate,

RANK() OVER (ORDER BY engagement\_rate DESC) AS engagement\_rank

FROM EngagementRate

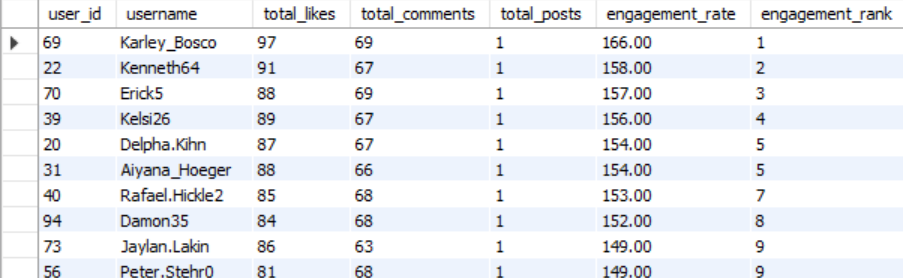
GROUP BY user\_id

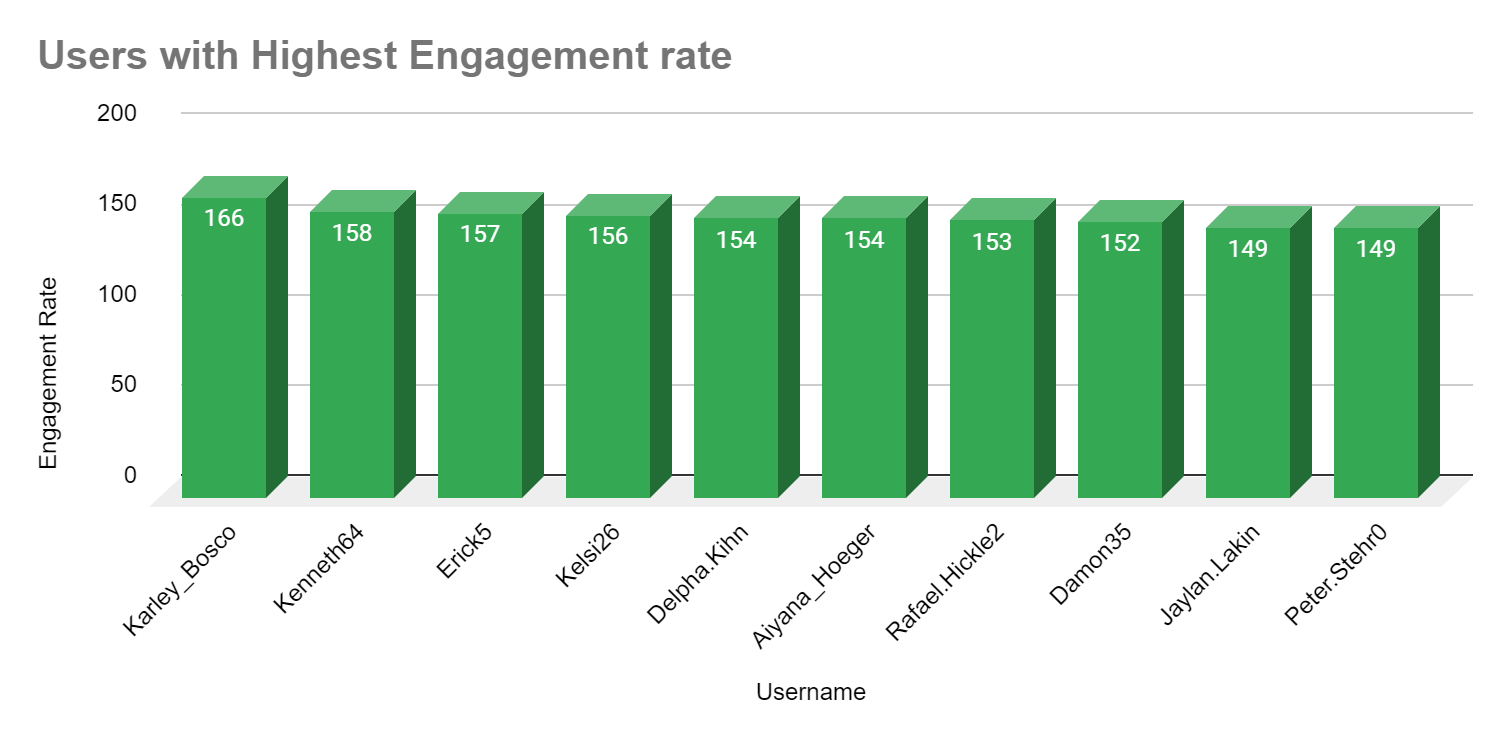
HAVING total\_posts >0

ORDER BY engagement\_rank

limit 10;

Following are the top users:





**—---------------------------------------------------------------------------------------------**

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

**Answer:**

In provided data there are almost the same number of followers and followings.

* By analysing data, 23 users have the highest number of followers that is 77 followers and remaining 77 users have 76 followers.

SELECT

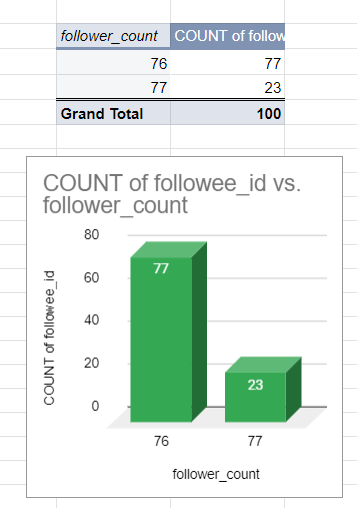
followee\_id,

COUNT(follower\_id) AS follower\_count

FROM follows

GROUP BY followee\_id

ORDER BY follower\_count DESC;

****

* AS per provided data all users have the same number of following which is following count = 99.

