**Social Media Analysis**

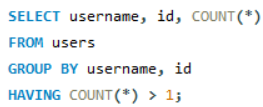
**Problem statement**

You are hired as a data analyst at Meta and asked to collaborate with the Marketing team. Marketing teams want to leverage Instagram's user data to develop targeted marketing strategies that will increase user engagement, retention, and acquisition. Provide insights and recommendations to address the following objectives.

**Objective Questions**

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

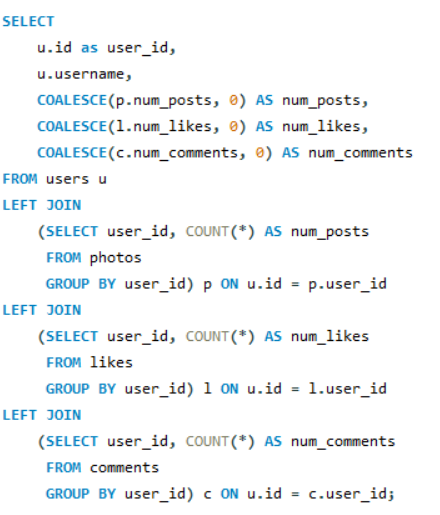
Query Snippet:



There are no duplicates or missing null values in any table, we checked it manually as well as by querying this query for every table by changing necessary selections for each and every table but there is no such missing data and null values present in the table

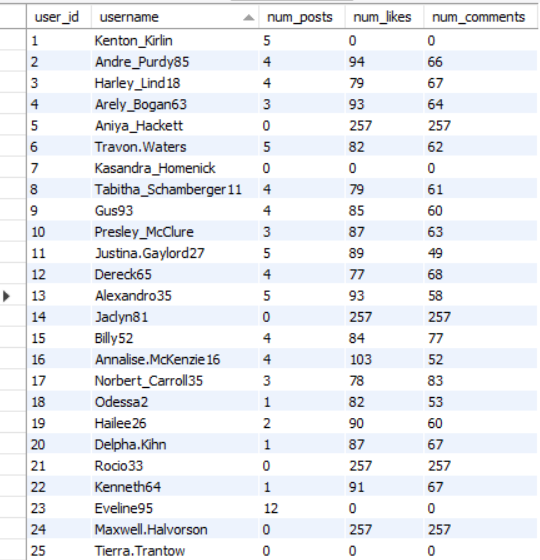
1. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?

Query Snippet:



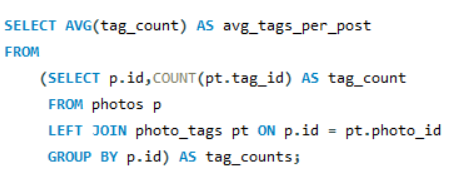
This query retrieves user information along with their activity data (number of posts, likes, and comments) from the users. The subqueries in the joins are calculating total\_likes, total\_comments,

total\_posts and, the final selection query is selecting user\_id, username, with total likes, comments and posts. **COALESCE** is used to counter null values and give 0 where no posts, likes, comments. This is the output.



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

Query Snippet:

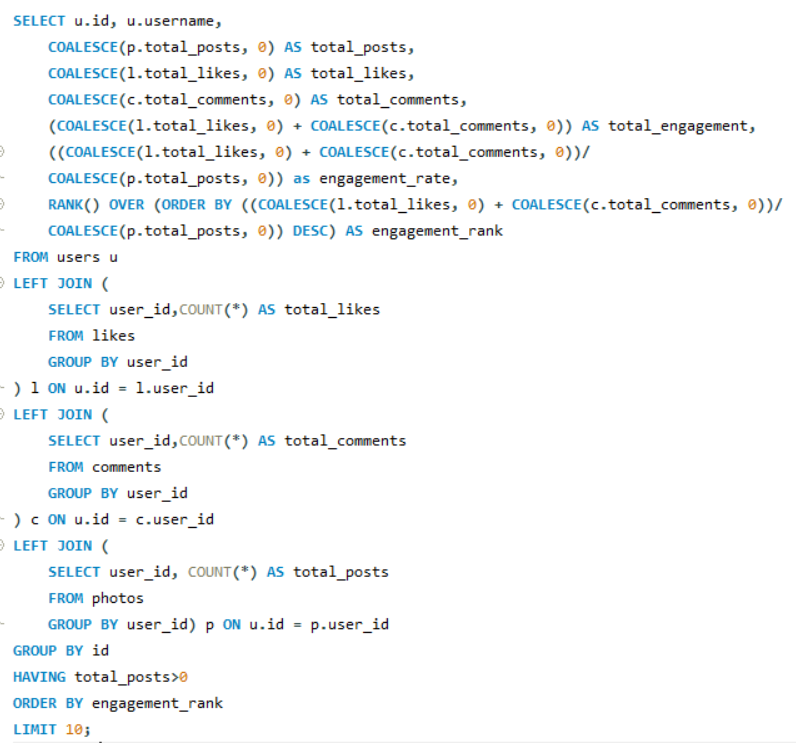


This query provides the average number of tags used per photo, in the subquery we selected posts and counted total tags associated with the tags as tag\_count and in the final query we calculated the average of that tag\_count.



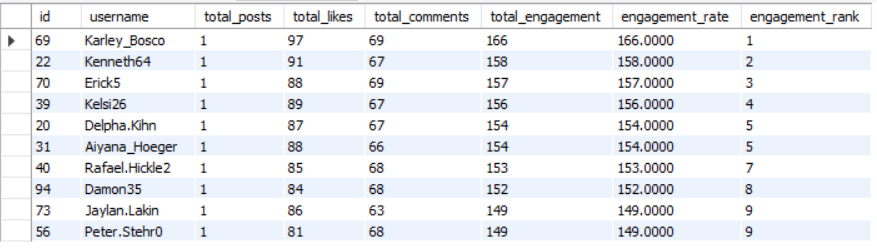
1. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.

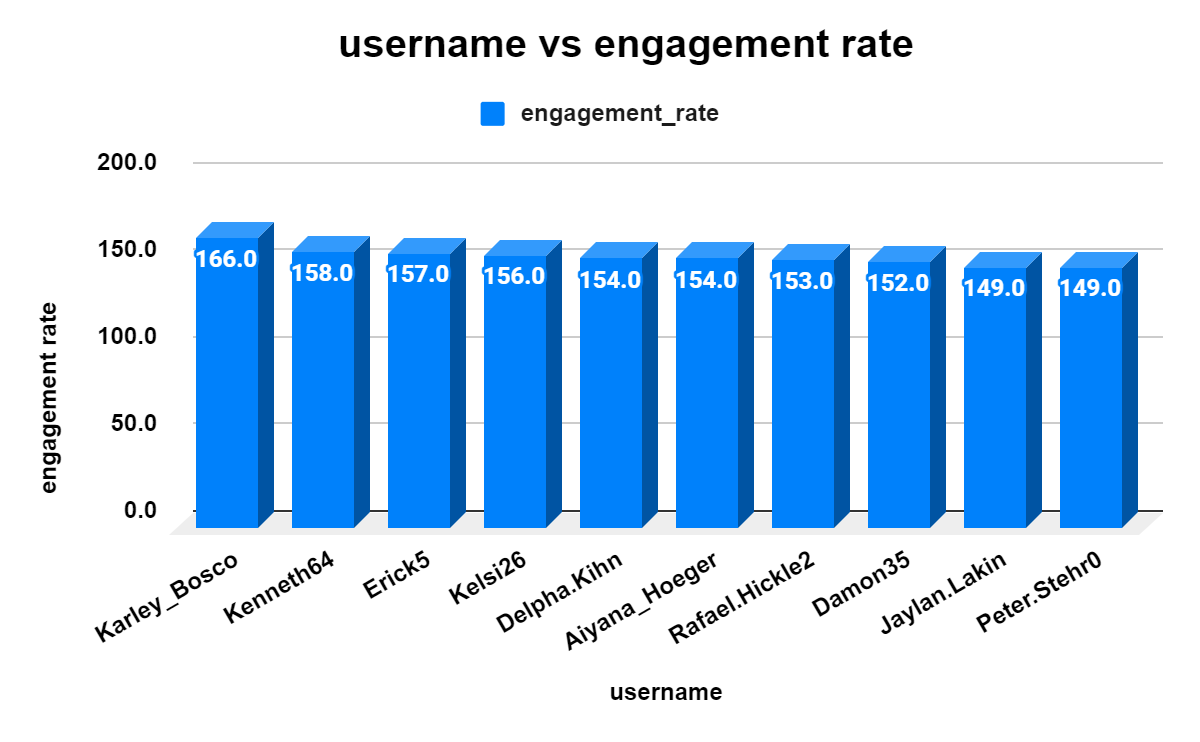
Query Snippet:



This query identifies the top 10 users with the highest engagement rates on their posts and ranks them accordingly using **RANK ()** function.

