

**Social Media Analysis**

**By: Nitin Kumar**

**Problem Statement**

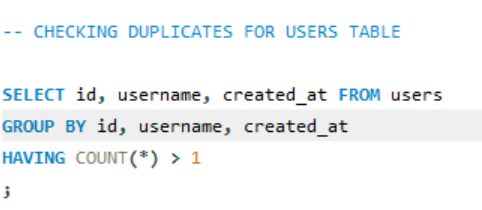
You are hired as a data analyst at Meta and asked to collaborate with Marketing team. Marketing teams wants to leverage Instagram's user data to develop targeted marketing strategies that will increase user engagement, retention, and acquisition. Provide insights and recommendations to address the following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

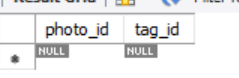
**Objective Questions:**

**\*\* For SQL references please refer to the SQL file where all the queries are mentioned along with the proper numbering to which objective/subjective question they correspond to.**

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

From the table definitions we can clearly observe that all the columns have either **NOT NULL constraint** or **DEFAULT constraint**. So, there would be no null values in our data.

Also, we used the following query to check for duplicates in the data.



As we can see clearly that we found no duplicate rows in ‘users’ table.

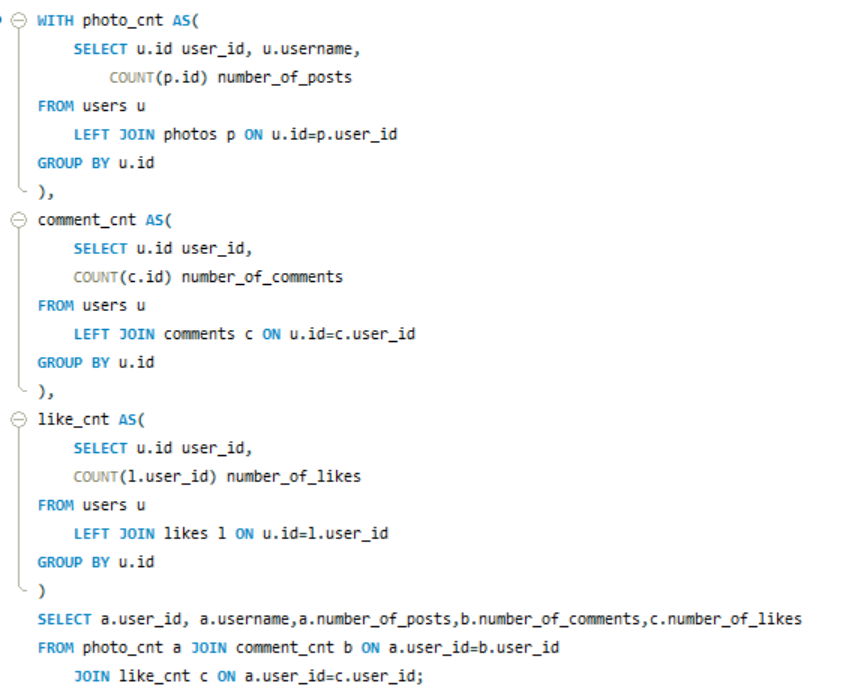
Subsequently upon performing similar steps for the other tables we conclude there are no duplicate values or null values in the data set that we are analysing.

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

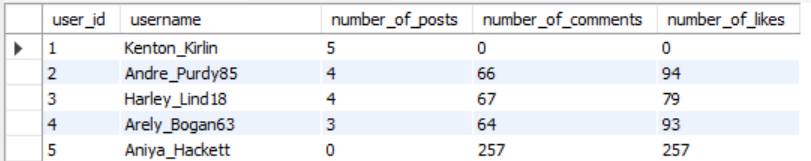
We calculated the number of posts, likes, comments per user in order to see the distribution of user activity levels across the user base.

We joined different tables to user tables and group the data based on users to get the aggregated counts as per our requirement.

**Query Used:**



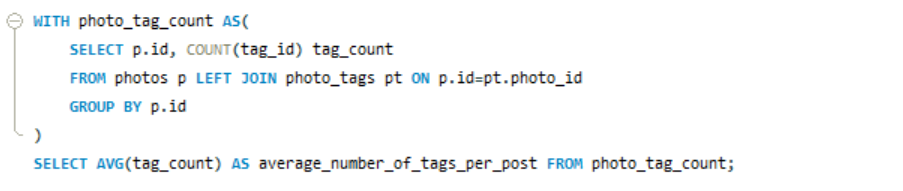
**RESULT:** Following are the top few rows of the result set.