SELECT

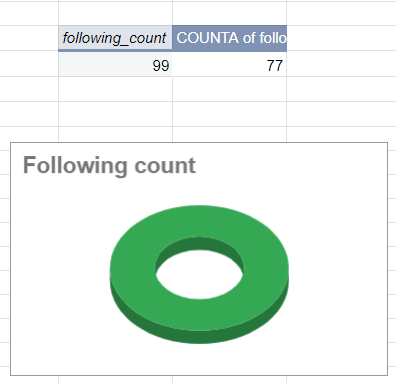
follower\_id,

COUNT(followee\_id) AS following\_count

FROM follows

GROUP BY follower\_id

ORDER BY following\_count DESC;



**—---------------------------------------------------------------------------------------------**

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

**Answer:**

Query for calculating average engagement rate per post for each user:

SELECT

u.id as user\_id,

u.username,

COALESCE(p.num\_posts, 0) AS Total\_Post,

COALESCE(l.num\_likes, 0) AS Total\_likes,

COALESCE(c.num\_comments, 0) AS Total\_comments,

CASE

WHEN COALESCE(p.num\_posts, 0) = 0 THEN 0

ELSE (COALESCE(l.num\_likes, 0) + COALESCE(c.num\_comments, 0)) / COALESCE(p.num\_posts, 0)

END AS avg\_engagement\_rate

FROM users u

LEFT JOIN

(SELECT user\_id, COUNT(\*) AS num\_posts

FROM photos

GROUP BY user\_id) p ON u.id = p.user\_id

LEFT JOIN

(SELECT user\_id, COUNT(\*) AS num\_likes

FROM likes

GROUP BY user\_id) l ON u.id = l.user\_id

LEFT JOIN

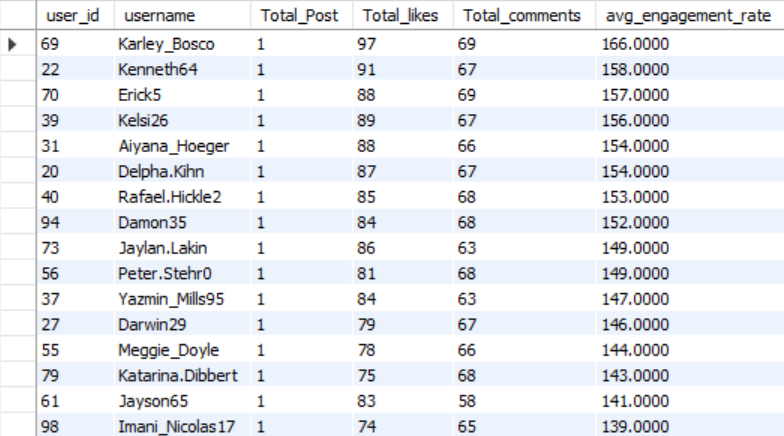
(SELECT user\_id, COUNT(\*) AS num\_comments

FROM comments

GROUP BY user\_id) c ON u.id = c.user\_id

ORDER BY avg\_engagement\_rate DESC;

Please see complete output from workbench, Here is data of some of users having highest(average engagement rate):



**—---------------------------------------------------------------------------------------------**

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

**Answer:**

For checking users who have never liked any post I have written a simple query by using subquery. I found a total of 23 users who did not like any post.

SELECT id as user\_id , username

FROM users

where id not in(

select user\_id from likes);

List of users who have never liked any post:

****

**------------------------------------------------------------------------------------------------**

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

**Answer:**

To create more personalised and engaging ad campaigns using user-generated content:

SELECT t.tag\_name, COUNT(\*) AS total\_posts, AVG(COALESCE(likes\_count.likes, 0) + COALESCE(comments\_count.comments, 0)) AS avg\_engagement

FROM tags t

JOIN photo\_tags pt ON t.id = pt.tag\_id

JOIN photos p ON pt.photo\_id = p.id

LEFT JOIN

(SELECT photo\_id, COUNT(\*) AS likes FROM likes GROUP BY photo\_id) AS likes\_count ON p.id = likes\_count.photo\_id

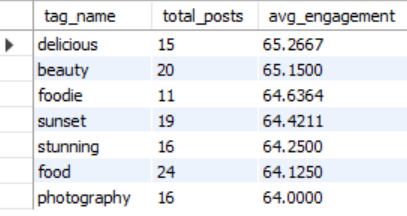
LEFT JOIN

(SELECT photo\_id, COUNT(\*) AS comments FROM comments GROUP BY photo\_id) AS comments\_count ON p.id = comments\_count.photo\_id

GROUP BY t.tag\_name

having avg\_engagement >= 64

ORDER BY avg\_engagement DESC;



By analysing this content, I can gain insights into what users like about the brand, identify trends, and understand the audience's preferences. Using user-generated content in ad campaigns will make them more relatable and authentic by showcasing real customer experiences. Additionally, I would collaborate with influencers to expand the reach and credibility of the campaigns. Finally, I would track the campaign's performance and make improvements based on feedback to ensure it resonates with the target audience.

**—---------------------------------------------------------------------------------------------**

1. **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?**

**Answer:**

If we would have given “content\_type” column to the “photos” table then we can find the correlation between user activity levels and specific content types:

SELECT p.content\_type,

COUNT(DISTINCT p.id) AS total\_posts, COUNT(DISTINCT l.user\_id) AS total\_likes, COUNT(DISTINCT c.user\_id) AS total\_comments, (COUNT(DISTINCT l.user\_id) + COUNT(DISTINCT c.user\_id)) AS total\_engagement

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.content\_type

ORDER BY total\_engagement DESC;

By analysing the results, we can identify which content types generate the most engagement. This insight can guide the focus of future content creation efforts towards the most engaging formats.