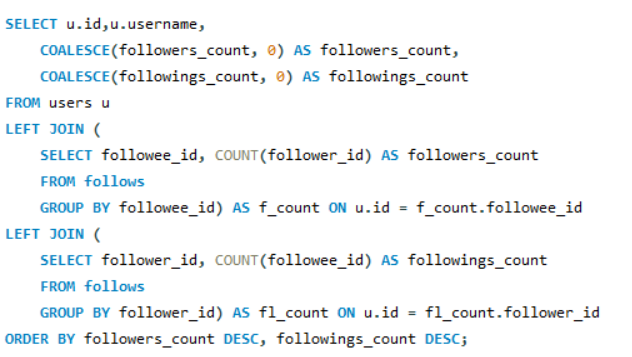
Here we calculated total\_posts, total\_likes, total\_comments in the join subquery and joined them on the users table, and in the main selection query selected total posts, likes, comments, total\_engagement, engagement\_rate by dividing total\_engagement by total\_posts and filtered those users who have more than 0 posts by HAVING clause.





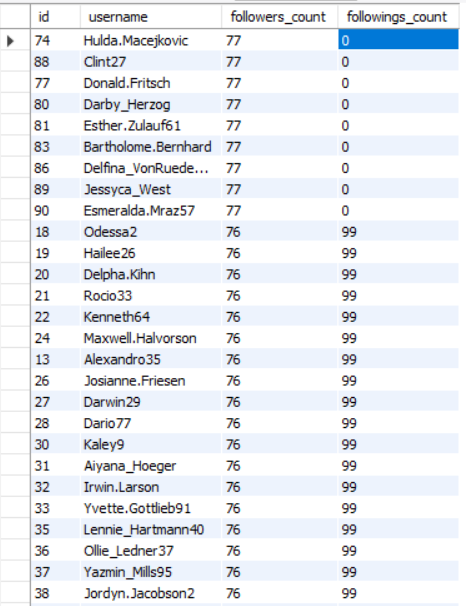
1. Which users have the highest number of followers and followings?

Query Snippet:



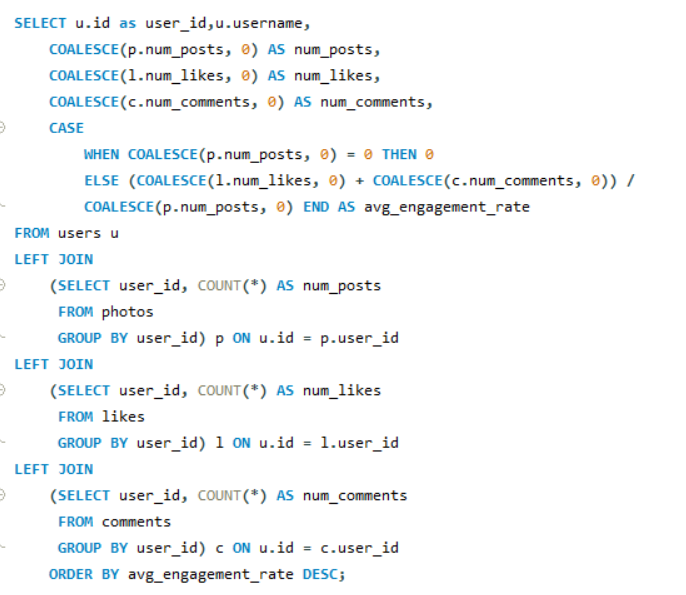
The query returns id, usernames and their follower and following counts, sorting them by the highest number of followers and then by the highest number of followings.

In the Joins we have created subqueries for calculating followers and followings count and then joined both on users table and in the main selection query we selected id, username, followers count and followings count, we used COALESCE on both followers count and followings count to negate with null values.



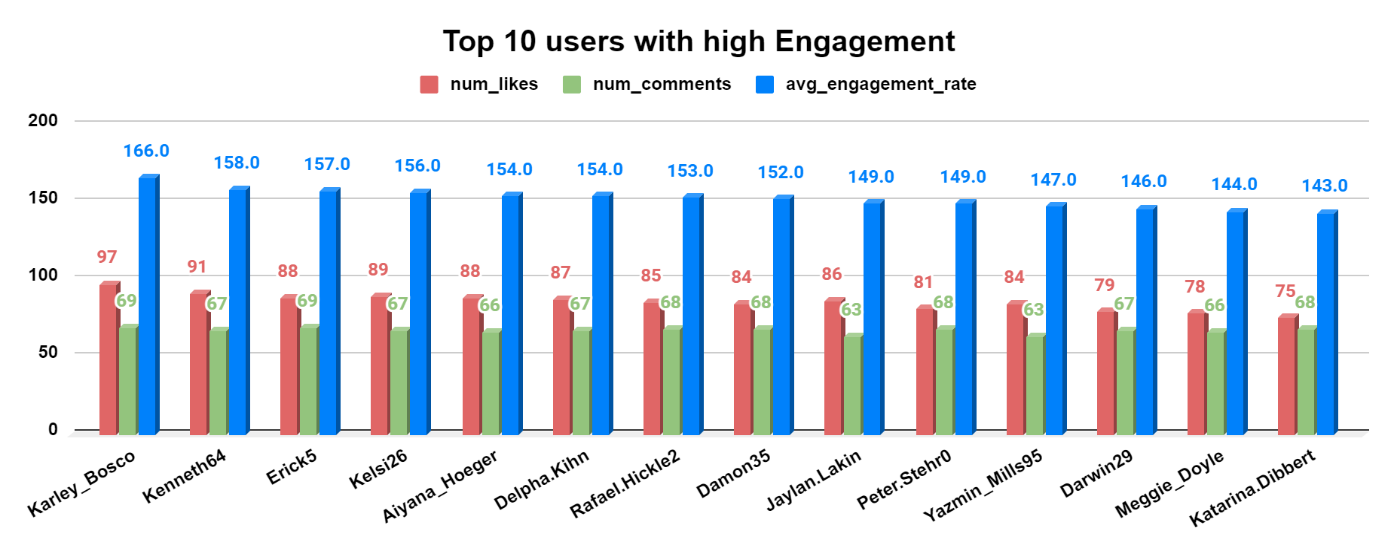
1. Calculate the average engagement rate (likes, comments) per post for each user.

Query Snippet:



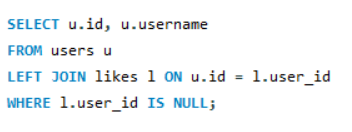
This query calculates and ranks users by their average engagement rate (likes + comments) per post, showing the most engaging users in the charts but in query output it is calculated for each user.

We are selecting the number of posts, likes and comments from their respective tables in the subquery inside joins and joining them all on users table and in final selection we selected id, username, number of posts, likes, comments and average engagement rate as sum of number of posts and likes divided by total posts. if there is no post engagement rate is set to be 0 to avoid division by 0, and sorting is done on avg\_engagement\_rate in descending order.



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

Query Snippet:



The query lists users who have not liked any posts, effectively identifying inactive or less engaged users in terms of liking content.

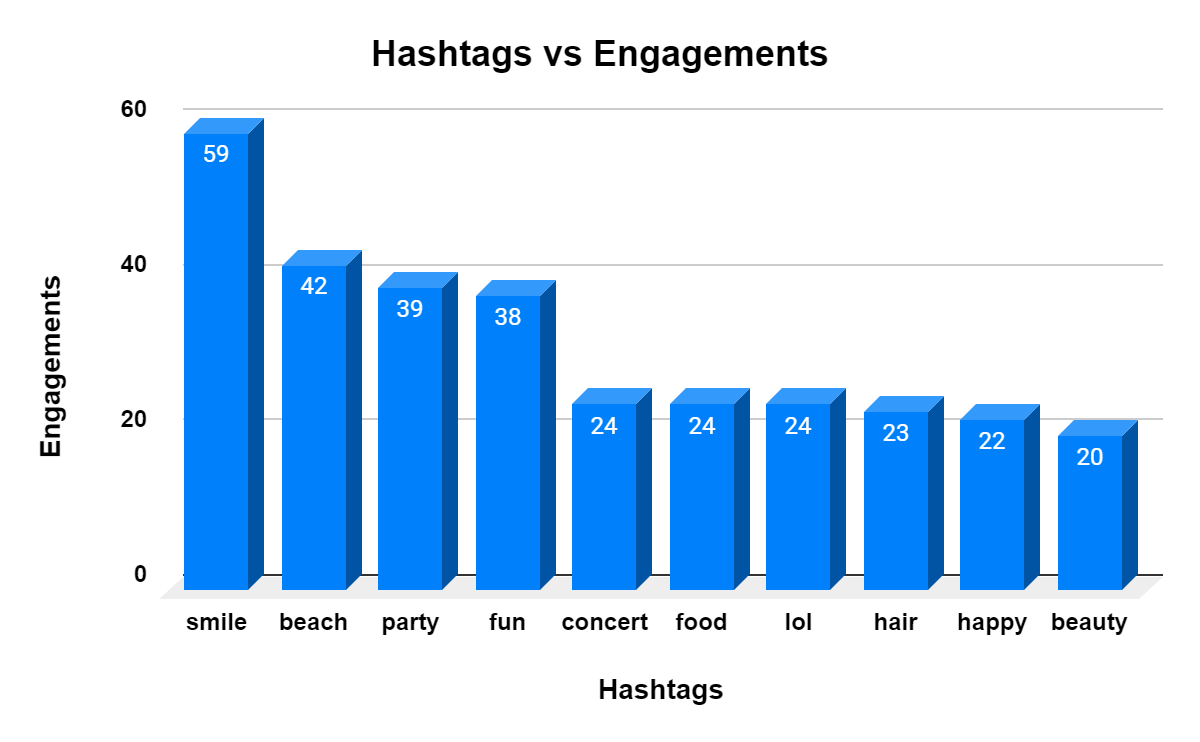
Selecting id and username from users table and joining likes table on user table with user\_id and filtering condition where l.user\_id is null