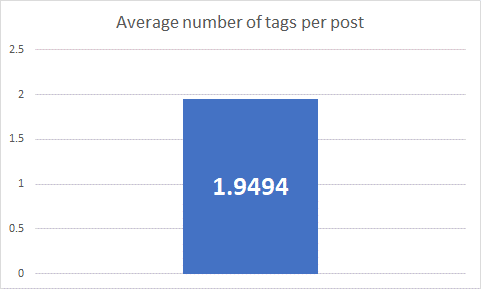
1. **Calculate the average number of tags per post (photo\_tags and photos tables).**

To get the average number of tags per post we calculated the number of tags for each post and subsequently got the average of it.

**QUERY USED:**

****

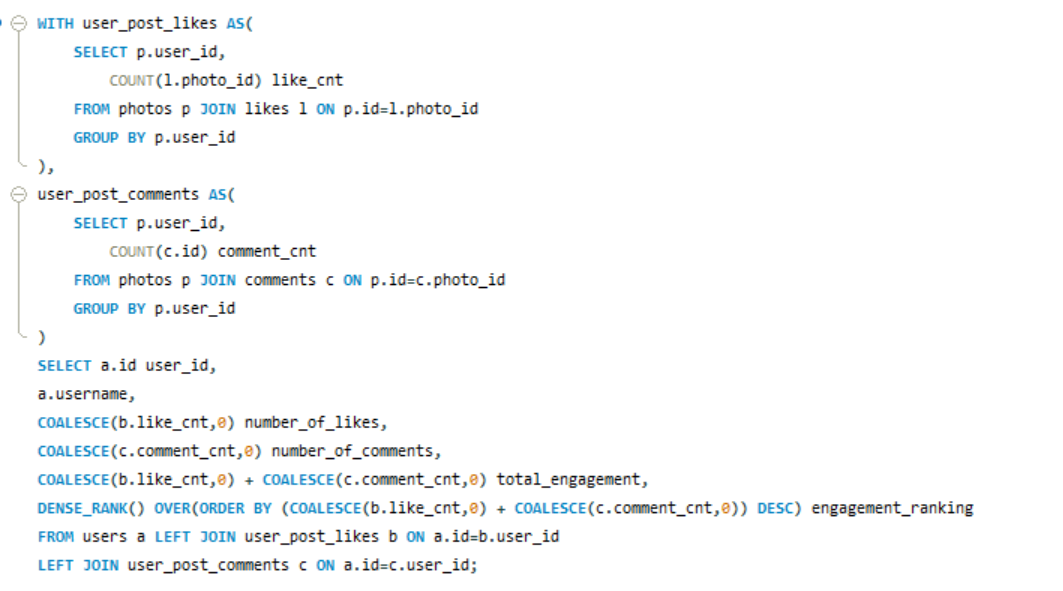
**RESULT:** The average number of tags per post came out to be around 2.



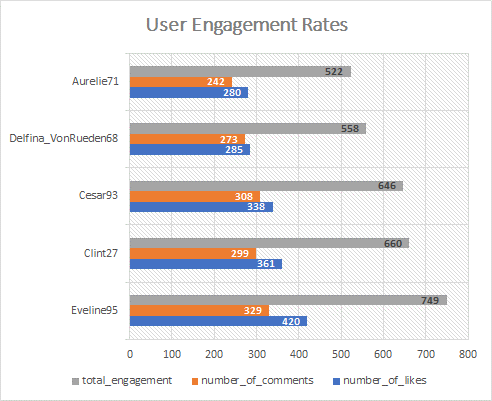
1. **Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.**

To identify the top users with the highest engagement rates we calculated the total number of likes and comments on each user’s post and subsequently ranked the users based on them.

**QUERY USED:**

****

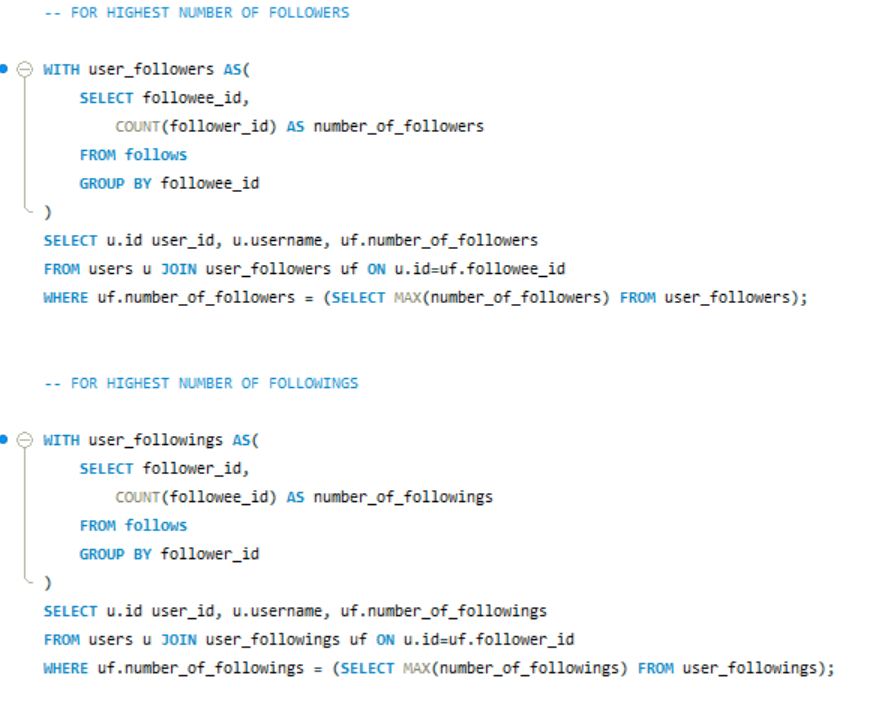
**RESULT:** The top five users with the highest engagement rates came out to be as follows.