This information guides content creation. For example, if videos receive the highest engagement, the strategy might include producing more video content. **—---------------------------------------------------------------------------------------------**

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

**Answer:**

Query to calculate total number of likes, comments, photo tags:

WITH LikesCount AS (

SELECT user\_id, COUNT(\*) AS total\_likes

FROM likes

GROUP BY user\_id

),

CommentsCount AS (

SELECT user\_id, COUNT(\*) AS total\_comments

FROM comments

GROUP BY user\_id

),

PhotoTagsCount AS (

SELECT tag\_id, COUNT(\*) AS total\_photo\_tags

FROM photo\_tags

GROUP BY tag\_id

)

SELECT

u.id as id,

u.username,

COALESCE(lc.total\_likes, 0) AS total\_likes,

COALESCE(cc.total\_comments, 0) AS total\_comments,

COALESCE(ptc.total\_photo\_tags, 0) AS total\_photo\_tags

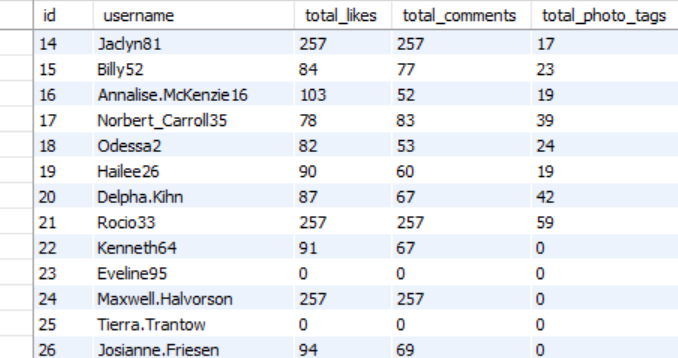
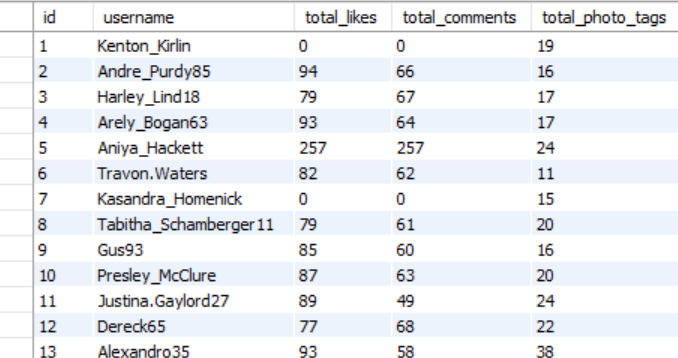
FROM users u

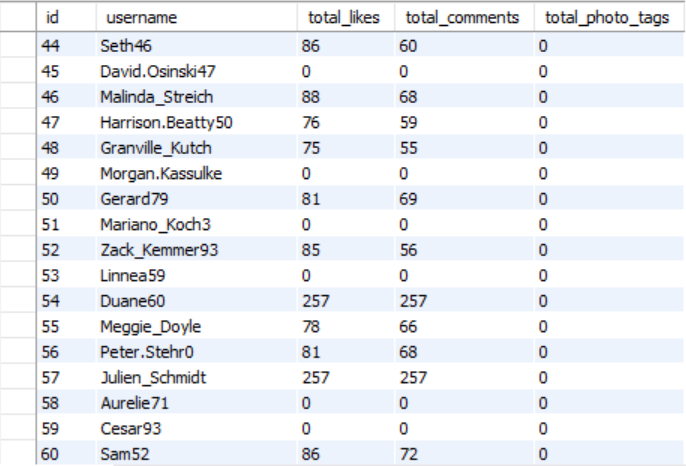
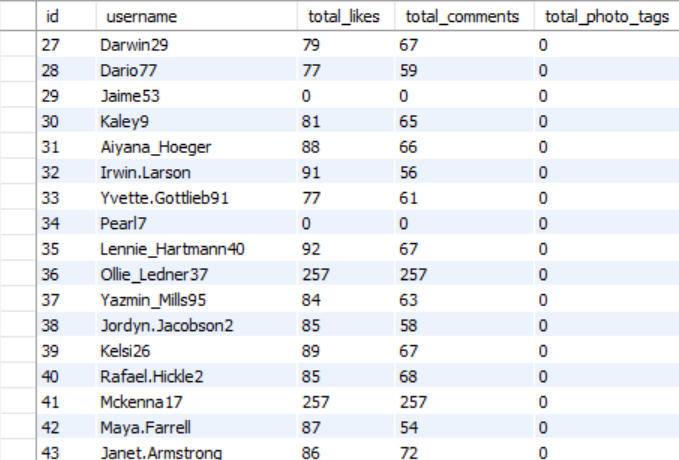
LEFT JOIN LikesCount lc ON u.id = lc.user\_id

LEFT JOIN CommentsCount cc ON u.id = cc.user\_id

LEFT JOIN PhotoTagsCount ptc ON u.id = ptc.tag\_id;

Total number of likes, comments, photo tags for each user:





**—---------------------------------------------------------------------------------------------**

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

**Answer:**

To rank users based on their total engagement over a month :

WITH MonthlyEngagement AS (

SELECT u.id AS user\_id,

u.username,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

(COALESCE(l.total\_likes, 0) + COALESCE(c.total\_comments, 0)) AS total\_engagement

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(photo\_id) AS total\_likes

FROM likes

WHERE DATE(created\_at) >= '2024-07-01' OR DATE(created\_at) <= '2024-07-31'

GROUP BY user\_id

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

LEFT JOIN (

SELECT user\_id, COUNT(id) AS total\_comments

FROM comments

WHERE DATE(created\_at) >= '2024-07-01' OR DATE(created\_at) <= '2024-07-31'

GROUP BY user\_id

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

)