The output table is given below with the list of users who have never liked any post.



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

Reference Image:



* We can analyse the popular hashtags used in the posts, and then we will see the trending topics related to the tags.
* Segment the users on the basis of hashtags and their interests to make more personalized ad campaigns.
* Segment users into different interest clusters based on the types of content they generate (e.g., travel, fashion, fitness).
* Understand user behaviour, such as what type of content they engage with the most, to run ad campaigns that match with their interests.
* Use the tag analysis to create personalized ad content that aligns with users' interests. For example, if a user frequently tags fitness-related content, show them ads related to fitness products.

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?

* **User Activity Levels:**
* **Photos**: Users might engage more with photos due to their quick visual appeal. Higher likes but moderate comments could indicate this.
* **Videos**: Longer content like videos might attract more comments and shares, reflecting deeper engagement.
* **Reels**: Quick, engaging reels may lead to a high number of likes and shares, indicating viral potential.
* **Specific Contents:** 
  + **Frequent Users**: Users who post more often may prefer certain content types, like reels for quick engagement or videos for storytelling.
  + **Casual Users:** Users with lower activity may lean towards consuming content like photos or short videos but not posting as much.

Yes, there is correlations between user activity levels and specific content type because if more photos, videos and reels are getting engagement then it means user activity levels are increasing and if engagement is increasing means the specific type of content is also increasing, so according to the rule if both thing increases it means there is correlation.

This information can guide us for content creation and curation strategies by many ways such as:

* Post mixed content and see what is getting higher engagement, Post them more.
* Use trends and hashtags related to trends for better engagement on posts.
* Recommend content according to the users interests and follow cross engagement strategy like watch a reel just after liking a photo to diversify user engagement.

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

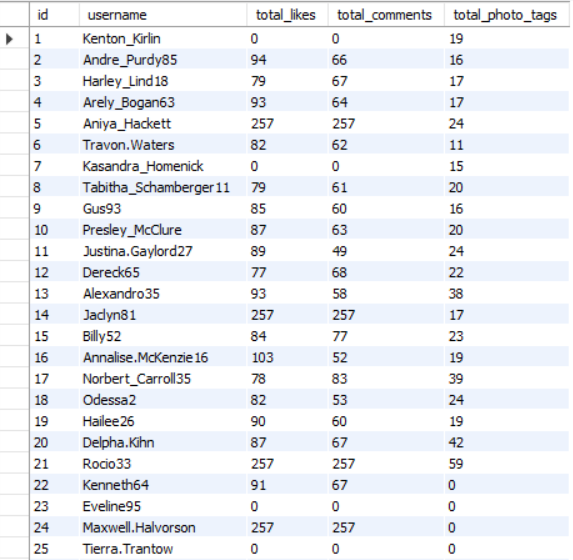
Query Snippet:

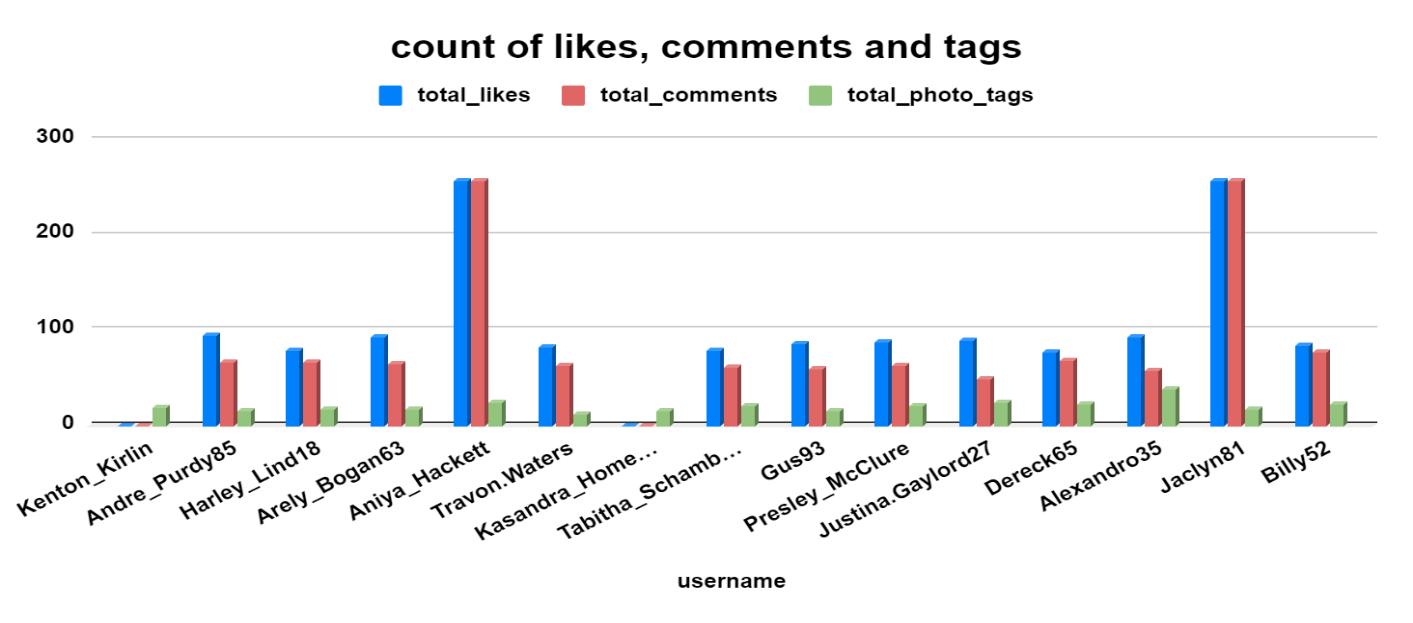


This query gives user\_id, username along with the total number of likes, comments, and photo tags for each user.

We have written a subquery for the total likes, comments and photo tags from their respective tables and joined all tables and then selected id, username, total\_likes, total\_comments and total\_photo\_tags avoiding null values for all these using COALESCE.

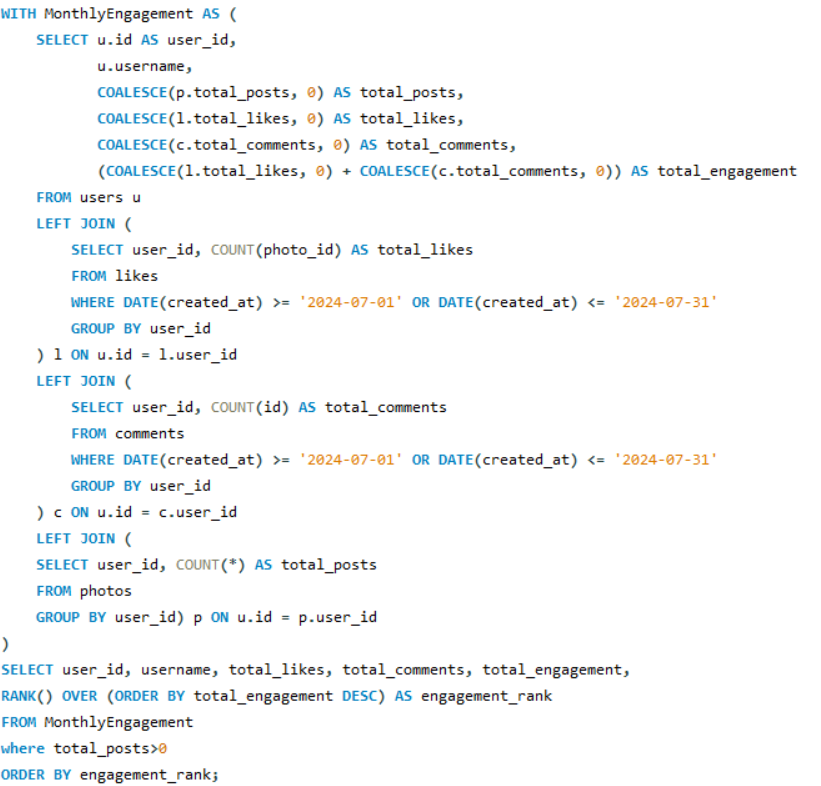
This is a snippet of top 25 users but in output the query is giving it for each user.