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

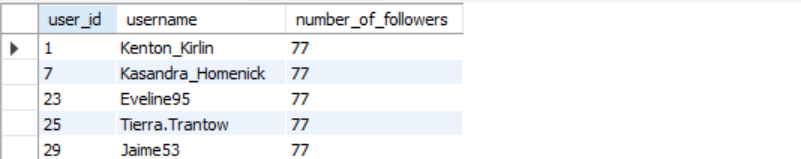
To identify users having the highest number of followers and followings we calculated the total number of followers and followings for each user ordered the results based on them.

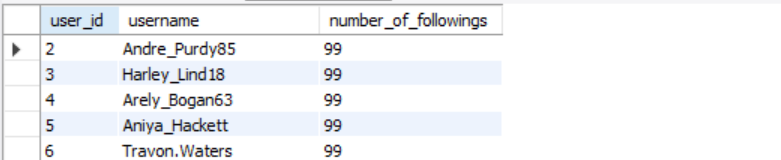
**QUERY USED:**

****

**RESULT:**

* The **highest number of followers** for any single user turned out to be **77**.
* The **highest number of followings** for any single user turned out to be **99**.
* In the given data set multiple number of users have same number of followers / followings. So top few rows of the result set are:

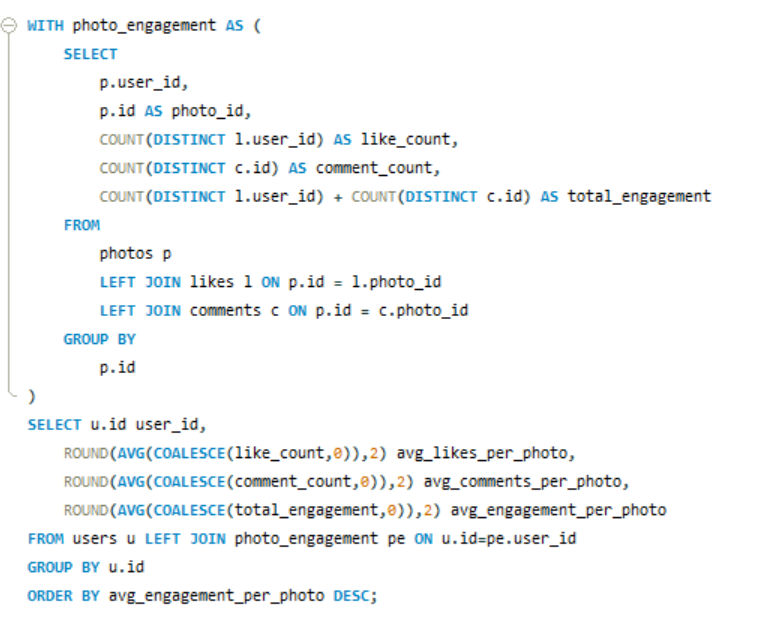




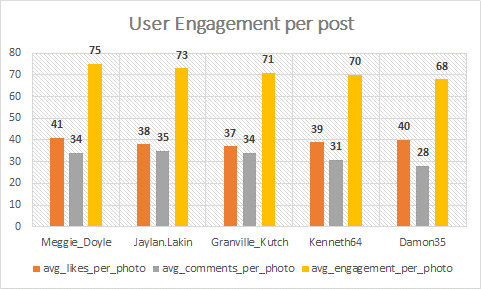
1. **Calculate the average engagement rate (likes, comments) per post for each user.**

To calculate the average engagement rate per post for each user we joined users table with likes and comments in order to get number of likes and comments on each post and subsequently group data based on users to get the required averages.