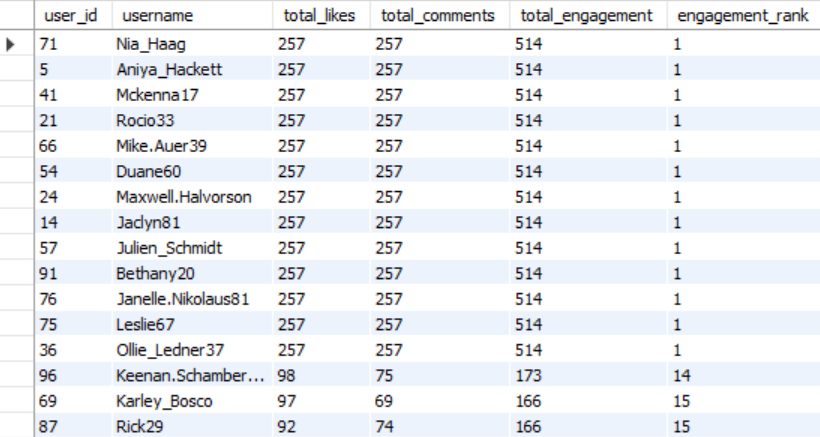
SELECT user\_id, username, total\_likes, total\_comments, total\_engagement,

RANK() OVER (ORDER BY total\_engagement DESC) AS engagement\_rank

FROM MonthlyEngagement

ORDER BY engagement\_rank;

Here are some initial users with total engagement over a month, for compleet output please visit workbench:



**—---------------------------------------------------------------------------------------------**

1. **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.**

**Answer:**

WITH HashtagLikes AS (

SELECT ht.tag\_name, COUNT(l.photo\_id) AS total\_likes, COUNT(DISTINCT p.id) AS total\_posts

FROM tags ht

JOIN photo\_tags pt ON ht.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 ht.tag\_name

),

AverageLikesPerHashtag AS (

SELECT tag\_name, ROUND((CAST(total\_likes AS FLOAT) / total\_posts),2) AS avg\_likes

FROM HashtagLikes

)

SELECT tag\_name, avg\_likes

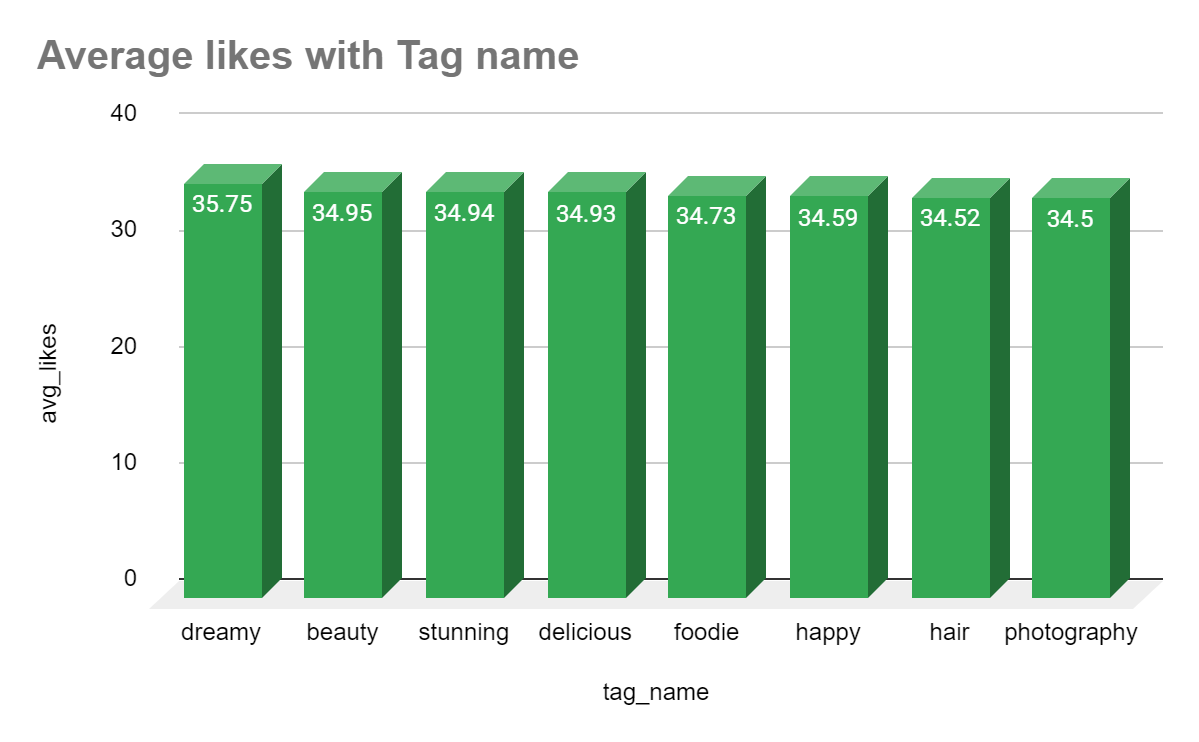
FROM AverageLikesPerHashtag

group by tag\_name

having avg\_likes >= 34.5

ORDER BY avg\_likes DESC;

Hashtags that have been used in posts with the highest average number of likes:



**—---------------------------------------------------------------------------------------------**

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

**Answer:**

I have written the following query for retrieving users who have started following someone after being followed by that person. But in the given data there is no such entry hence the query returns 0 rows.

WITH FollowRelationships AS (

SELECT

f1.follower\_id AS user\_id\_1,

f1.followee\_id AS user\_id\_2,

f1.created\_at AS followee\_follow\_date,

f2.created\_at AS follower\_follow\_date

FROM follows f1

JOIN follows f2 ON f1.followee\_id = f2.follower\_id

WHERE f1.follower\_id <> f2.followee\_id

)

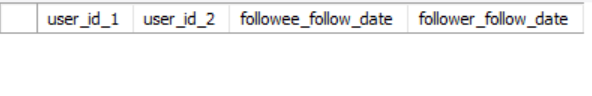
SELECT user\_id\_1, user\_id\_2, followee\_follow\_date, follower\_follow\_date

FROM FollowRelationships

WHERE follower\_follow\_date > followee\_follow\_date

ORDER BY user\_id\_1, user\_id\_2;

This query identifies users who follow each other indirectly by finding pairs of users where one follows the other, and that person also follows a third user.