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

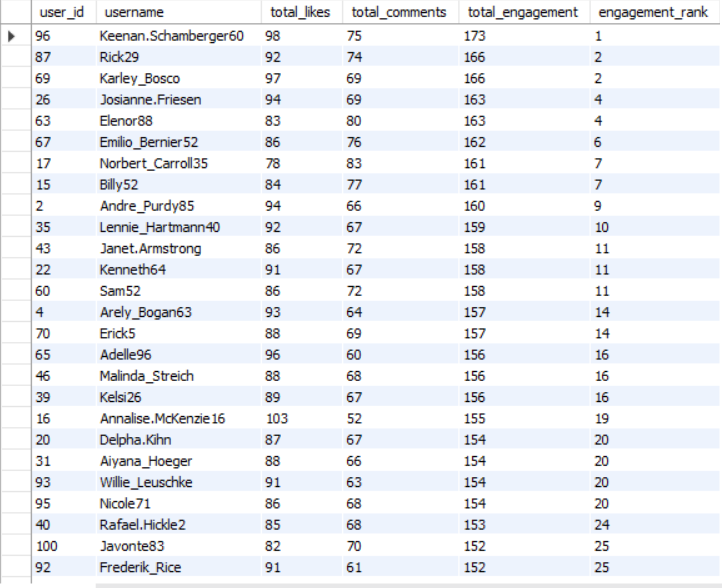
**Query Snippet:**

****

The query will calculate and rank the total engagement in the month of July 2024.

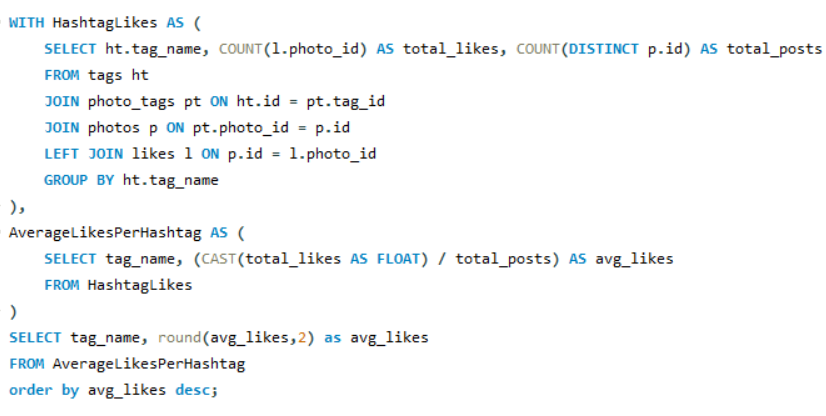
The main cte is calculating total likes, comments, total posts from subquery of join and all are joining on user table, and then selecting id, username, total likes, comments and total engagement from the main cte and filtered for no posts and ordered by engagement rank.

The output snippet here is for a few users here but it's there for each user in the query output.



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.

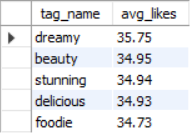
Query Snippet:

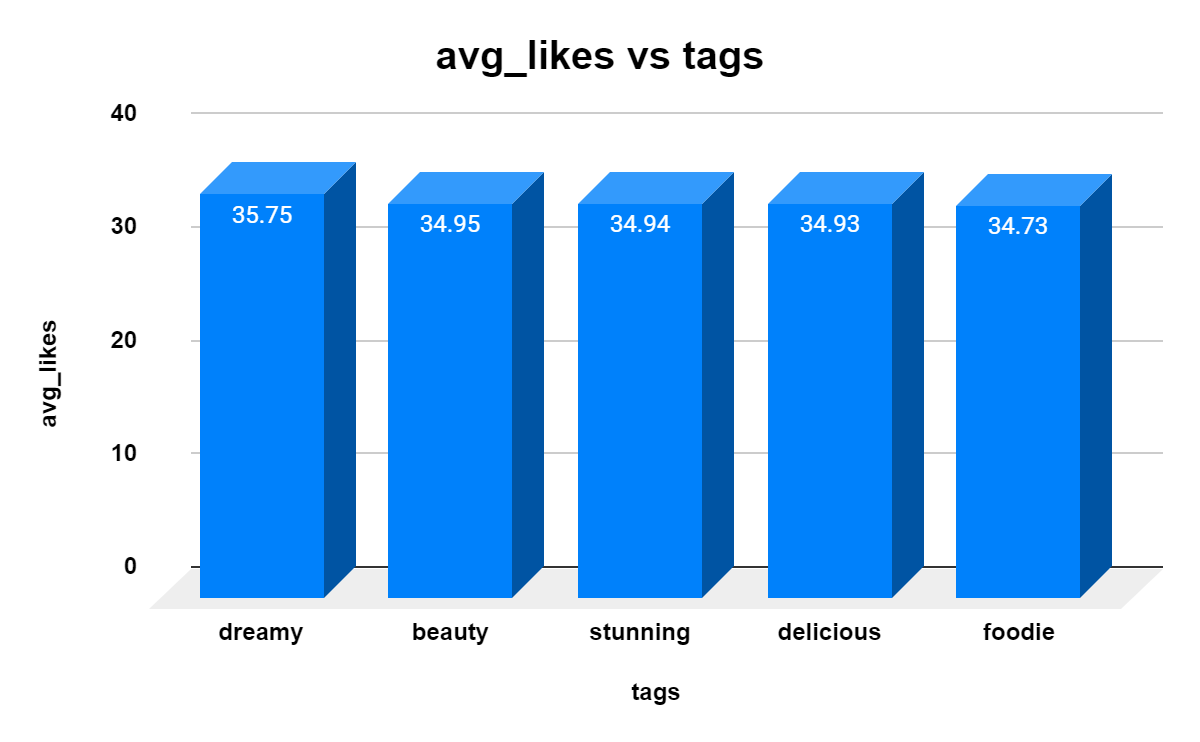


This query is calculating average no. of likes per post for different hashtags. In the first cte we are calculating likes and posts per hashtags using tags table and in second cte we are calculating average likes per hashtags.

And in the end, we are selecting hashtags and the average likes and rounding it off two decimal places and ordering the result by average likes in descending order.

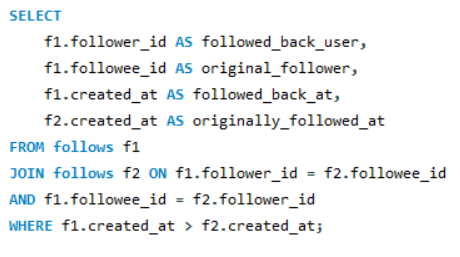
These are the top 5 hashtags but in the output all are present.





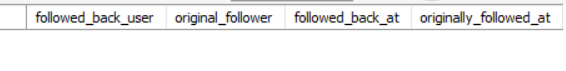
1. Retrieve the users who have started following someone after being followed by that person

Query Snippet:



This query is retrieving the users who have started following someone after followed by that user but there is no one doing the same because the data we have in that the created at time in the follows table is same

so, it is retrieving nothing as per the output snippet below.