**QUERY USED:**

****

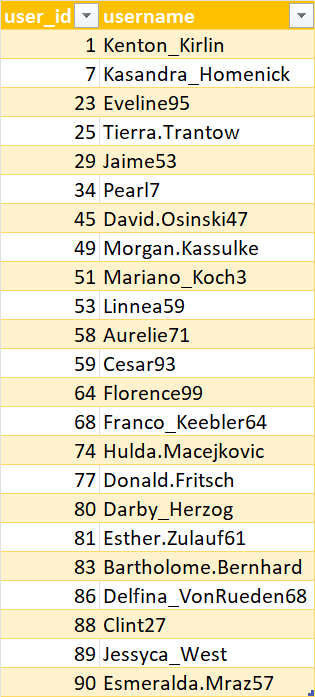
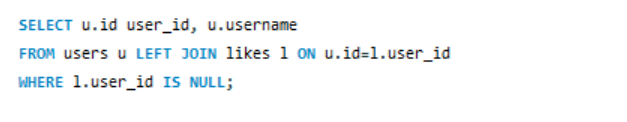
**RESULT:** The top five users with the highest average engagement rates on their posts came out to be as follows.



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

The users who never liked any post are the none that have no occurrence in likes table.

**QUERY USED:**

****

**RESULT:** Thelist of users who have never liked any post is as follows:

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

To create more personalised and engaging ad campaigns we can leverage user-generated content and use the data available with us in the different tables in various ways.

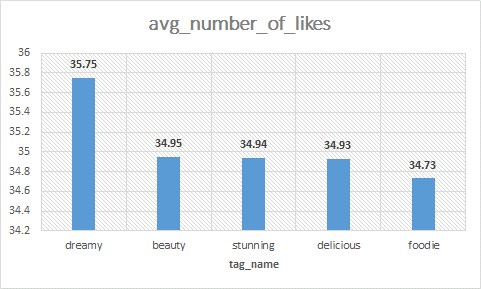
* We can segment the audience based on Engagement. We can identify users who frequently like comment or share UGC related to our specific brand and target them accordingly.
* The content can be customised and personalised as per the target audience. This could involve analysing the types of posts, tags our users are most engaged in.
* We can also aim at retargeting the audience that previously visited /liked our page/content with personalised ads based on the interaction.
* We can group audiences based on common interests with the help of content they interact with and the followers.
* We can try to collaborate with the influencers that majority of the user base is following hence gaining more interests from our users.

By effectively leveraging Instagram user data associated with UGC, brands can create highly personalized and engaging ad campaigns that resonate with their audience, drive meaningful interactions, and ultimately improve campaign performance metrics.

**QUERY USED:**

****

For example, we used the data to get the top hashtags with max engagement as follows:



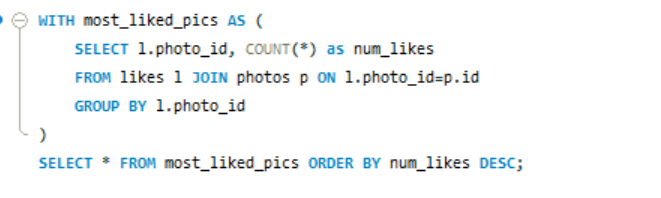
1. **Are there any correlations between user activity levels and specific content types (e.g., photos, videos, reels)? How can this information guide content creation and curation strategies?  
     
   We do not have any data related to videos and reels here in our data set so we can’t comment on those specific content types. But in general we can say,**

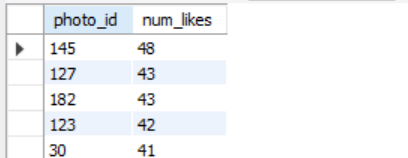
Yes, there are usually correlations between user activity levels and specific types of content (e.g., photos, videos, reels). Understanding these correlations can indeed guide content creation and curation strategies effectively

* Analysing the preference of your audience for the various types of content can often prove helpful. We can analyse metrics such as like, comments, share, view etc to see which type of contents our target audience often engage in.
* For example, if your audience shows a preference for videos by consistently engaging with them more than photos or reels, focusing on creating more video content can help maximize engagement.
* Allocating our resources in creating content mix curated by the user preference can often result in higher engagement rates.
* Continuously experimenting with different content types to understand evolving user preferences and adapting your strategy accordingly plays a vital role as well.
* Utilizing analytics tools provided by social media platforms to track engagement metrics for different content types. These insights can provide quantitative data to support content decisions and optimizations.

So, by leveraging these insights, brands and content creators can refine their content creation and curation strategies to maximize engagement, reach, and overall effectiveness on the platform.

**QUERY USED:**

  
  
Based on the data we figured out following were the top most liked pictures.



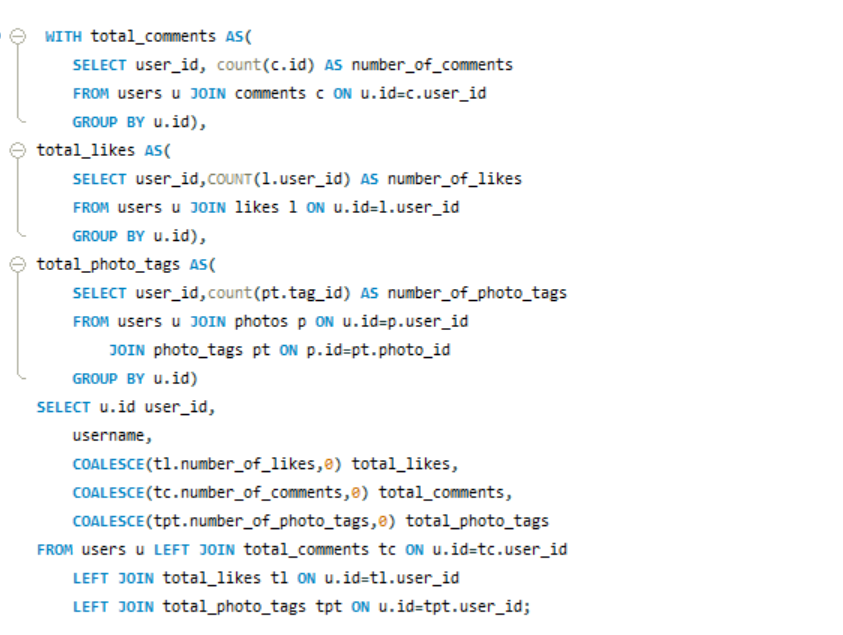
* Top of Form
* Bottom of Form

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

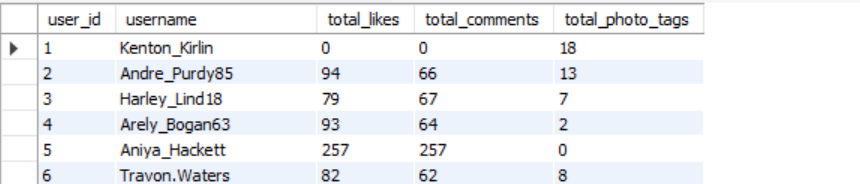
We calculated the number of likes, comments and photo tags for each user by creating different CTE’s where we joined the users table with likes, comments and photo\_tags respectively.

We grouped the data based on user in order to get the required counts.

**QUERY USED:**

****

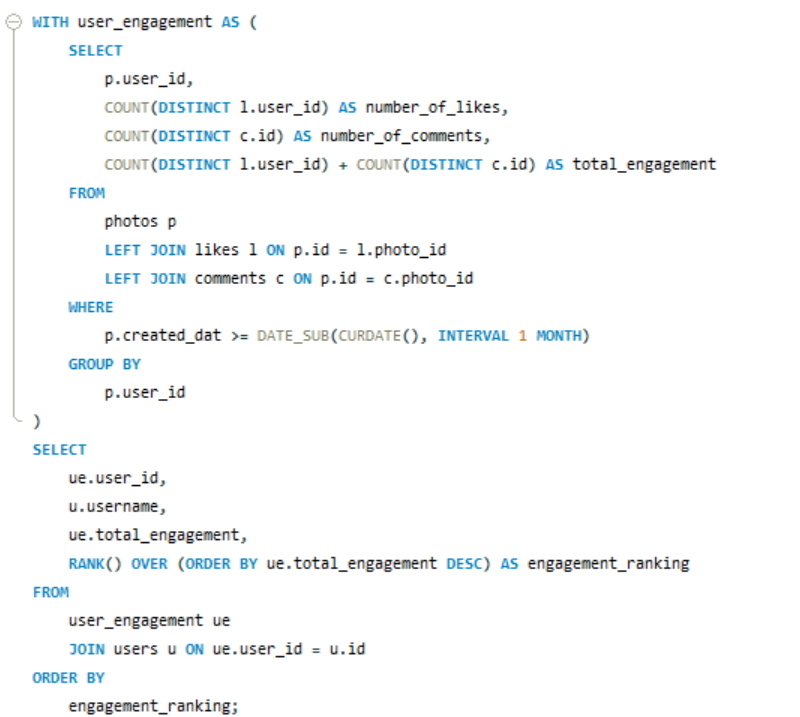
**RESULT:** Top few rows of the result set is as follows:

****

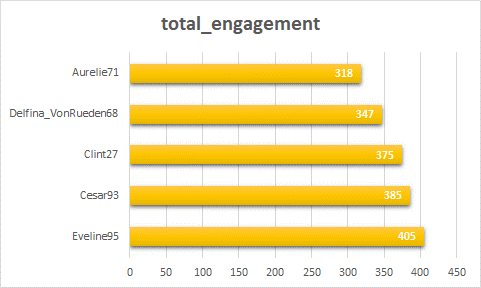
1. **Rank users based on their total engagement (likes, comments, shares) over a month.**

To identify the top users with the highest total engagement we calculated the total number of likes and comments on each user’s post and subsequently ranked the users based on them collectively as their total engagement.

**QUERY USED:**

****

**RESULT:** The top five users with the highest engagement came out to be as follows.

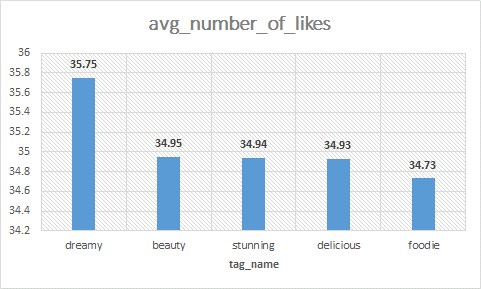


1. **Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.**

To identify the hashtags that have been used in posts with the highest average number of likes we created a CTE to calculate the average likes for each hashtag first and then subsequently calculated the average number of likes for each hashtag and ordered the results in decreasing order of these averages.

**QUERY USED:**

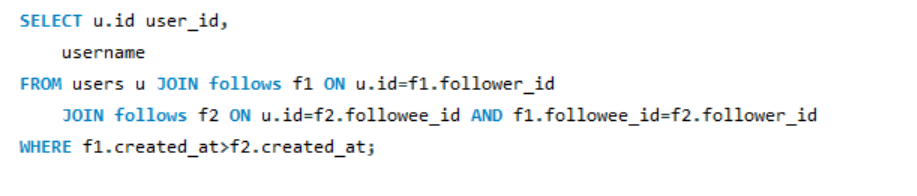
****

**RESULT:** The top five hashtags that have been used in posts with the highest average number of likes came out to be as follows.

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

To retrieve the users who have started following someone after being followed by that person we joined the users table with follows table twice once to get the follower time and then to get the following time on the condition that table one’s following id is same as second tables follower id.

And finally, we filtered the results based on created\_at column in order to get the desired result.

****

**RESULT:** In our data set all the values of created column are defaulted to NOW(). So the time for following and followers are same of each user and hence we got do records meeting the required conditions.



**Subjective Questions:**

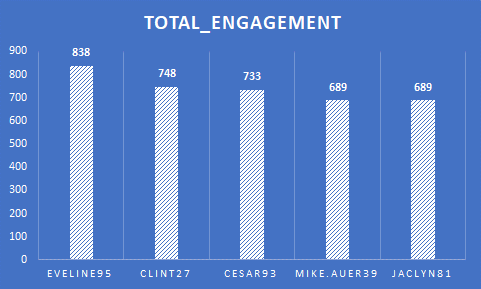
1. **Based on user engagement and activity levels, which users would you consider the most loyal or valuable? How would you reward or incentivize these users?**

To determine the most loyal or valuable customers we can consider various parameter to assess their engagement and activity levels. Few of them could be:

* Content liked by them.
* Comments posted by them.
* Photos posted by them.
* Number of likes received on their posts.
* Number of comments received on their posts.
* Number of followers.
* Number of followings.

We considered are above factors and accumulated them together to rank the users based on their total engagement and activity levels.

**RESULT:** The top 5 users based on this ranking came out to be as follows:



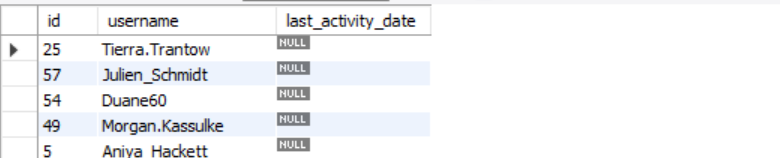
For the high engagement and activity levels we can incentivize these users in order to increase customer retention as well as attracting new customers. We can adapt the following strategies:

* Featuring tops users in some newsletters or a leaderboard.
* We can highlight their status through some special badges.
* Provide exclusive and early access to new features.
* Invitation for exclusive events like meetups and webinars.
* Discounts and gift cards.
* Personalized gift hampers.

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

First of all, to identify the inactive users we tried to find the last activity date of each user and then sorted them in ascending order.

The user with no entry in activity date or one with the last\_activity\_date as oldest can be classified as inactive user.

Some of the results are as follows:

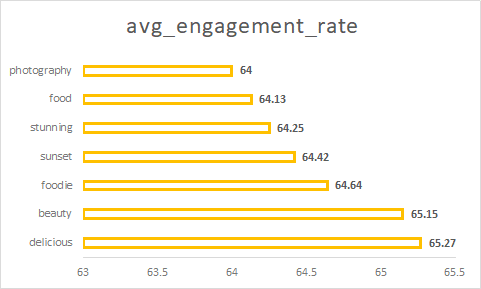
Some of the suggested re-engagement activities to engage the inactive users and encourage them to start posting or engaging again could be:

* Personalised email campaigns to intrigue users to engage again.
* Reach out with a personal touch, such as a message from a community manager or a special offer.
* We can send survey request to understand the reason for being inactive and make improvements accordingly.
* Provide incentives and special offers.
* Community engagement programs like challenges and contests with interesting rewards.
* Introduce gamification feature like points and badges to provide recognition and gain user interest.

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

To determine the hashtags / topics with highest engagement rates we calculated the number of likes and comments for each post with those tags. Subsequently we added the number of likes and comments and then calculated average engagement per post for each tag.

**RESULT:** The top 7 tags with highest engagement rates came out to be as follows:



This information can prove to be helpful in guiding content strategy and ad campaigns. We can leverage this data as:

* We can incorporate high-engagement hashtags into posts to increase visibility and reach.
* Create content around topics that are currently trending and of high interest to the audience.
* Create ad campaigns around trending topics that resonate with the target audience.
* Monitor the performance of ads using high-engagement hashtags and topics to assess their effectiveness.
* Invest more in ad campaigns and content creation related to high-engagement topics to maximize ROI.

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?**

Yes, there are often trends that can be noticed in user engagement based on demographics (age, location, gender) as well as posting time.

* Different age groups may prefer different types of content (e.g., younger audiences might favour short-form videos or memes, while older audiences might prefer detailed articles or tutorials).
* Engagement can even vary based on location as time zone changes and posting times accordingly.
* Besides, posting time also varies amount different age groups like youngsters these days are often night owls who post at late nights contrary to the elder group.
* Also, different genders might show varied engagement levels with different types of content. For instance, product preferences, content topics, and even visual styles can differ based on gender.

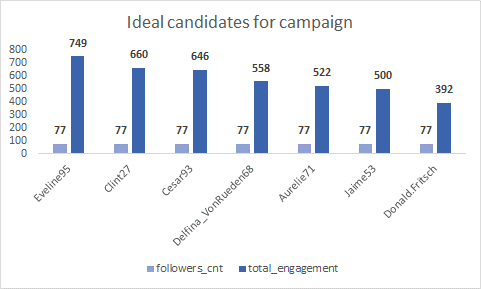
These insights can be helpful in targeted marketing campaigns as well:

* Tailor content to different age groups.
* Adapt content for regional cultural preferences and time zones to maximize engagement.
* Customize content and campaigns to resonate with different genders’ interests.
* Target ad campaigns based on times with maximum user engagement.
* Use demographic targeting in ad platforms to reach specific audience segments based on age, gender, and location.

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?**

Users with high number of followers and high engagement on their posts would be the ideal candidates for influencer marketing campaigns.

So, in our data we ranked the users based on their number of followers and then engagement (likes and comments) on their posts.

**RESULT:** Top 7 candidates with high number of followers and content engagement are:

While approaching and collaborating with the influencers we can keep the following strategy in mind:

* Start with sending a personalised message to the influencer regarding how they are a good fit for the campaign.
* Layout the objectives and guidelines of the campaign. Also, highlight the benefits that they get.
* Discuss about the compensation and incentive details.
* Work in a collaborative manner with scope for creativity. We can present our brand in a manner that resonates with their style and thus leaving an impact on the customer base as well.
* Feedback loops are very important. So, we can keep revisiting our strategies with scope for improvisations.

Also, we can try to foster long term relationships with the influencers through exclusive offers, appreciation for their work and proper communication.

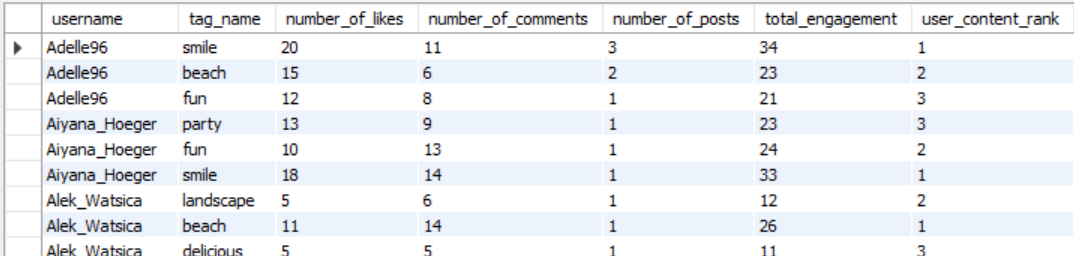
1. **Based on user behaviour and engagement data, how would you segment the user base for targeted marketing campaigns or personalized recommendations?**

User behaviour and engagement data can prove to be very helpful in segmenting the user base for targeted marketing campaigns or personalized recommendations.

From our data set we followed these steps for each user to know what type of content the users engage in:

* We calculated the number of posts the user has posted containing the particular tag.
* We calculated the number of posts the user has liked corresponding to the tag.
* We calculated the number of posts the user has commented on corresponding to the tag.
* We added all three value to get user’s total engagement corresponding to the tag.
* And finally, we ranked all the tags for each user based on total engagement and filtered the top 3 tags for each user.

**RESULT:** We got top three tags for each user based on their engagement. The example result set looks like this.



We can segment the user base accordingly for targeted marketing campaigns or personalized recommendations. We can consider the following:

* Group users based on their content preferences and include them in the targeted marketing campaign.
* Tailor content and recommendations to gender-specific interests and preferences.
* We can segment users based on their activity level as highly active, moderately active or inactive users.
* We can send personalised recommendations based on the type of content the user engages most in.

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

If data on ad campaigns (impressions, clicks, conversions) is available we can measure the effectiveness using different KPI’s and hence adapt measures to optimize the future campaigns. It could involve following steps:

* We can calculate CTR (Click Through Rate) to measure how well the ad attracts the user to click.
* We can calculate Conversion Rate to measures how well the ad converts clicks into desired actions like purchases, sign-ups etc.
* We can calculate ROI in terms of amount spent on ads and corresponding returns.
* We can evaluate metrics across different ad campaigns to identify which campaigns perform best.
* We can test out ads with slight variation to see which variation performs best.
* Based on performance, we can adjust the messaging to better align with what resonates with the audience.
* We can further refine our target audience based on the kind of response we get.
* Reallocate budget from underperforming campaigns to those with higher CTR and conversion rates.
* Monitor the performance regularly and reiterate through the process to keep optimising.
* Utilize analytics tools and platforms (e.g., Google Analytics, Facebook Ads Manager) to gain deeper insights into campaign performance.

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

We can consider various factors using the user activity data to identify potential brand ambassadors or advocates who could help promote Instagram’s initiatives or events. These include:

* Identify the users with high level of engagement like considering the users who consistently like comment or share content on Instagram and take part in such events and initiatives.
* Identify the users with a large follower base. Such users can potentially influence a large number of audiences whose interests align well with them.
* Identify users who use relevant hashtags associated with Instagram’s campaigns or initiatives.
* Assess if the user’s content has gone viral or significantly impacted their followers, indicating their potential as an ambassador.

Thus, by leveraging user activity data to identify and engage with potential brand ambassadors, Instagram can effectively promote its initiatives and events through influential and enthusiastic advocates, amplifying its reach and impact.

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

**Problem Statement:** You are hired as a data analyst at Meta and asked to collaborate with Marketing team. Marketing teams wants to leverage Instagram's user data to develop targeted marketing strategies that will increase user engagement, retention, and acquisition.

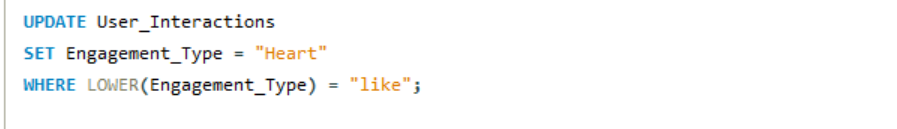
**Suggested Approach for the given problem statement:**

* First of all, we need to understand the data completely. And we need to thoroughly understand all the attributes in data.
* Next, we need to understand the significance of those attributes in determining different metrics and relevance in drawing insights.
* Further we can join tables to draw insights based on collective data.
* We can consider different sorts of metrics to understand customer behaviour.
  + Identify the activity level of users to filter out the active user base.
  + Engagement metrics to see which users like comment or post maximum content.
  + The number of likes and comments on user’s post to see the content popularity.
  + The number of followers any user has so as to identify the influential status
  + Demographic segmentation
  + The tags associated with any users post to identify the common interest points as well as most popular topics.
* After analysing the data, we can develop Targeted Marketing Strategies
  + Create personalized content recommendations based on user interests and engagement history.
  + Determine the best times to post based on when users are most active and engaged.
  + Develop campaigns to re-engage inactive users, such as personalized re-engagement emails or notifications.
  + Implement loyalty or rewards programs to incentivize continued engagement and reward frequent users.
  + Collect feedback from users to understand their needs and improve the user experience based on their input.
* Finally, we can strategize re iterations and improvement plans for our campaigns.
  + Regularly track and analyze the performance of marketing campaigns and strategies using relevant metrics.
  + Evaluate the impact of different strategies on user engagement, retention, and acquisition.
  + Expand and scale successful campaigns or strategies that show positive results.

Hence, we can effectively leverage Instagram user data to develop targeted marketing strategies that will increase user engagement, retention, and acquisition. Thus, proving the significance of data driven decision in growth and success of any business model.

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 can use UPDATE statement in SQL that is a DML (Data Manipulation Language) statement in order to achieve the required result.

 We can use the following Query:

\*\* Since the query is written on an assumed table which does not exist in the given data we can’t provide an output for the same.