**—---------------------------------------------------------------------------------------------**

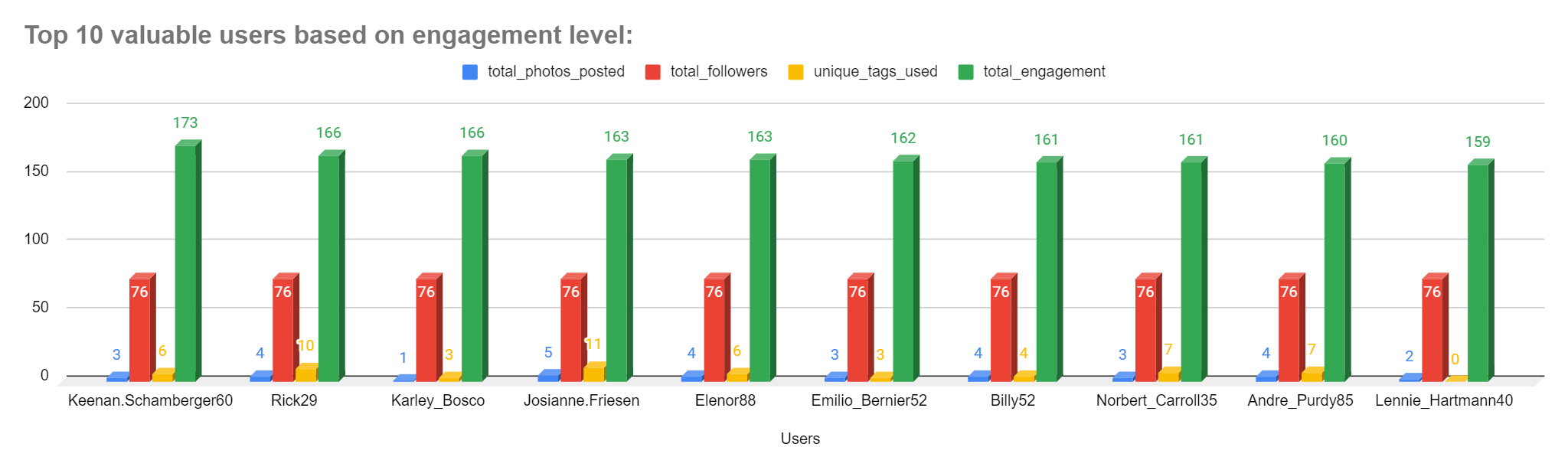
**SUBJECTIVE QUESTIONS**

1. **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?**

**Answer:**

I have shown top 10 user s who are most loyal and valuable:





To reward and incentivize the most valuable users on our platform:

* We can offer subscription services to the top users, such as ad-free experiences, early access to new content, or additional features. By giving discounts on these subscriptions to our most active users, we show them that their participation is important to us and reward them with more value from the platform.
* We can provide them early access to new features, by giving them this opportunity to try out upcoming updates we can make them feel like insiders and part of the community.
* Also we can provide them interesting goodies for appreciating their work or engagement on social media

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

For inactive users, we can make several strategies like

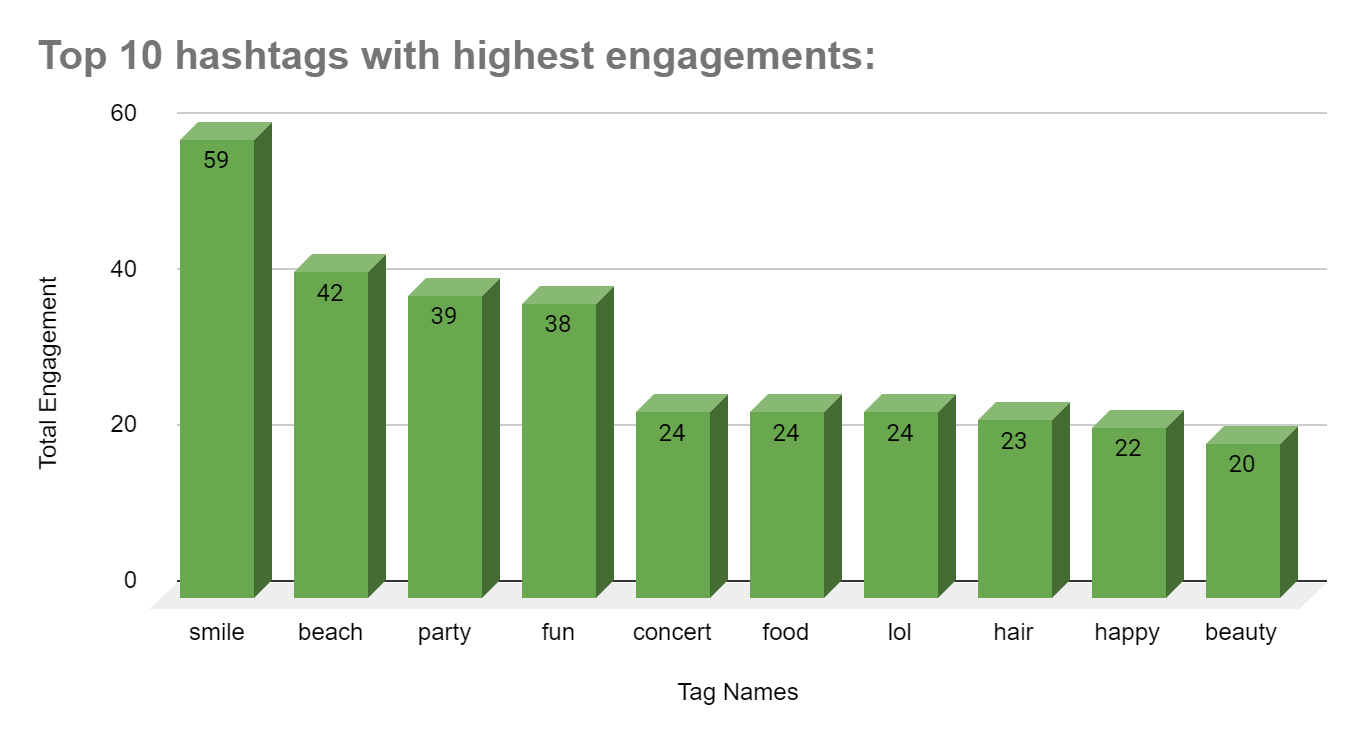
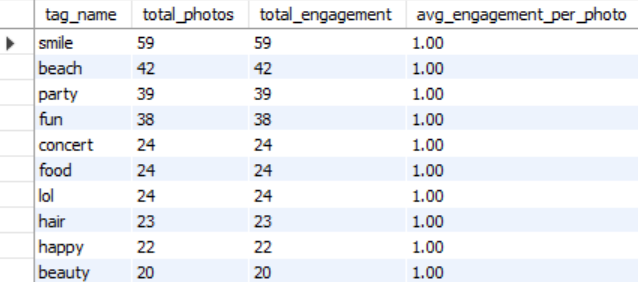
* To motivate inactive users, we can set specific targets that award badges based on their level of activity. For example, after posting a certain number of images or content pieces, users will earn Bronze, Silver, and Gold badges that will be displayed around their profile picture. It will encourage users to become more active on the platform.
* We can offer them 30 days of free subscriptions so that they get engaged on the platform. It might increase their interest in our platform.
* We can send personalised notifications regarding trending posts or content, new followers, comments on their posts, or events they might be interested in, by this we can take users back to the platform by reminding them of their existing connections.