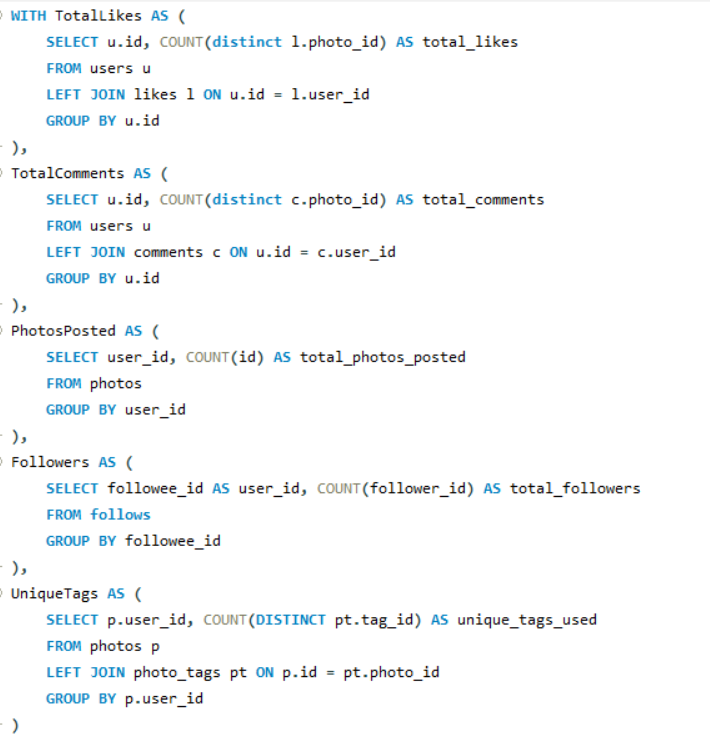


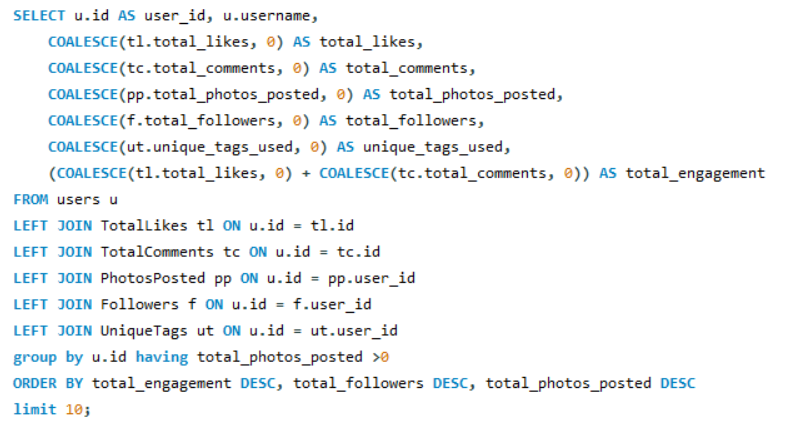
**Note: Subjective Questions from next page.**

**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?

Query Snippet:



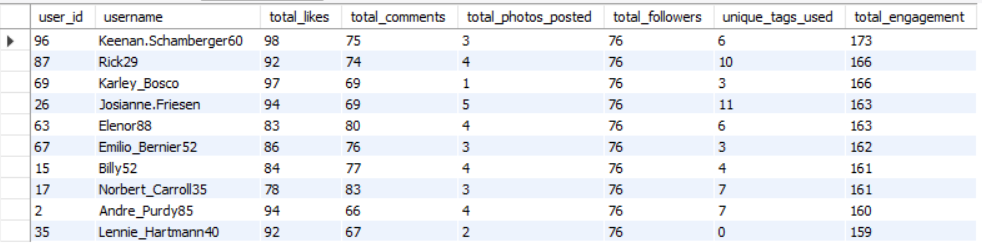


Insights:

The SQL query mentioned in the file identifies the most loyal or valuable users based on their engagement and activity levels. It calculates various things, such as the total likes, total comments, total photos posted, total followers, and unique tags used by each user. The final selection filters out users who have not posted any photos and ranks the remaining users by their total engagement, followers, and photos posted.

We have filtered the top 10 users only on the basis of the total engagement.

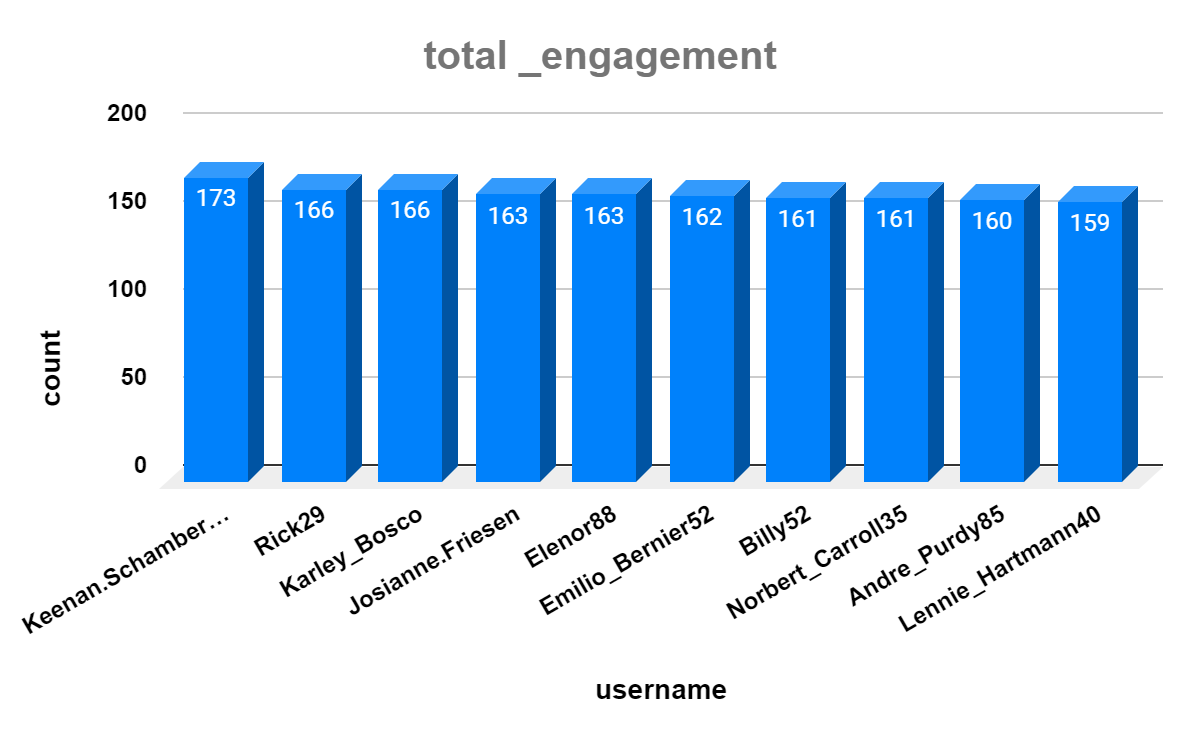
* **total\_engagement** (sum of likes and comments) is a good measure of user activity and interaction with content.
* **total\_followers** reflect the user's influence or popularity within the platform.
* **total\_photos\_posted** indicates the user's content contribution.
* **unique\_tags\_used** shows the diversity of the content, which might correlate with the user's creativity or engagement across different topics.



**Recommendations:**

**Suggestions for Rewarding or Incentivize these top Users:**

* **Offers:** Send offers or discounts based on their interests and engagement patterns.
* **Early Access:** Involve top users in beta testing new features or products, making them feel like valued contributors to the platform's development.
* **Reward Points:** Introduce a reward system where the most loyal users can earn points that can be redeemed for discounts or other monetary benefits.



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

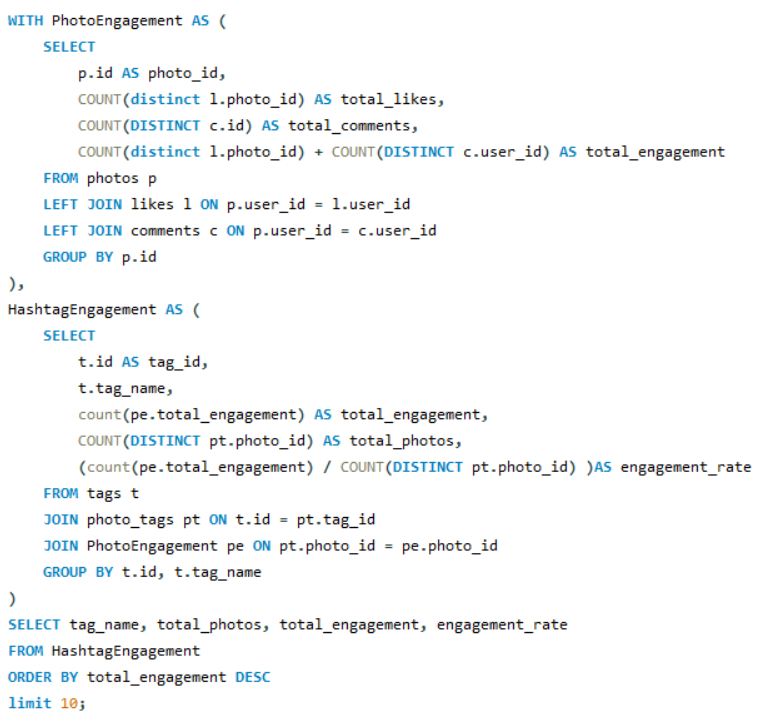
**Recommendations:**

We should focus on the points which are discussed below. May it could help in coping up with the problem of re-engaging the inactive users and it could also help in encouraging people to start posting or engaging again.

* **Notifications:**
* **Interactive Notifications**: Allow users to take action directly from the notification, such as liking a post or joining a conversation or such as swiggy and blinkit make their notifications interactive to catch their users back.
* **Timely Reminders**: Send push notifications that remind users of upcoming events, new features, or content they’ve expressed interest in.
* **UI/UX Improvements:**
  + **Interactive UI**: Make the user interface Interactive, Dynamic and smooth in the best possible way, there should be no glitches in any process performed by users.
  + **Updates**: If there have been significant updates or changes to the platform, guide inactive users through a quick onboarding process to familiarize them with new features and benefits.
* **New Trend Notifications:** 
  + **Highlight Trending Topics**: Showcase trending topics that might interest them, emphasizing the social aspects of the platform they’re missing out on.
  + **Friend Activity Notifications**: Notify users when their friends or connections are active on the platform, encouraging them to rejoin the conversation.
* **Ad Campaigns** 
  + **Customised Ads:** Show ads according to the type of content they are engaging with mostly.
  + **Targeted Email Campaigns**: Send personalized emails or notifications reminding users of what they’ve missed, including updates, new features, or content that aligns with their past interests.

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

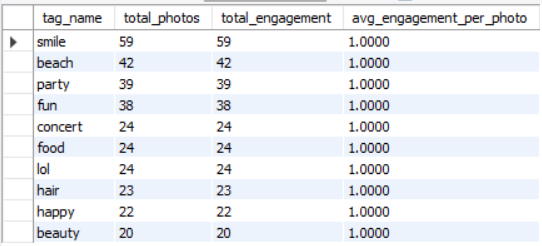
Query Snippet:



Insights:

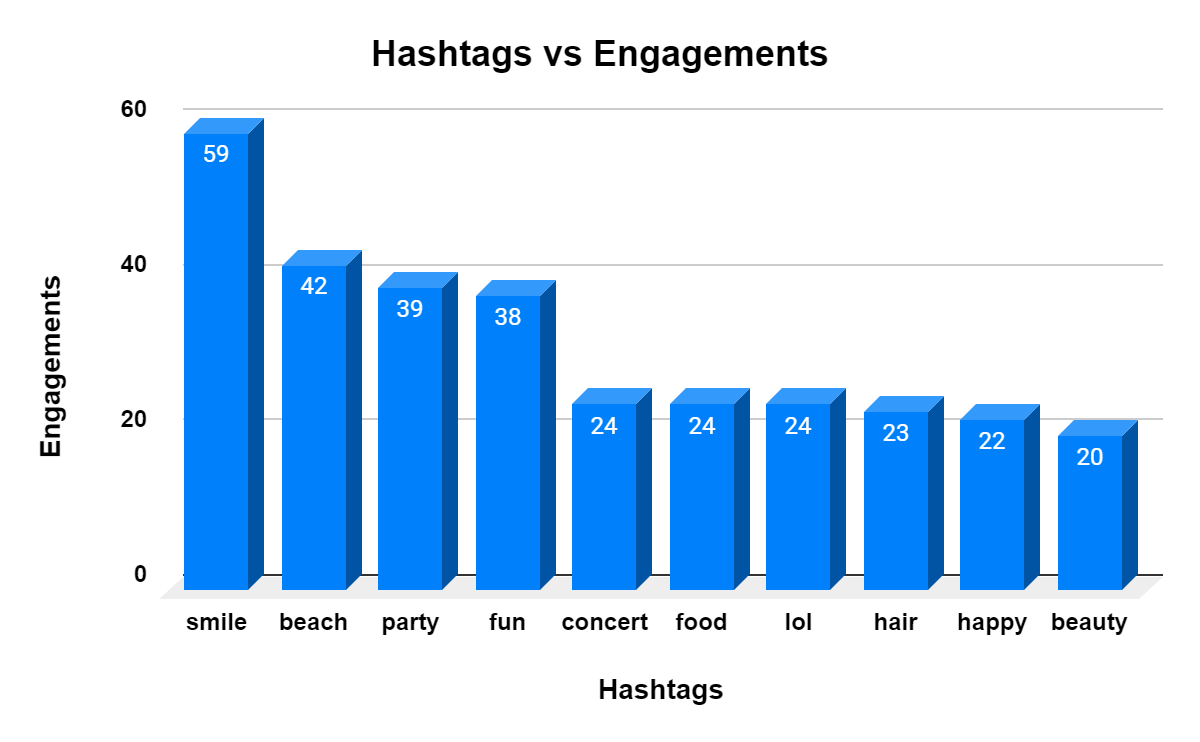
This Query is giving hashtags with highest no. of engagement (likes and comments) by:

* + **Calculating Engagement**: Summing up likes and comments for each photo.
  + **Aggregating by Hashtag**: Linking these engagement queries to hashtags used in those photos.
  + **Engagement Rate**: Total engagement divided by total photos with that hashtag.
  + **Ranking Hashtags**: Ordering the hashtags by total engagement and calculating the average engagement rate per photo for each hashtag.



**Recommendations:**

* **Content Strategy**:
* **Prioritize High-Engagement Topics:** Focus on creating content around the hashtags that have the highest engagement. This could mean posting more content related to these topics or encouraging users to tag their posts with these popular hashtags to boost visibility and interaction.
* **Content Optimization**: Analyse why certain hashtags are performing well and replicate those elements in future content, such as the type of content, timing of posts, or accompanying text.
* **Ad Campaigns:**
* **Targeted Advertising**: Use the high-engagement hashtags in your ad copy, targeting users who are already interested in these topics. Ads featuring popular hashtags are likely to attract more attention and clicks.
* **Influencer Marketing**: Partner with influencers who are already using these high-engagement hashtags. Their content, when aligned with these trending topics, can amplify your campaign's reach and effectiveness, leading to higher engagement and conversions.



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?

Query Snippet:



Insights:

Since demographics aren't available and posting times are the same for every post, we should focus on the engagement patterns across different hours of the day and days of the week. We got to know that the post timing was around 11 and day of the week was 6 for each and every post, there were 257 posts at that time.

**Recommendations:**

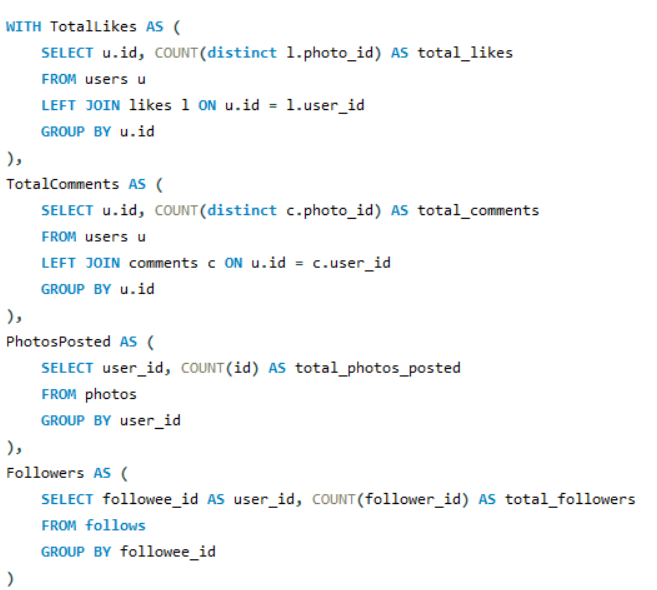
We can see these points for identifying if the data was available.

* + Identify hours of the day when users are most active (e.g., posting, liking, and commenting). This could reveal peak engagement times.
  + See if there are specific days of the week with higher engagement. For example, weekends might see more activity as users have more free time.
  + Also see if there are times or days with lower activity. knowing these off-peak periods can help in deciding when to post content to avoid overcrowded timelines.



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?

Query Snippet:



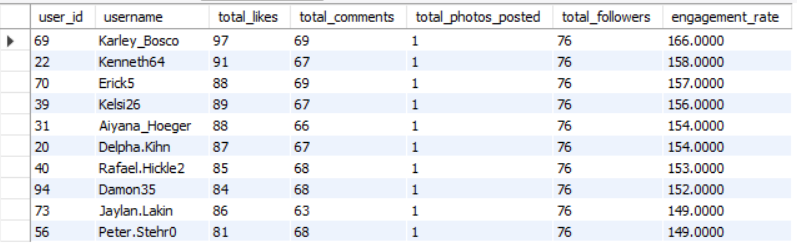


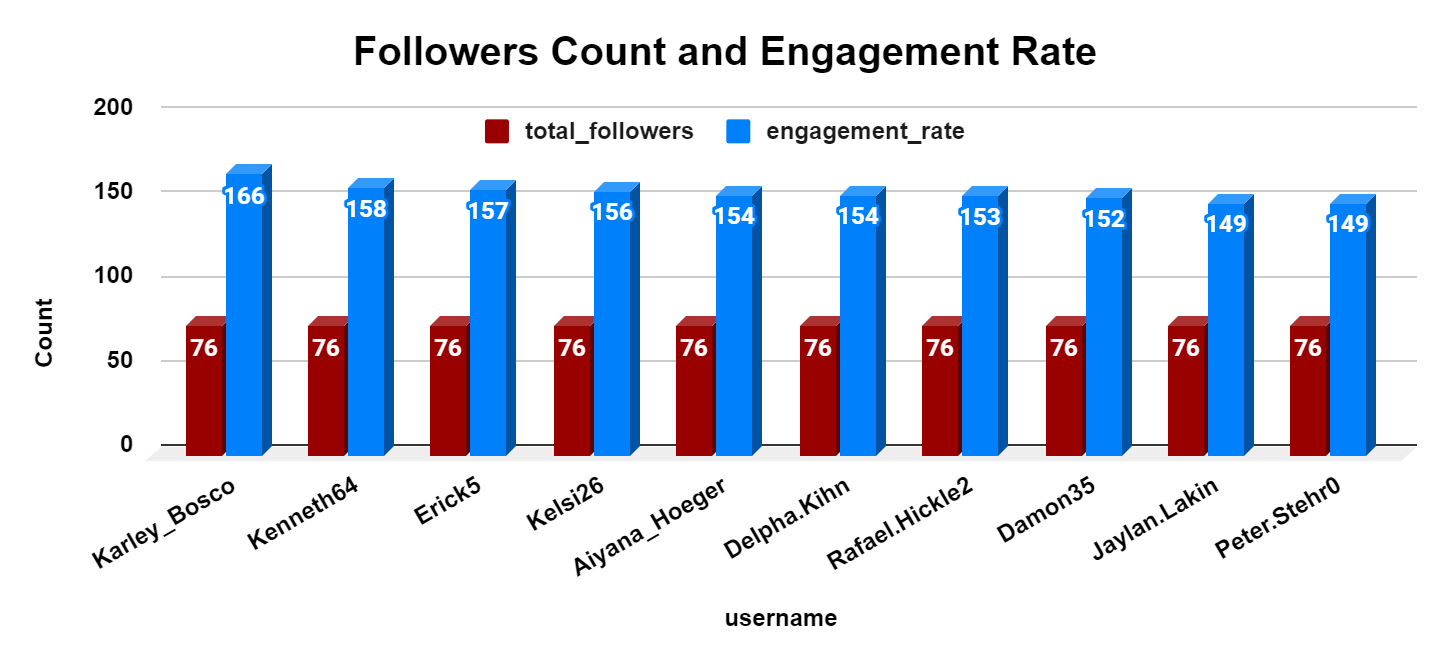
Insights:

The Query will return the ideal candidates for influencer marketing campaigns based on their follower counts and engagements rates.

We have calculated total likes, total comments, total photos posted, total followers and we have calculated the engagement rate by summing up total likes and comments and dividing by total photos posted.

At the end we filtered the query for users who have at least one photo and Users are ordered by engagement rate (from highest to lowest), followed by total followers, and then total photos posted. The top 10 users are selected.



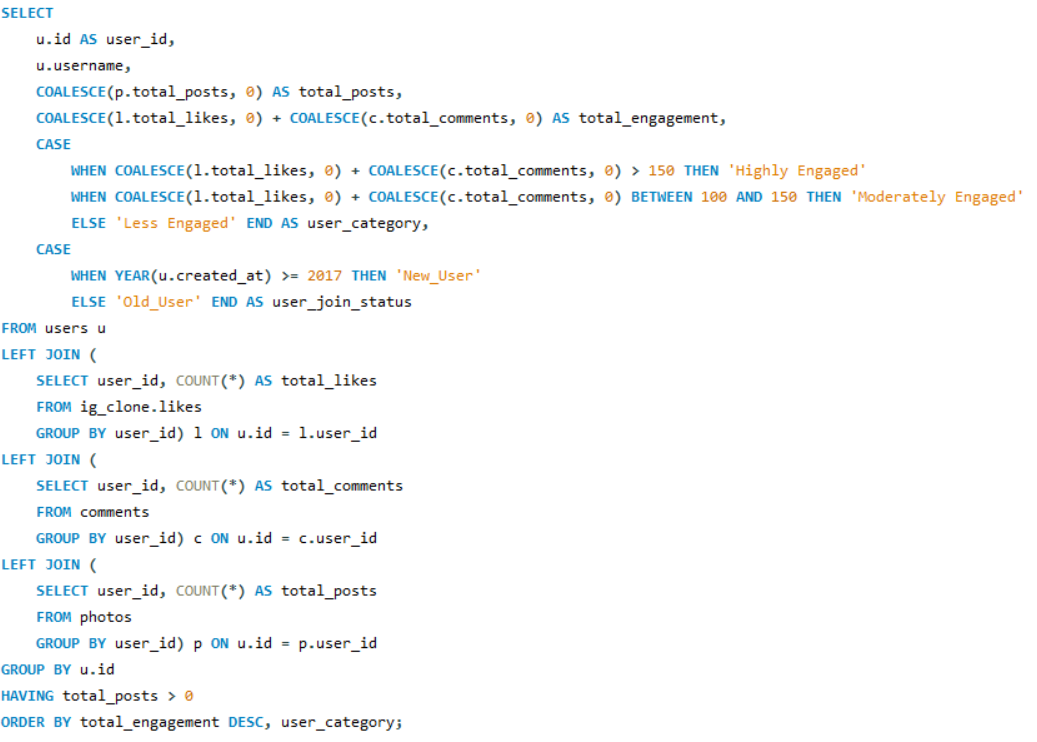


**Recommendations:**

* We have to Approach these users by personally messaging them and why we want to collaborate with them, and we have to offer clear value to these influencers, such as compensation, exclusive products, or opportunities to co-create content.
* And Work with influencers to create authentic content that matches with their personal brand and engages their audience. This could include product reviews or sponsored posts.
* Involve influencers in broader marketing campaigns, giving them creative control to ensure the content has meaning for their followers while promoting your brand effectively.

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

Query Snippet:

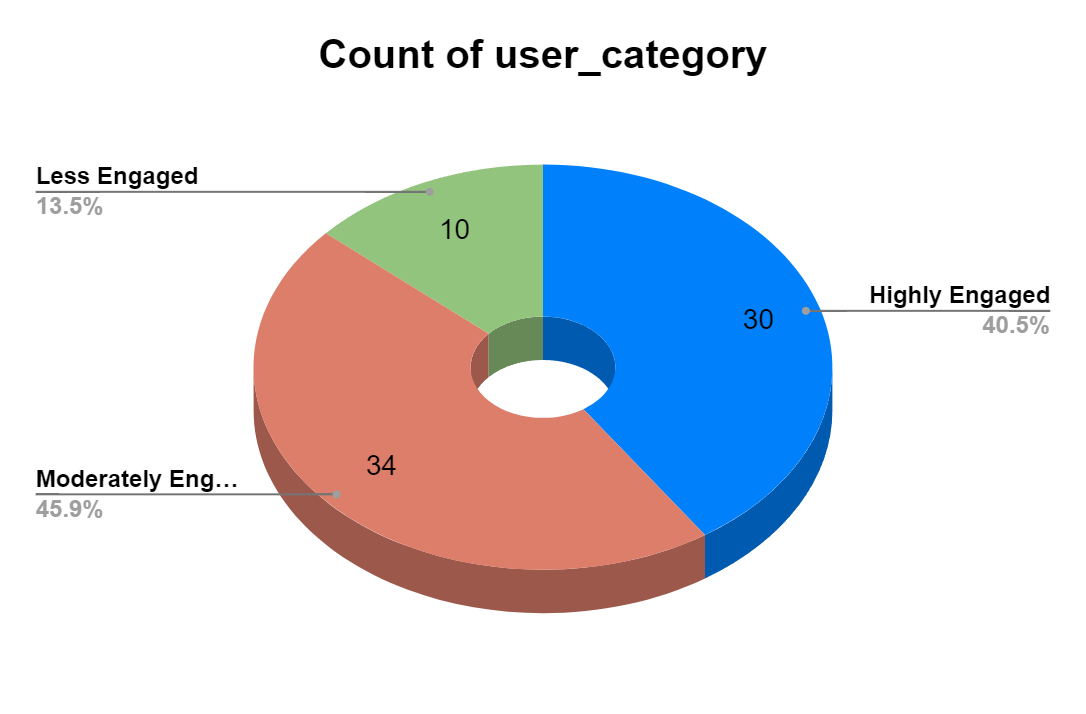


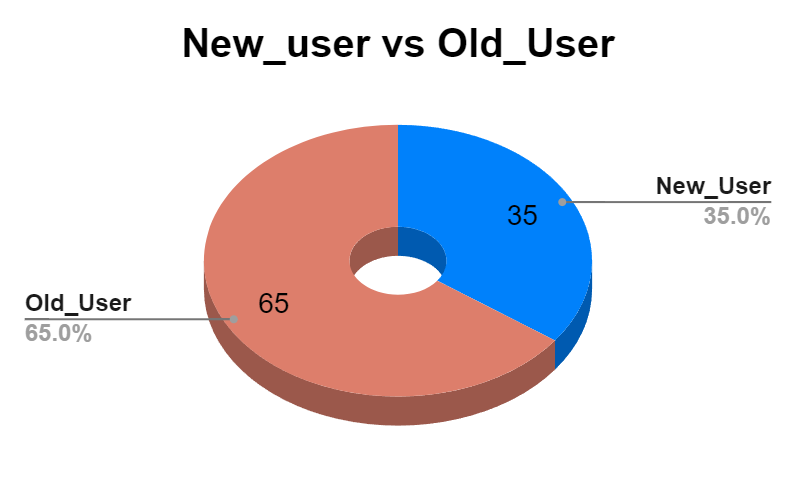
Insights:

This Query will give you the analysis of user engagement as well as tenure on the platform, meaning from how long they are using this platform. You can identify which segments of users are highly engaged and relatively new, and develop campaigns that specifically address their behaviours and needs.

In the query we have selected user\_id, username, total\_posts, total\_engagement, User Category on the basis of total\_engagement and User Join Status means from when the user is using this platform.

We **Join** the users table on likes, comments and photos table and use **CASE** statement to create **segments** for user engagement and join status and in the end **group by** user\_id and filtered out for no posts using **HAVING** and ordering the result by total engagement and user category using **ORDER BY** clause.





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

Insights:

If data on ad campaigns (impressions, clicks, conversions) is available

* + **id**: Unique identifier for each ad campaign.
  + **impressions:** Number of times the ad was shown.
  + **clicks**: Number of times the ad was clicked.
  + **conversions**: Number of completed actions that resulted from the ad.
  + **cost**: Total cost spent on the ad campaign.

We will calculate:

* + **cost\_rate:** (clicks / impressions) \*100**,** this will help us to get insights that ad is generating interest if the cost\_rate is high
  + **conversion\_rate:** (conversions / clicks) \* 100, this will also help us to know about the conversions according to the ads if the rate is high means the ad is doing good.
  + **cost\_per\_clicks:** cost / clicks, it will give us the amount which we are spending on one ad click.
  + **ROI:** (conversions \* revenue\_per\_conversion - cost), if the **Return On Investment (ROI)** is positive that means the ad is doing good.

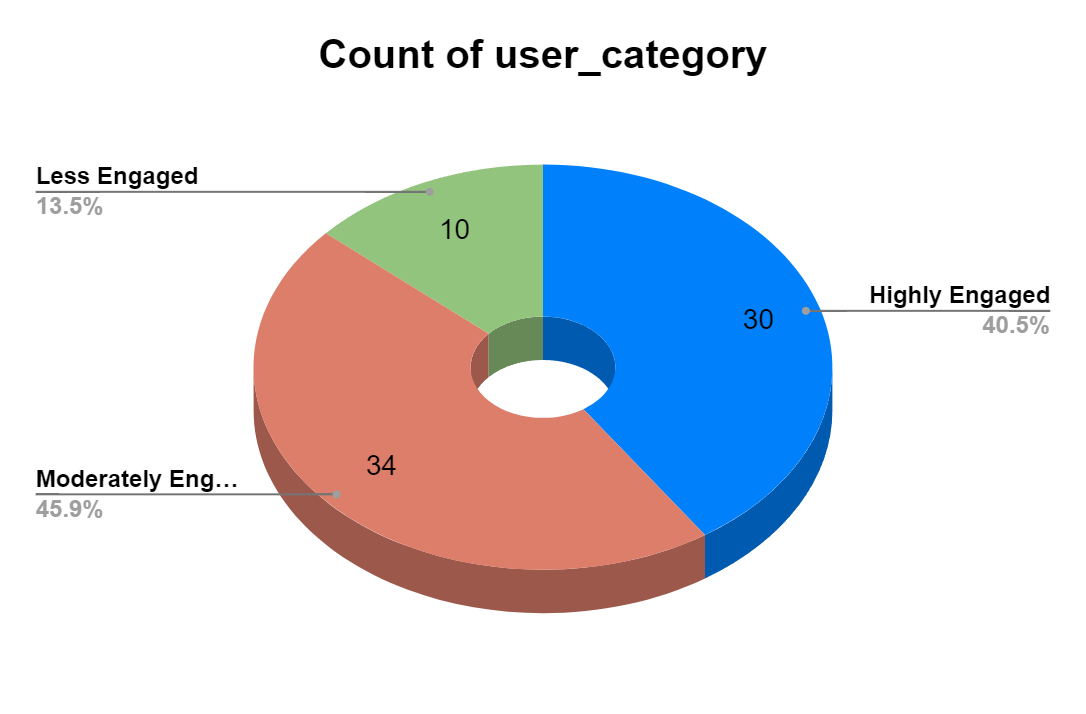
**Recommendations:**

We can optimize future ad campaigns by taking few things like that:

* Analyse those ad campaigns which have higher ROI and conversion rate then try to run more campaigns like that.
* Analyse those ad campaigns also which have lower ROI and conversion rate then focus on why these ads are not working, which aspects are missing here and improve that part and optimize those ads also.
* At last spend more cost on the low performing ads and make them work and this will make campaigns more effective.

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

Reference Image:



Insights:

As we have the user activity data we have calculated earlier about the Engagement, Number of followers, User Join Status, Unique tags so now we will see these aspects to identify potential brand ambassadors.

We have the categories **highly engaged**, **moderately engaged**, and **less engaged** so we will now see which of these users are highly engaged, have more followers and are posting frequently, we will prioritize these users as potential brand ambassadors or advocates.

**Recommendations:**

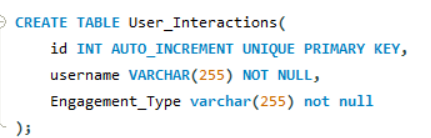
* Send a personal message appreciating their engagement and inviting them to participate in a specific campaign or event.
* Provide exclusive access to features, events, or content, or offer money in exchange for promotion.
* Maintain long-term partnerships with highly engaged users, make them official brand ambassadors for ongoing events.
* Work with them to create content that matches with both their style and **Instagram’s** brand, it will increase authenticity and reach.

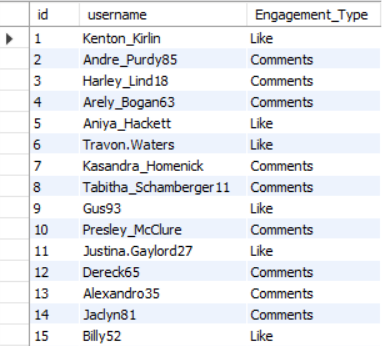
1. How would you approach this problem, if the objective and subjective questions weren't given?

* First you have to analyse what are the broader goals of the marketing team here. They are trying to increase engagement, re-engage inactive users and attract new users.
* After that check the type of information you can collect for analysis such as number of followers, likes count, comments count, age group, type of content.
* **Segment the users** based on their followers count and engagement (likes and comments).
* Analyze the type of posts they are posting, what tags they are mostly using and what are their preferences.
* Analysing users who are posting frequently and who are not posting, who are not posting give them recommendations of the latest **trends** to **encourage them to re-engaging** and continue posting again.
* Users who are new or just joined guide them properly about what they can do, make the **User Interface** smooth and interactive.
* **Content driven strategies** for users among the different categories, suggest content related to their interests.
* Develop **target marketing strategies** via **personalised content recommendations**, **email campaigns**, **collaborations.**
* Segment users for different campaigns according to their needs.
* **Monitor** the campaigns whether it is doing good or not for effectiveness.
* Take **user feedback** from the users for better suggestions and implement it for effective engagement.

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?

We created a table User\_Interactions by this query and inserted data into it and in Engagement\_Type column we added Like and Comments randomly





After that as per the question says we have to update Engagement\_Type column’s Like as ❤ ️ so we updated that by this query.