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

Hashtags that have the highest engagement rates are:

****

This information guide content strategy and ad campaigns as follows:

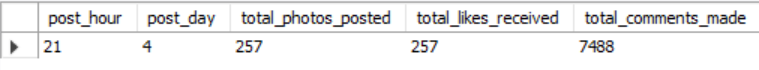
* For content strategy: By using this information we got to know which topics or hashtags are popular, after watching total\_posts and total\_engagemnet column which indicates how many number of photos got tagged and how many comments and likes have got for the particular hashtag, so that content creators can make posts that include these trending hashtags. This increases the likelihood that people will engage with their content, like, comment, or share it.
* For ad campaigns: Advertisers can use this information to place their ads alongside content that is currently trending (by looking at the total\_posts and total\_engagemnet with respect to hashtags), which means more people are likely to see these ads.

**—---------------------------------------------------------------------------------------------**

1. **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?**

**Answer:**

In the provided data, there is no any data related to age, location, gender so we could not find any trend related to these factors so I have shown a pattern using posting times.



Approach about How can these insights inform targeted marketing campaigns:

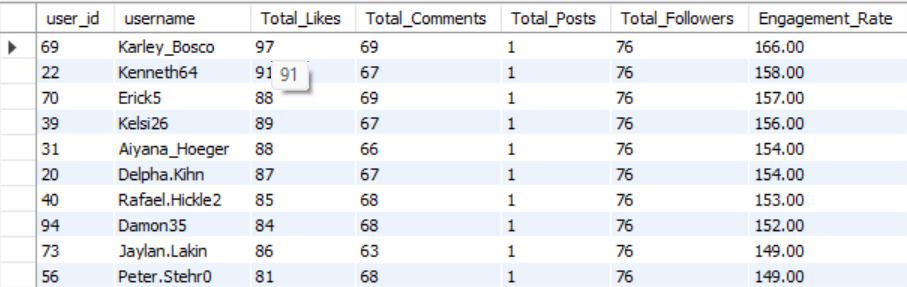
* By using this analysis, marketers can identify when users are most active (e.g., post\_hour and post\_day with the highest total\_likes\_received or total\_comments\_made), marketers can schedule their campaigns during these peak engagement times to maximize visibility and interaction.
* If data related to age and gender would have been given then marketers could have analysed habits of different audience segments. For instance, younger audiences might engage more during late hours, while professionals might be more active during lunch breaks or evenings. Campaigns can be tailored to target these specific groups at the right times.

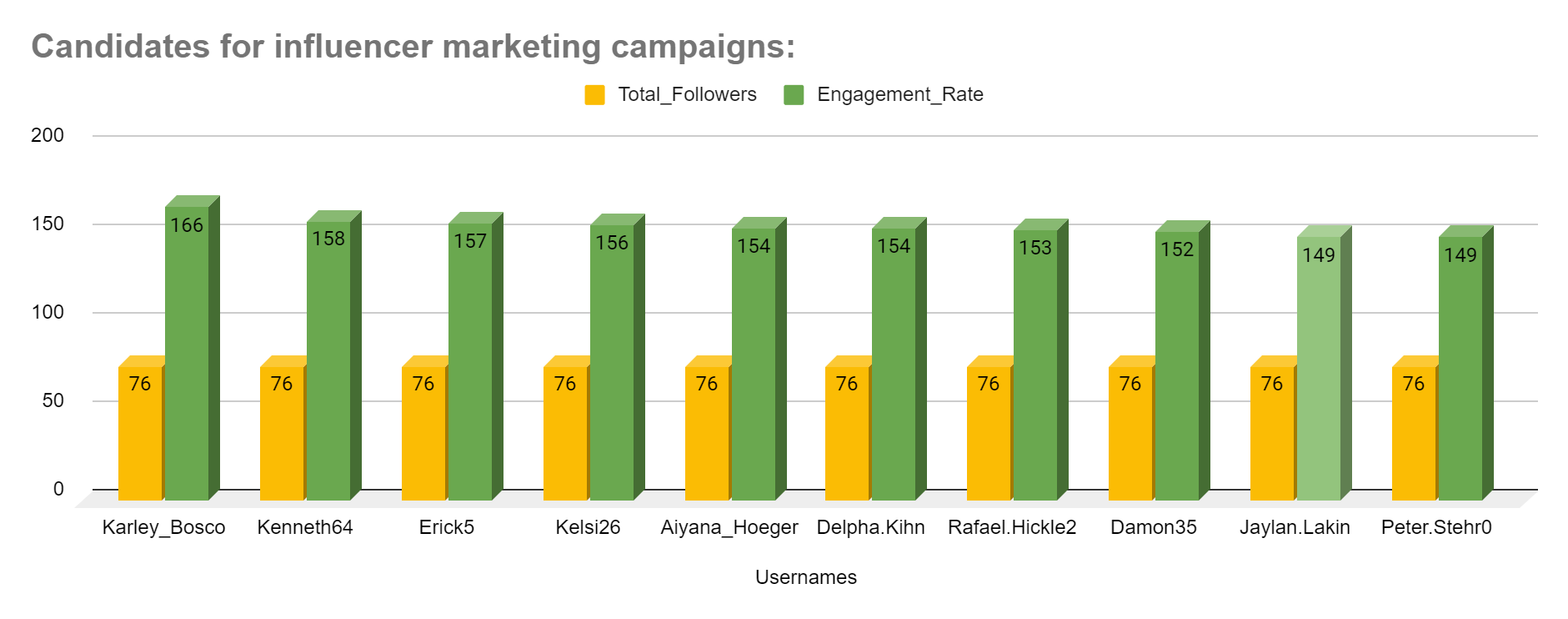
**—---------------------------------------------------------------------------------------------**

1. **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?**

**Answer:**

Based on follower counts and engagement rates, I found 10 users to be ideal candidates for influencer campaigns.





To collaborate effectively with influencers based on our metrics:

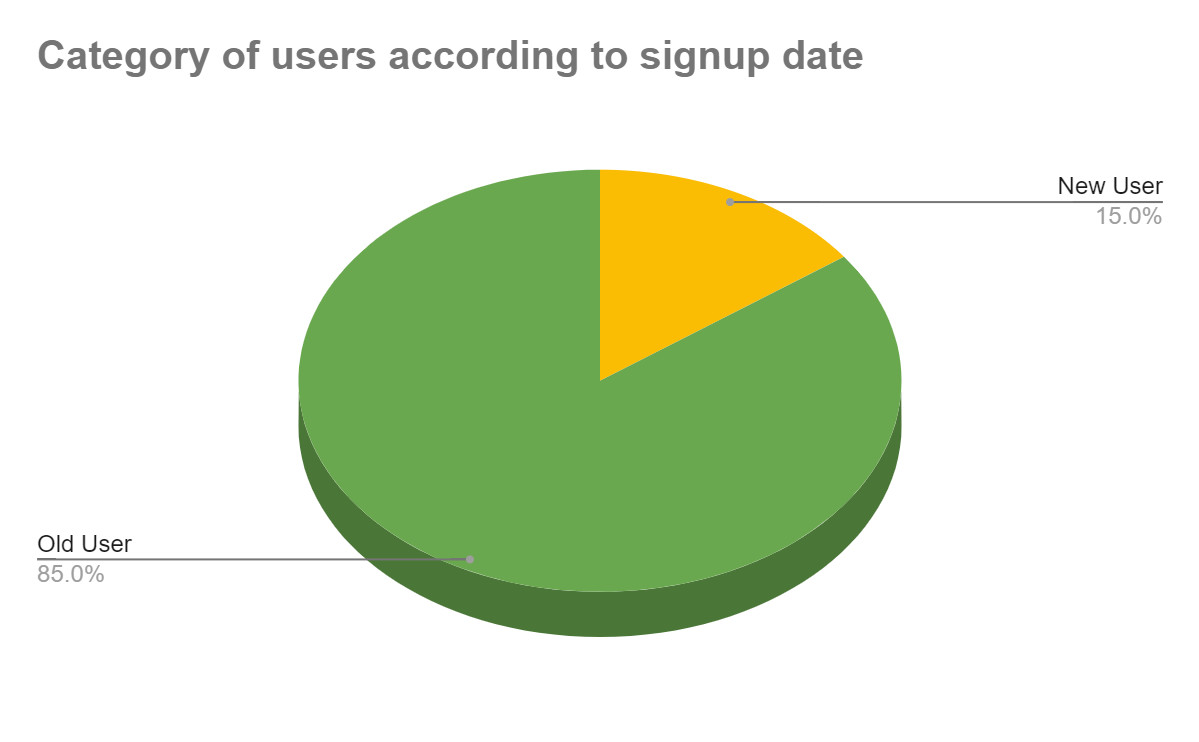
* We would first identify top performers by analyzing their engagement rates, follower counts, and activity levels. We would then reach out with personalized proposals, highlighting the mutual benefits of the partnership.
* Our approach would include setting clear objectives and deliverables, agreeing on compensation, and supporting the influencers with necessary resources. Throughout the campaign, we would monitor performance, provide feedback, and make adjustments as needed to ensure a successful collaboration.

**—---------------------------------------------------------------------------------------------**

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

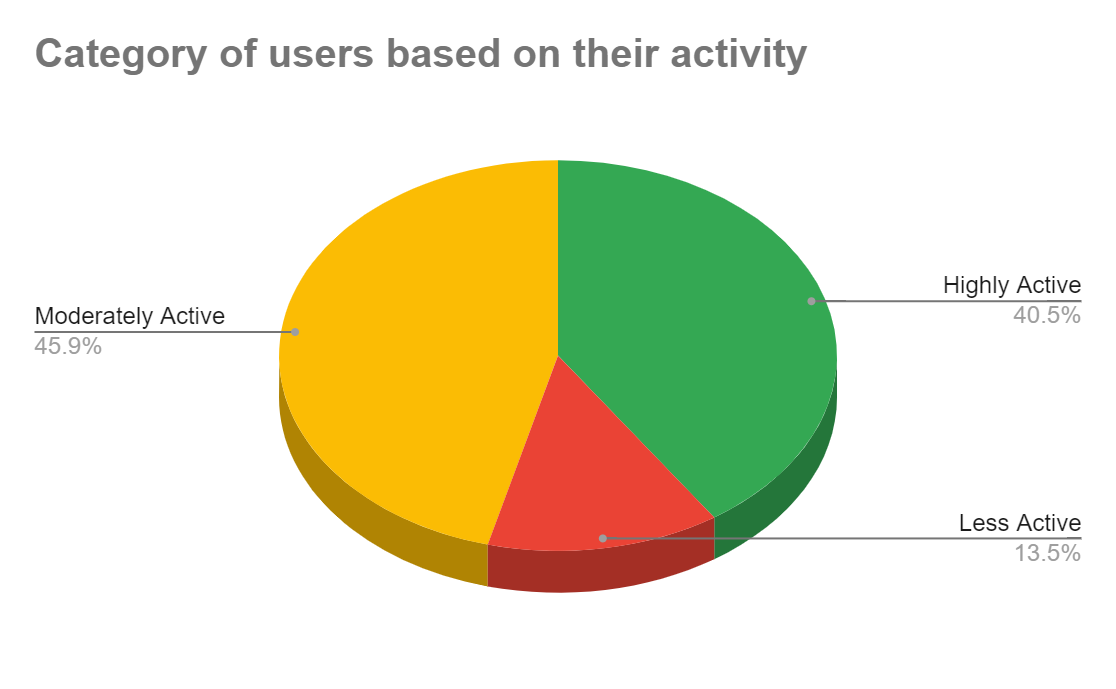
**Answer:**

For categorising old users and new users, by studying data I found that most recent date is “2017-05-04” by referring that I have categorised New Users and Old Users.



This helps us understand how long users have been with us, which can guide marketing and personalised recommendations.

For categorizing users activity level, users are categorized as Highly Active, Moderately Active, and Less Active



This segmentation helps in tailoring targeted marketing campaigns and personalized recommendations by identifying users who are more likely to engage with content, thus enabling more focused and effective marketing strategies.

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

Below is the pseudo code to measure the effectiveness of ad campaigns:

SELECT

campaign\_id,

impressions,

clicks,

conversions,

(clicks \* 1.0 / impressions) AS click\_through\_rate,

(conversions \* 1.0 / clicks) AS conversion\_rate

FROM AdCampaigns

WHERE (conversions \* 1.0 / clicks) < targetConversionRate;

* I would calculate the click rate which tells us how often people clicked ad after seeing it. This we can calculate by dividing the number of clicks by the number of impressions. A higher click rate usually means the ad is appealing to the audience.
* I would also calculate the Conversion Rate by dividing conversions by clicks. By this we can find out how many clicks actually lead to the desired action like purchase, etc. This helps assess if the ad effectively encourages people to take action.
* By looking data of impressions, we can also observe and analyse that which type of ad is getting liked by users most and by analysing it we can identify areas of improvements.
* For optimising future campaigns, we would target the right audience by observing factors like age, gender, etc for particular ad. We would have to keep our website easy to use after someone clicks the ad. Also we would focus on designs of different ads. By constantly checking what works and what doesn't, we can improve our ad results over time.

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

To identify potential brand ambassadors or advocates from user activity data:

* We can look for users who got more number of likes and comments. Which indicates the good quality of content user posting.

SELECT

u.id AS user\_id,

COALESCE(SUM(l.user\_id IS NOT NULL), 0) AS total\_likes,

COALESCE(SUM(c.user\_id IS NOT NULL), 0) AS total\_comments

FROM users u

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

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

GROUP BY u.id;

* We can look for the users with large number of followers because these users have a broad reach and can influence others.

SELECT

followee\_id AS user\_id,

COUNT(follower\_id) AS total\_followers

FROM follows

GROUP BY followee\_id

* We can analyse the content quantity like who is posting more number of post per week or month which indicates interest in platform.

SELECT

p.user\_id,

EXTRACT(MONTH FROM p.created\_dat) AS post\_month,

COUNT(p.id) AS total\_posts\_per\_month

FROM photos p

GROUP BY p.user\_id, post\_month;

* Also, users who regularly promote or support Instagram’s initiatives are ideal, as their consistent enthusiasm shows they genuinely care about the brand.

**—---------------------------------------------------------------------------------------------**

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

**Answer:**

Approach for this problem if question would have not provided:

* Track the time when users are most active with respect to their age group by analysing likes, comments , posting time and age group. By using these data we can schedule marketing posts and ad campaigns when users are more likely to engage
* We can segment users based on their interests, behaviours so that we can create tailored marketing campaigns that resonate with each segment to improve relevance and engagement.
* Identifying users with the high followers count, high numbers likes and comments for partnership with these influencers to reach a broader audience and enhance credibility.
* By analysing which type of content (images, videos, stories) receives the most engagement and most number of impressions so that we can use this insight to focus on creating high-performing content types.
* We can measure the performance of past marketing campaigns by analyzing impressions, clicks, and conversions and can use this data to refine future campaigns and allocate resources more effectively.
* By analysing segments of users(active user, non-active users) we can bring them back to the platform by sending customised messages or can increase their activity to the platform by offering memberships based on type of users.
* We can compare instagram data with other social media platforms (competitors) to identify opportunities and gaps. We would develop strategies to differentiate our brand and capitalise on market trends.
* We can improve ad performance by reaching users who are more likely to be interested in your products or services.

**—---------------------------------------------------------------------------------------------**

1. **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?**

**Answer:**

For updating "Engagement\_Type" to all instances of "Like" to "Heart" to align with Instagram's terminology we can use below query:

UPDATE User\_Interactions

SET Engagement\_Type = "Heart"

WHERE Engagement\_Type= "Likes";

For this question I have created a table named "User\_Interactions" and updated all the values of “Likes” with “heart” for column "Engagement\_Type".

Query Explaination: updates the User\_Interactions table, changing the value of Engagement\_Type from "Likes" to "Heart" for all records where the current Engagement\_Type is "Likes."

Table before updating: Table after updating:

