**Social Media Project Documentation**

**Vikas Upadhyay**

**Observations and insights for Objective and Subjective questions:**

**Objective Question:**

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

**Ans:**

* There are no null or missing values in any of the tables in our database. This indicates that the data is well-maintained and complete, which is essential for accurate analysis and reporting.
* "In the future, if data quality issues like null or duplicate values arise, we can use the following queries to identify them:
* **Check for Nulls:**

SELECT \*

FROM table\_name

WHERE column\_name IS NULL;

* **Solutions to Handle Potential Issues**:
* "Replace missing values with default values, where applicable (COALESCE in SQL)."
* "Update records based on related tables or business logic."
* "Remove records if the missing data makes them unusable."
* **Check for Duplicates**:

SELECT column\_name, COUNT(\*)

FROM table\_name

GROUP BY column\_name

HAVING COUNT(\*) > 1;

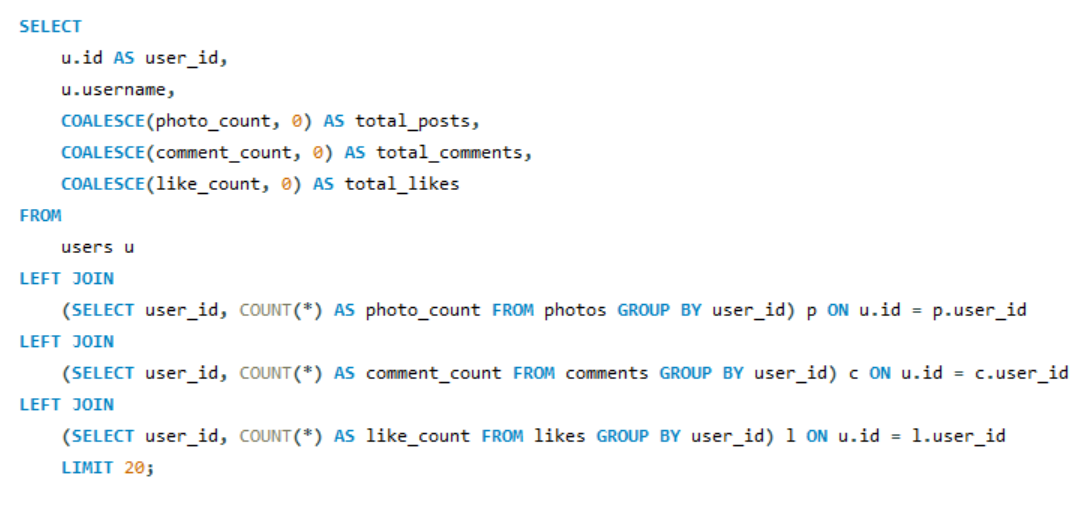
* **Solutions to Handle Potential Issues**:
* Identify the root cause of duplication, such as redundant inserts.
* Remove duplicates using ROW\_NUMBER() or DISTINCT in SQL queries.
* Enforce unique constraints on relevant columns.
* **Preventive Measures to Maintain Data Integrity**:
* Implement strict database constraints (e.g., NOT NULL, UNIQUE, and PRIMARY KEY constraints).
* Regularly validate data during ETL (Extract, Transform, Load) processes.

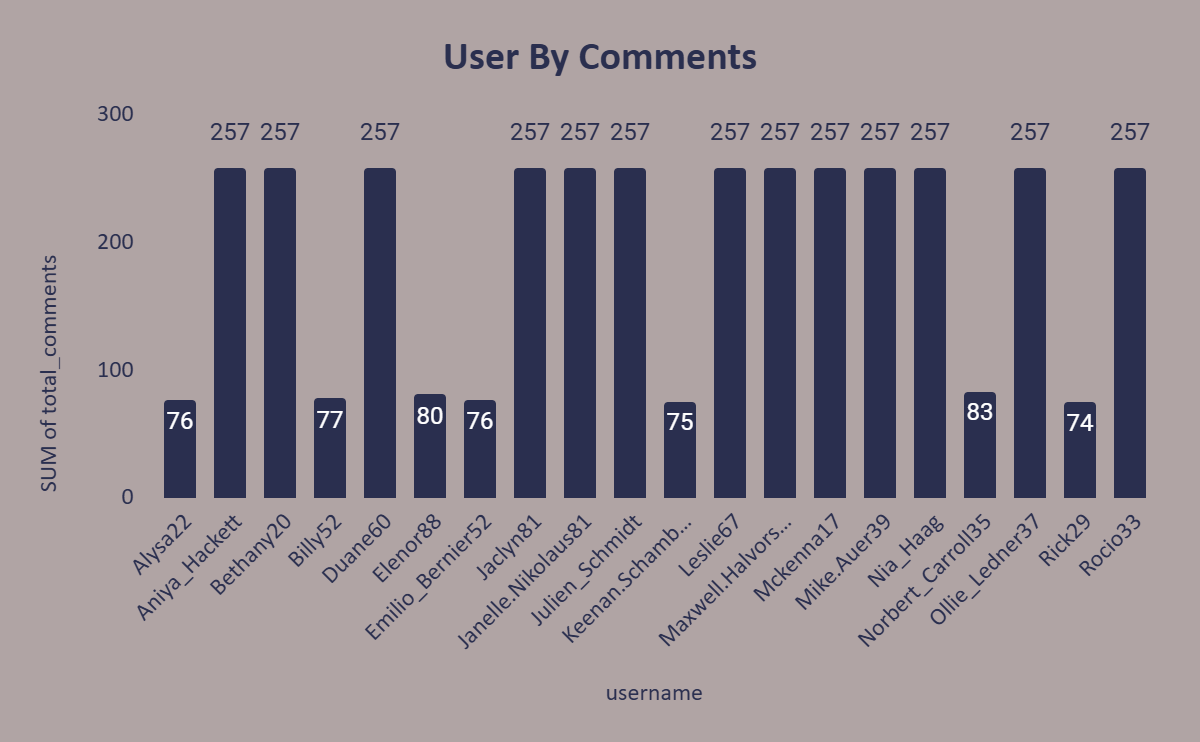
1. **What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?**

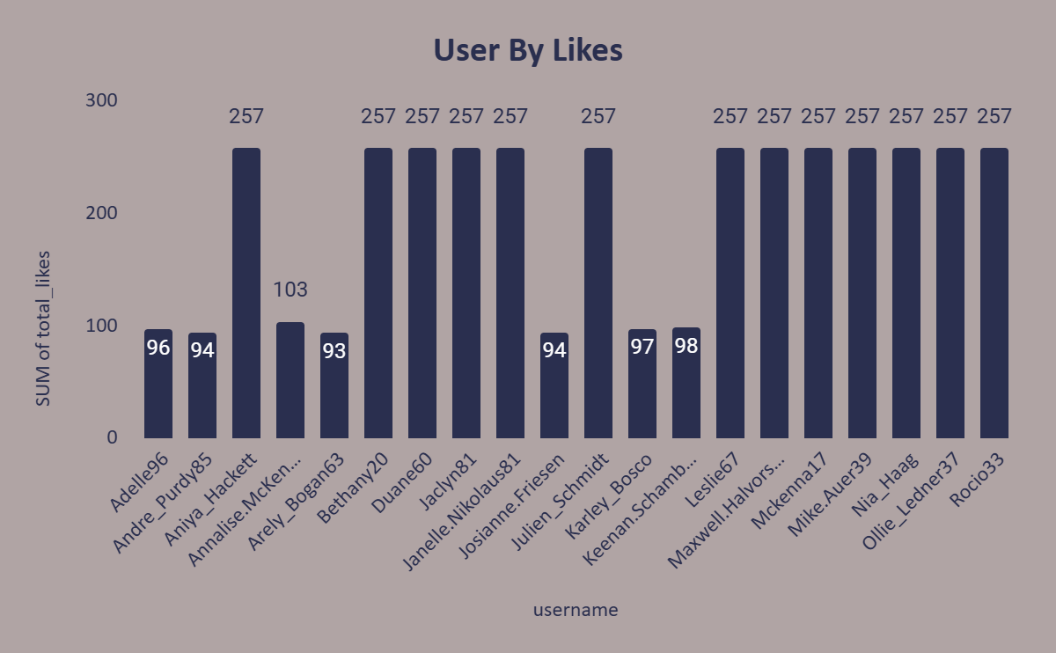
**Ans:**

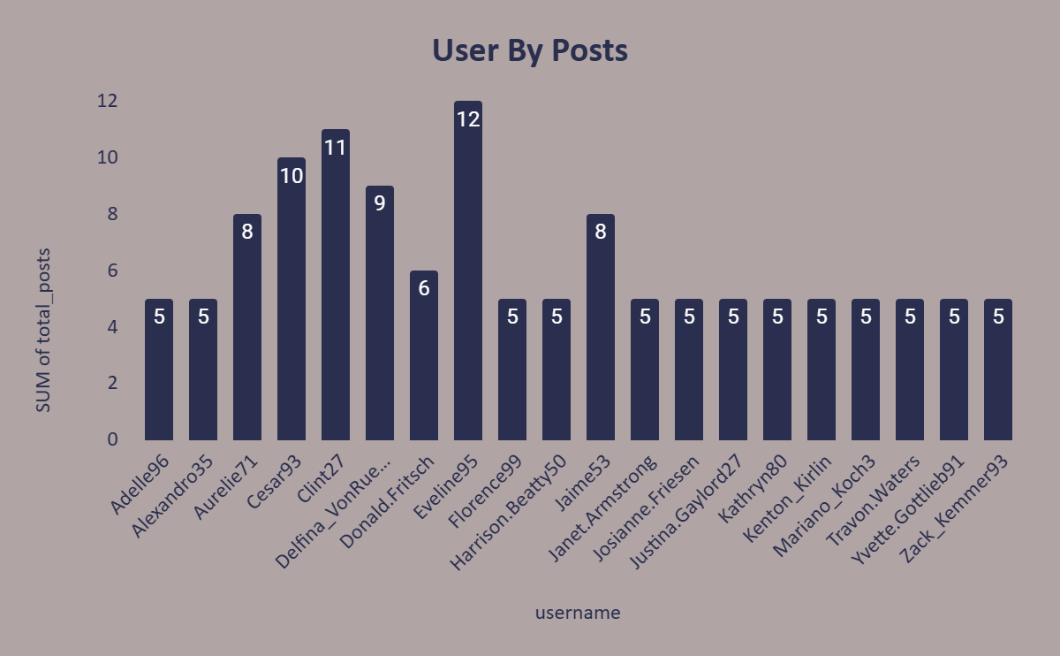
* To calculate the distribution of user activity levels (posts, likes, and comments) across the user base, we need to aggregates the number of photos posted, comments made, and likes given by each user.

**Visualization:**

****





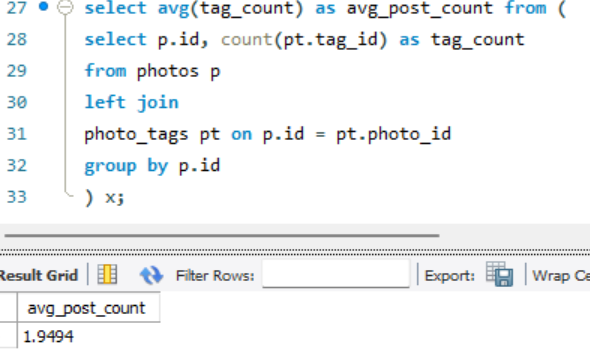


* **Observations:**
* This data provides a clear snapshot of user activity, including total posts, comments, and likes for each user.
* Users with higher counts are more active and could be key drivers of community engagement.
* Users with zero activity (photos, likes, or comments) can be targeted with engagement campaigns.
* Identifies users with high activity levels, which can indicate key contributors to the platform.

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

**Ans:**

* This query gives us the average number of tags per photo.
* **Visualization:**

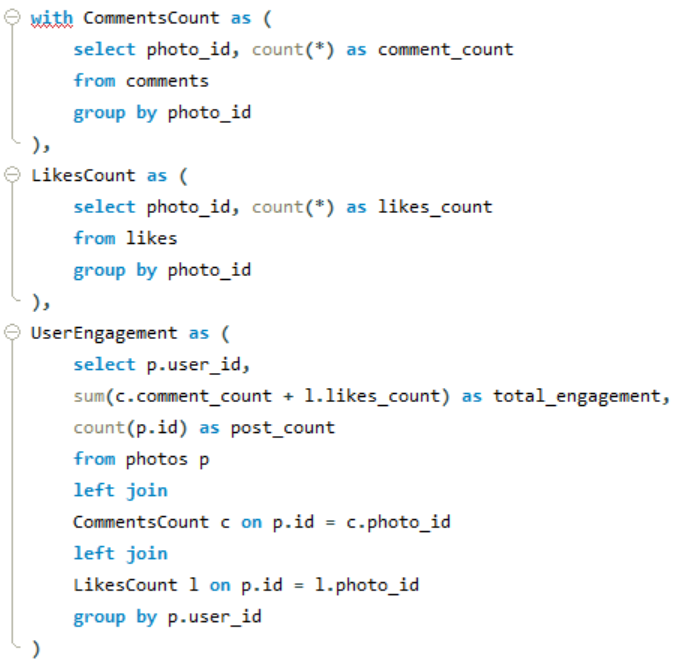
****

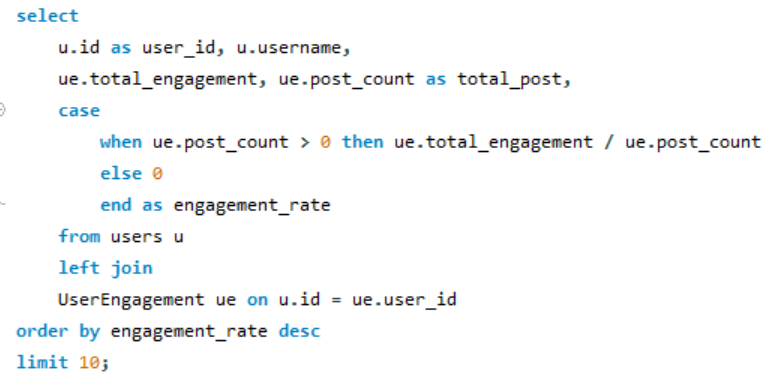
* **Observations:**
* The average number of tags per post is 1.9494.

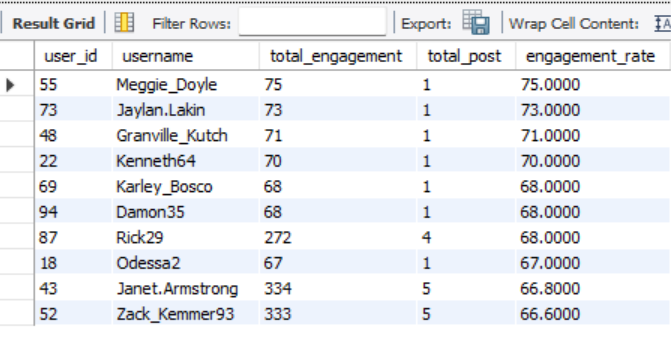
1. **Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.**

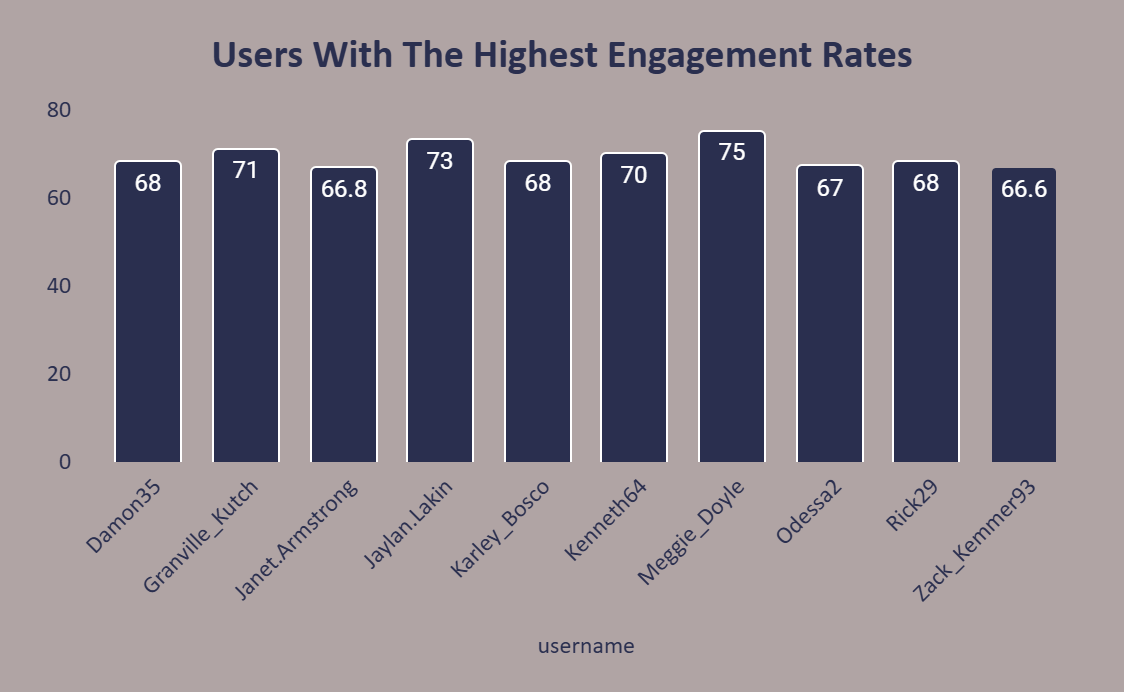
**Ans:**

* To identify and rank the top users with the highest engagement rates on their posts, we need to calculate the total engagement (likes and comments) for each user’s posts and divide that by the number of posts they have made.
* **Visualization:**

****

****



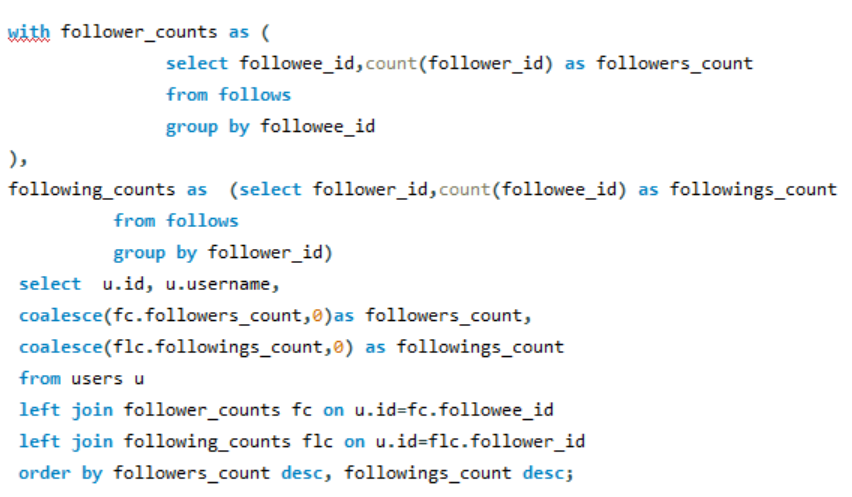


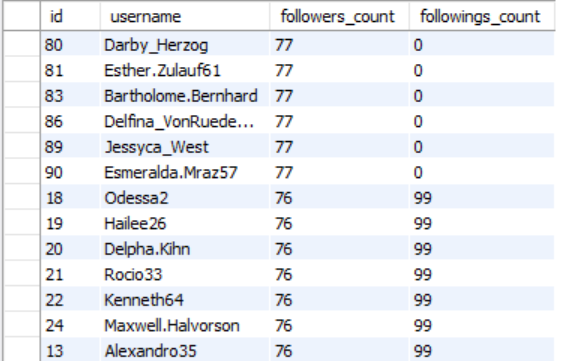
* **Observations:**
* Users like Janet Armstrong and Zack Kemmer have higher posts but their engagement rate per post is low compared to few others.
* While users like Meggie Doyle, Jaylan Lakin, and Granville Kutch have fewer posts but their engagement rate is higher. Users with lots of posts but low engagement might need content improvement.

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

**Ans:**

* To identify users with the highest number of followers and followings, we analyse the follows table.
* **Visualization:**

****

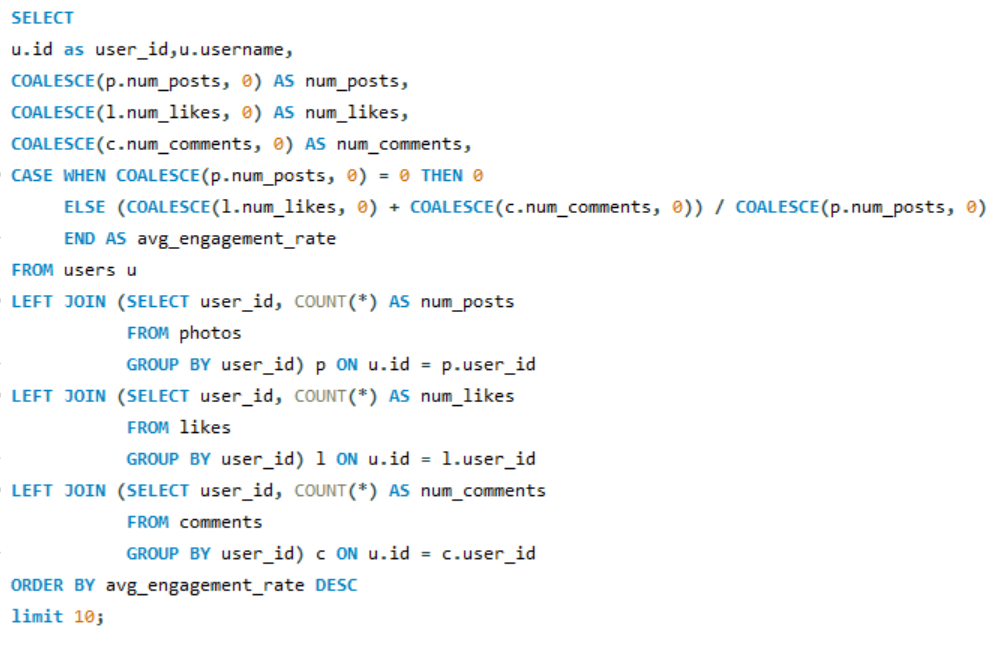
****

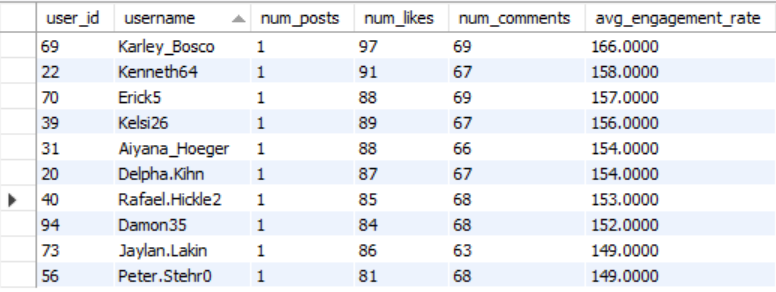
* **Observation:**
* This query provides actionable insights that can be directly applied to improve platform engagement and user retention strategies.
* **Total Followers**: Number of users following each individual.
* **Total Followings**: Number of other users each individual is following.
* **Comparing both followers and followings provides insights into user behaviour.**
* High followers but low followings = Influencers.
* High followings but low followers = Explorers.
* High in both = Social Connectors.

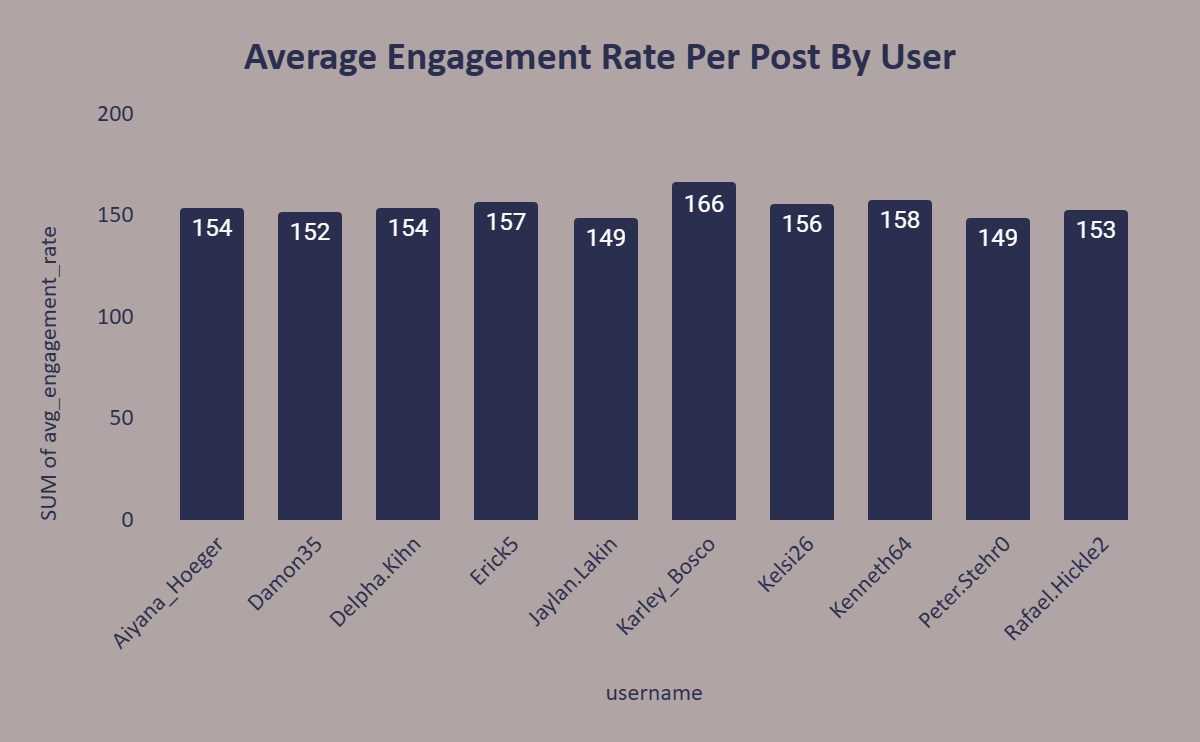
1. **Calculate the average engagement rate (likes, comments) per post for each user.**

**Ans:**

* To calculate the average engagement rate (likes and comments) per post for each user, we'll determine the total engagement (sum of likes and comments) for their posts, divide it by their total number of posts, and handle cases where users have no posts to ensure a meaningful result.
* Calculate the average engagement rate as:
* Average Engagement Rate = Total Engagement​ / Total Posts
* **Visualization:**

****

****

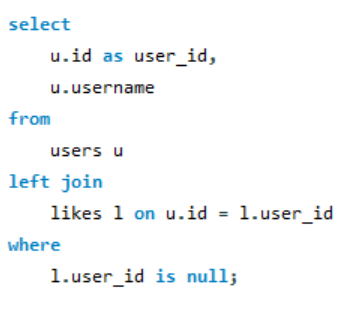


* **Observation:**
* This chart highlights the users with the highest engagement per post.
* Measure how effectively users generate interaction on their posts.
* Users with low average engagement rates may need to improving their content visibility or quality.

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

**Ans:**

* **Visualization:**

****

****

* **Observation:**
* The above table shows the list of users who have never liked any posts.
* These users may represent inactive accounts or individuals who consume content but do not engage.
* Some users might prefer passive participation (browsing without interaction) rather than active engagement.

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

**Ans:**

* User-generated content (such as posts, hashtags, and photo tags) can provide valuable insights into user behaviour, preferences, and trends.

you can leverage this data effectively to create personalized and engaging ad campaigns.

* **Analyse Popular Hashtags and Photo Tags**:
* Identify trending hashtags and tags associated with posts to understand the interests and themes that resonate most with your audience.
* Use these tags to create ads that align with these trends, making the content feel relevant and timely.
* **Segment Users Based on Content Themes**:
* Group users by the topics they engage with most (e.g., food, travel, tech) using hashtags and tagged content.
* Deliver ads tailored to these specific themes, increasing the likelihood of engagement.
* **Showcase UGC in Ads and Identify Influential Creators**:
* Highlight content from influential or highly engaged users to amplify reach.
* Collaborate with these creators to co-create or promote campaigns.
* **Geotag and Localize Ads**:
* Utilize geotags from posts to identify location-based trends and preferences.
* Create location-specific campaigns that reflect local culture, events, or needs.
* **Behavioural Targeting**:
* Analyse the types of posts users create or engage with (likes, comments) to build behavioural profiles.
* Use these insights to serve ads tailored to their habits, such as promoting outdoor gear to someone posting hiking photos.

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?**

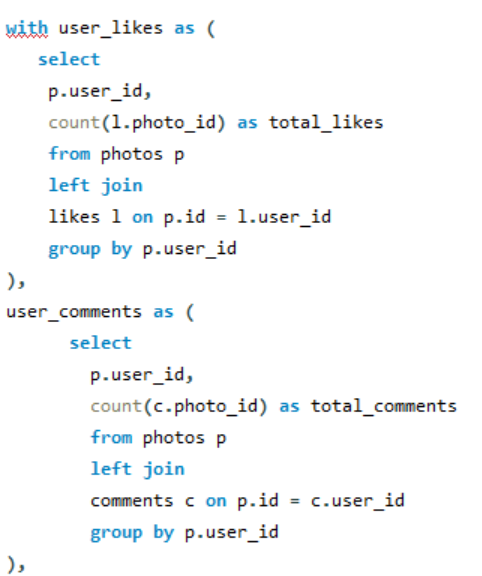
**Ans:**

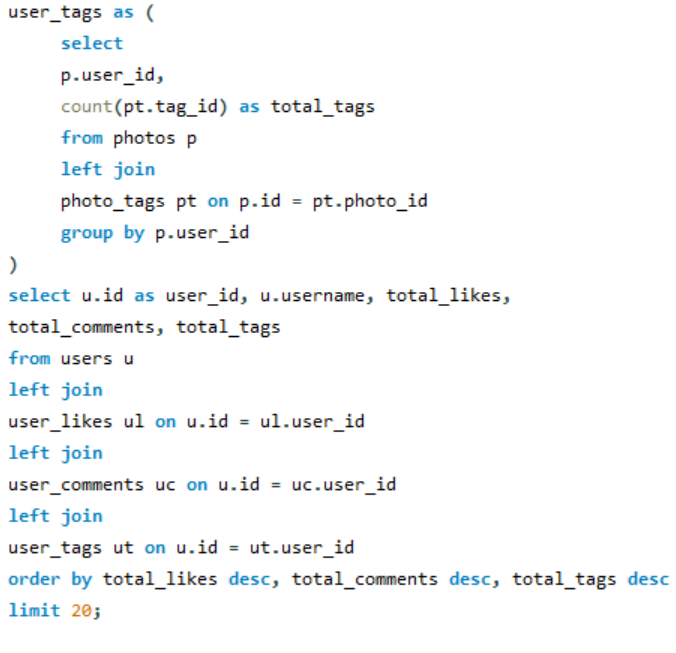
* While our existing database schema focuses on user activity (likes, comments, followers, etc.), it currently lacks details about content types (e.g., photos, videos, reels). Assuming content types could be tracked by adding a column (content\_type) to the photos table or using a similar mechanism, we can explore the correlation between user activity and content types. this can be approached conceptually and strategically. Correlating user activity with content types can lead to data-driven decisions that enhance user satisfaction and maximize engagement.
* **Correlation Analysis:**
* Measure engagement levels (likes, comments) for different content types.
* Compare metrics such as average likes per post or comments per post across content types.
* **Content Popularity:**
* Identify which content types drive the most engagement.
* Understand the user demographic or behavioural patterns associated with each content type.
* **Content Trends:**
* Track time-based trends to see how preferences for specific content types change over time.
* **Content Creation and Curation Strategies:**
* **Personalization:** Tailor content recommendations based on user activity history. For example, suggest more videos to users who engage with videos the most.
* **Diversity in Formats:** Offer a mix of content types to cater to diverse audience preferences, but prioritize high-engagement formats.
* **Promotion:** Boost engagement by promoting content types that already perform well organically.
* **Content Planning:** Use engagement trends to guide editorial calendars. For example, schedule more reels during peak engagement hours.
* **Experimentation:** Test new formats or variations (e.g., longer videos, animated photos) and analyse performance.

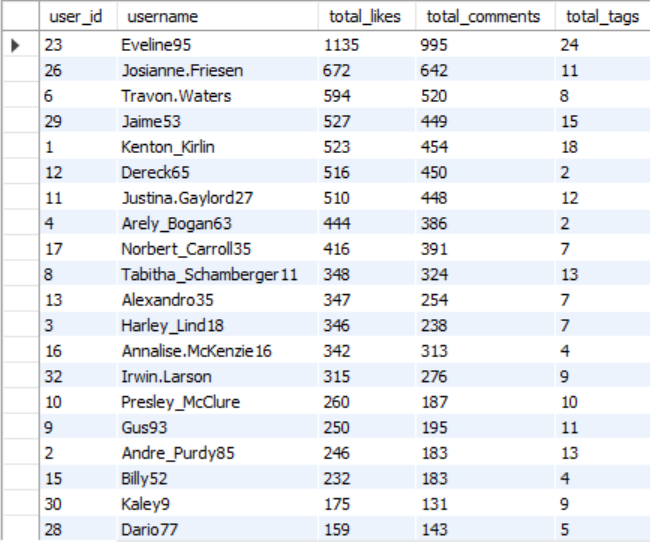
1. **Calculate the total number of likes, comments, and photo tags for each user.**

**Ans:** By analysing these metrics, we can identify key contributors and tailor strategies to encourage broader engagement**.** The user with the total number of likes, comments and photo tags are:

* **Visualization:**

****

****

****

* **Observation:**
* Users with high likes and comments indicate active contributors whose content resonates with others.
* Users with high tag counts are associated with diverse or frequently tagged content, making them ideal for campaigns tied to specific themes.
* Users with few likes, comments, or tags may represent opportunities to activate through engagement initiatives.
* Identify what type of content from high-engagement users drives interaction and replicate those themes.

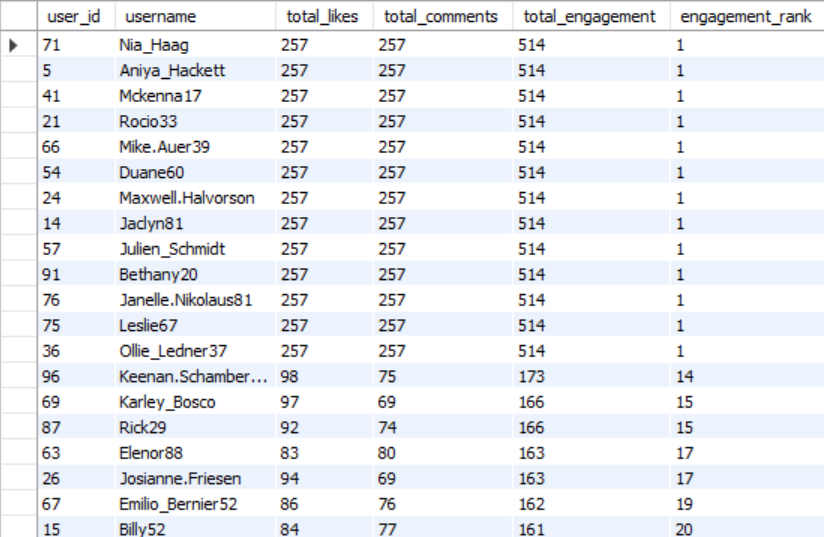
1. **Rank users based on their total engagement (likes, comments, shares) over a month.**

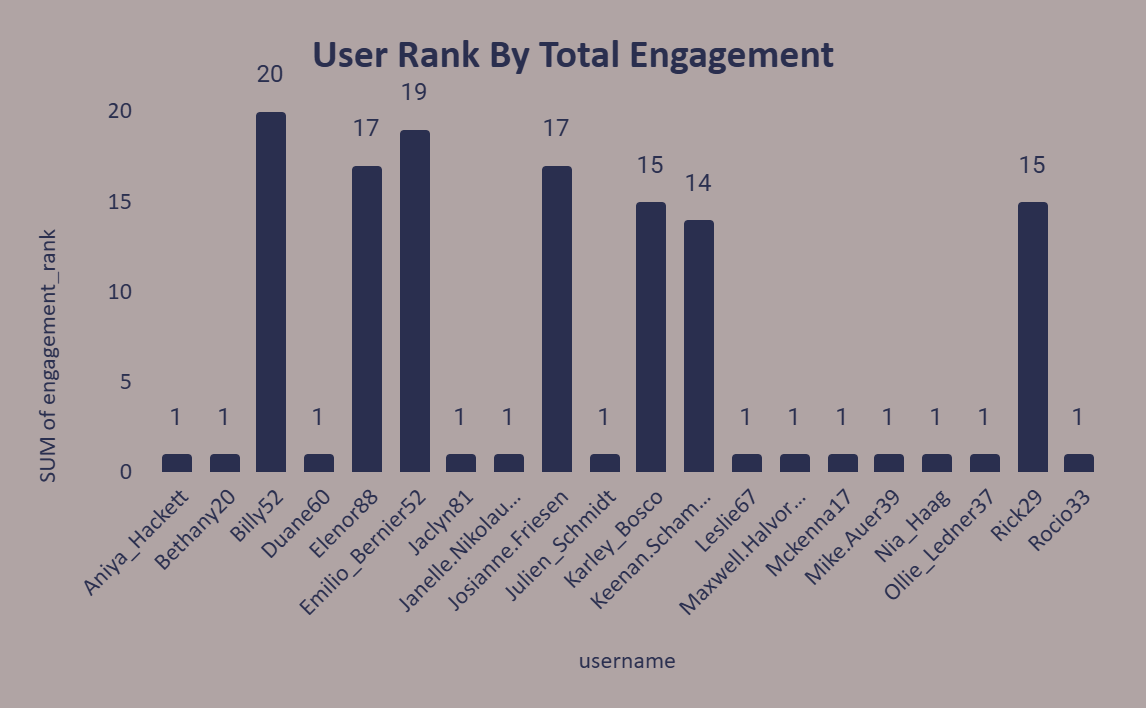
**Ans:**

To rank users based on their total engagement (likes, comments, and shares) over a one-month period, we'll need to consider the following:

* **Likes, comments, and shares** — aggregate these metrics for each user.
* **Time frame** — filter the data to include only the past month.
* **Ranking** — rank users based on total engagement.
* **Visualization:**

****

****

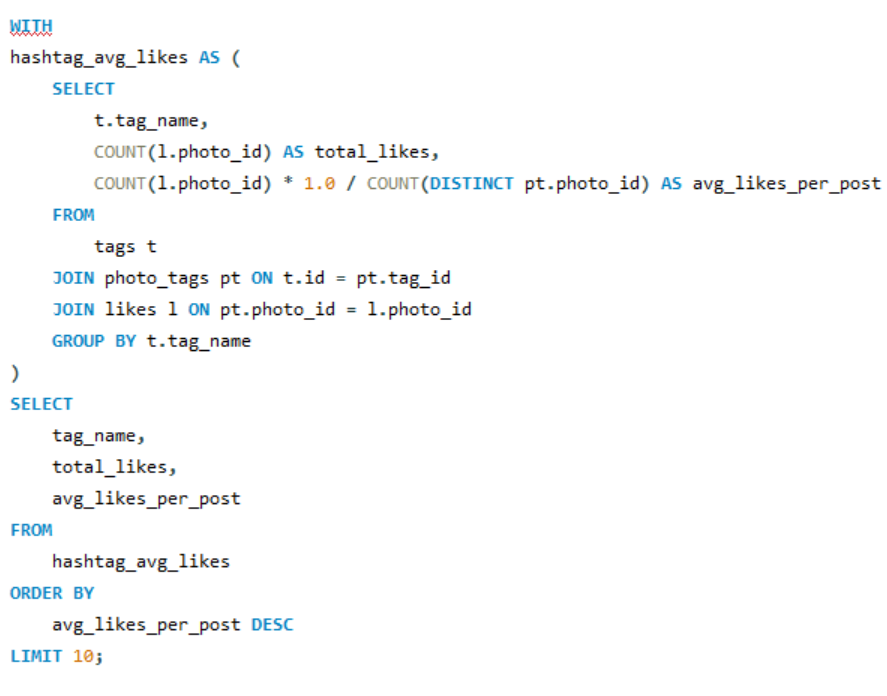


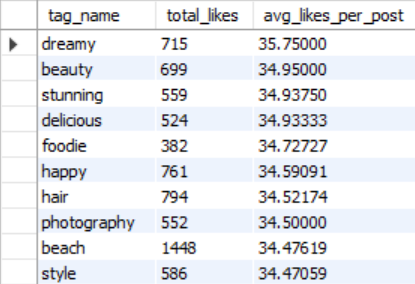
* **Observations:**
* Users who rank higher based on total engagement (likes, comments, and tags) are likely to be the most influential on the platform.
* By analysing the types of posts that attract the most likes, comments, and tags, we can identify trends and replicate those strategies to boost user engagement across the platform.
* Users with low engagement (ranked lower) may benefit from promotional strategies such as content contests, giveaways, or personalized engagement prompts to encourage activity.
* The total engagement per user can also indicate which content is resonating the most with the audience. This data is valuable for optimizing future posts.

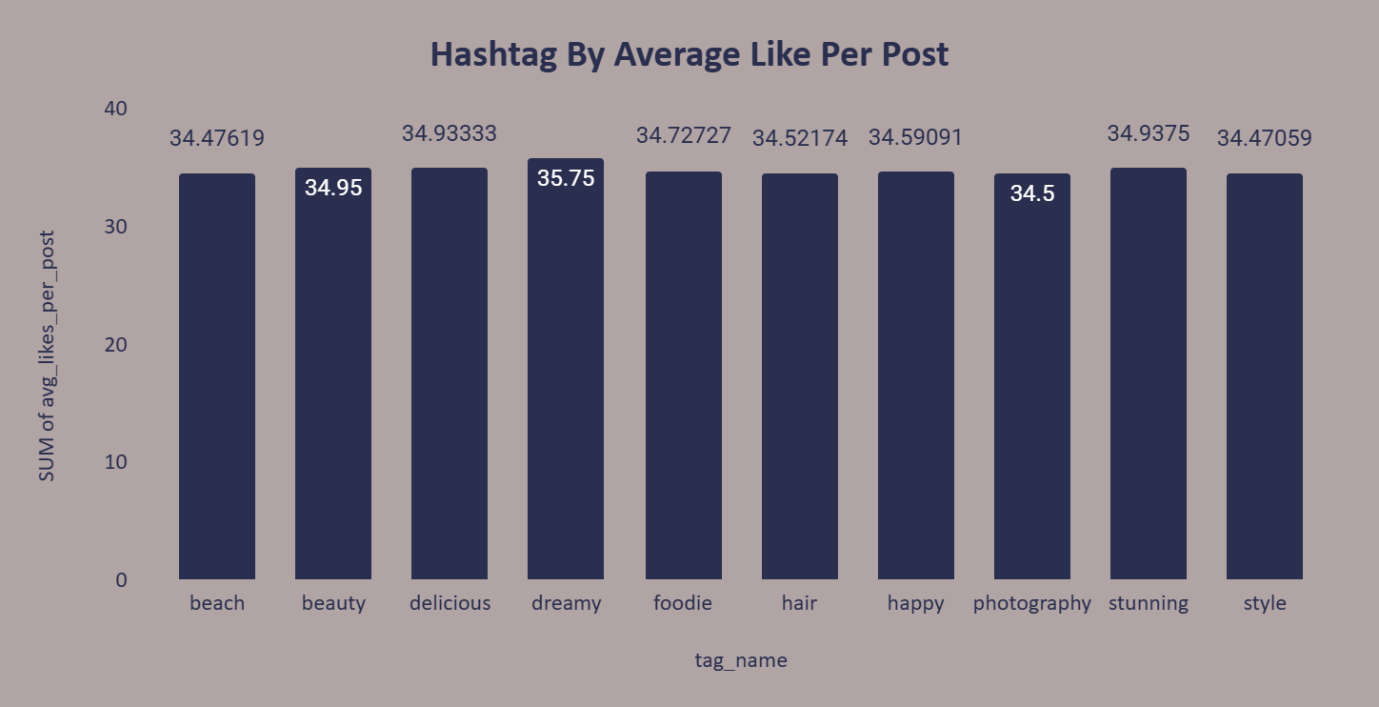
1. **Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.**

**Ans:** The hashtags that have been used in posts with the highest average number of likes are:

* **Visualization:**

****

****

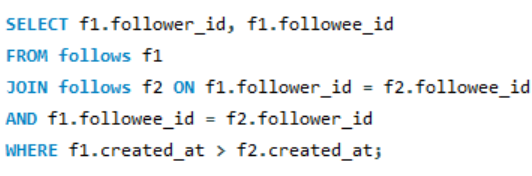


* **Observations:**
* The hashtags with the highest average likes per post are indicative of the most engaging topics or themes on the platform.
* Use these popular hashtags to increase visibility and engagement for new posts or campaigns.
* Monitoring these hashtags over time can help identify seasonal or emerging trends.

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

**Ans:**

* To retrieve the users who have started following someone after being followed by that person, we can use the follows table and perform a self-join.
* **Visualization:**

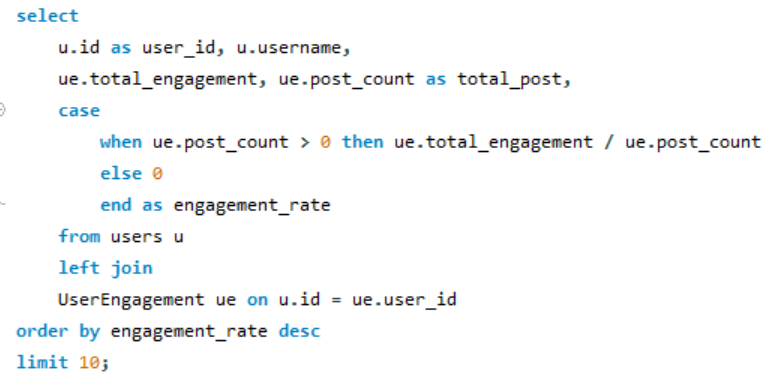


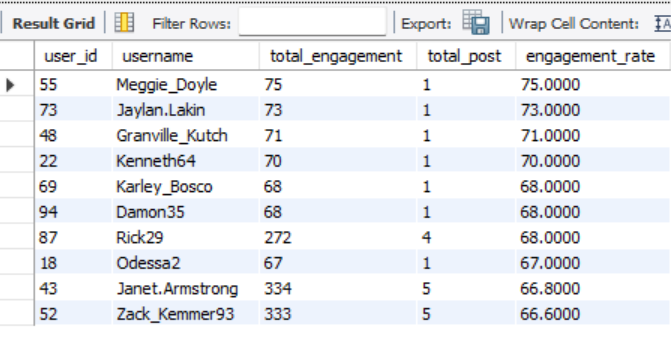
* **Observations:**
* This query identifies reciprocal follow actions where a user decided to follow back someone who followed them first.
* Such interactions indicate potential mutual interest or courtesy-based follow-backs.
* **Subjective Questions:**

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?**

**Ans:**

* Loyal or valuable users can be identified based on metrics that reflect high engagement, consistent activity, and positive contributions to the platform.
* **Observations:**

****

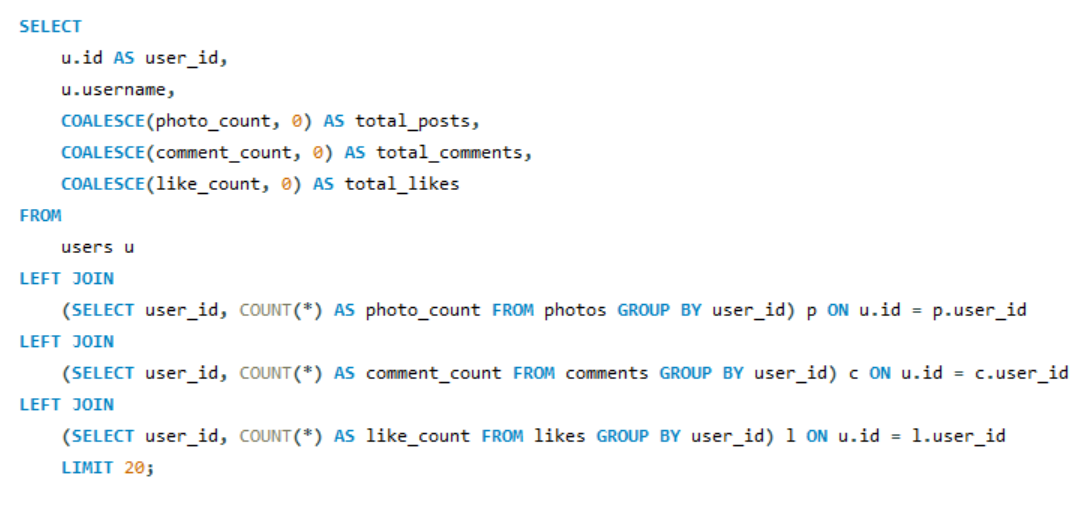


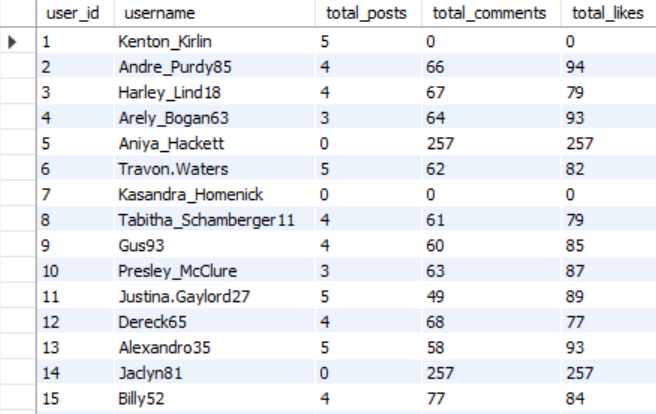
* **Engagement Rates:**
* The engagement rates for the users listed range between 66.6 and 75.
* **Meggie Doyle** has the highest sum of engagement at **75**, indicating strong interaction with their content.
* Users like **Jaylan Lakin** and **Granville Kutch** also have high engagement rates (**73** and **71**, respectively), positioning them as top contributors to the platform's engagement.
* Many users (e.g., **Damon35**, **Karley Bosco**, **Rick29**) have engagement rates around **68**, suggesting consistent, above-average activity.
* **Zack\_Kemmer93** and **Janet Armstrong** have the lowest engagement rates, indicating potential for improvement or different audience dynamics.
* **Insights:**
* **Top Performers Drive Engagement:**
* Users like Meggie Doyle, Jaylan Lakin, and Granville Kutch likely contribute significantly to community interactions, making them valuable to the platform.
* **Engagement Clusters**:
* Most users are clustered around similar engagement levels, suggesting consistent activity patterns across a segment of the user base.
* **Potential for Growth**:
* Users with lower engagement could benefit from targeted strategies to increase their contributions.

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

**Ans:**

* **Strategies to Re-engage Inactive Users:**
* Re-engagement strategies should focus on making the return easy, rewarding, and enjoyable, ensuring users feel valued and motivated to stay active.

****

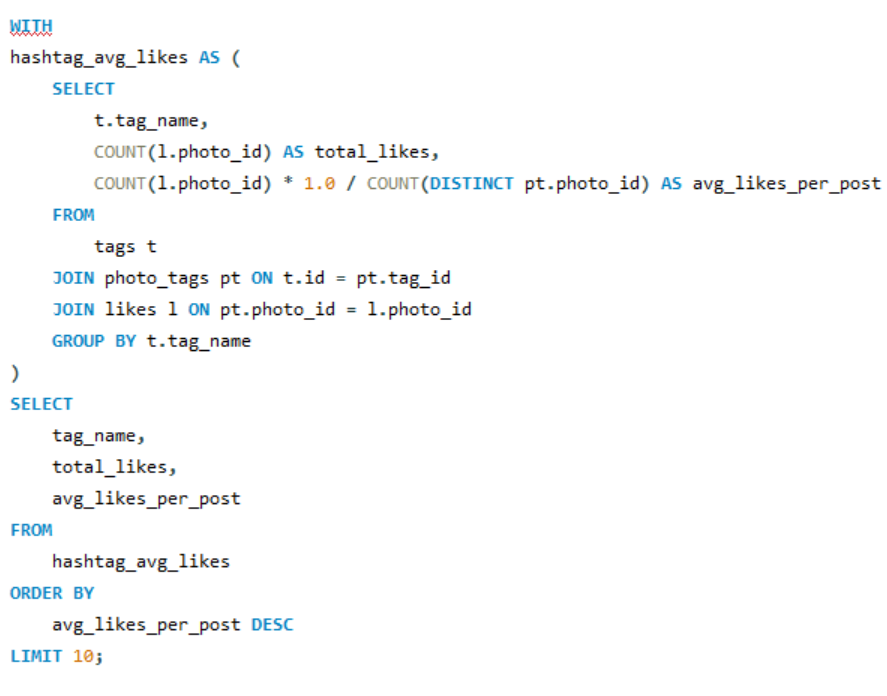
****

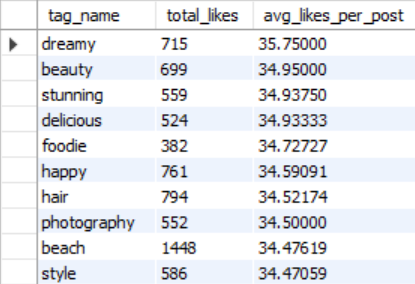
* **Personalized Outreach:**
* Send tailored emails or in-app notifications reminding them of the platform’s value and encouraging them to return.
* Highlight them the popular posts, new features, or trending topics.
* **Content Inspiration**:
* Share personalized content ideas or prompts based on their previous activity.
* Showcase successful posts from similar users to inspire them to create.
* **Feature Highlights:**
* Promote new platform features (e.g., reels, filters, or themes) that can enhance their posting experience.
* Offer tutorials or guides to help them explore the features.
* **Feedback and Surveys**:
* Reach out to understand why they became inactive and address their concerns (e.g., usability issues, lack of relevance).
* **Analyse Drop-off Patterns**:
* Study user data to identify common triggers for inactivity and address them proactively.
* Use push notifications, reminders, or streaks to make engagement a routine.
* **Onboarding Refresh**:
* Re-onboard inactive users by guiding them through updates or features they might have missed since their last activity.

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

**Ans:**

* **Insights:**

****

****

* **High Engagement Hashtags:**
* The hashtag "dreamy" has the highest average likes per post at 35.75, suggesting it resonates well with audiences.
* Other top performers include "beauty" (34.95) and "delicious" (34.93333).
* **Clustered Performance:**
* Most hashtags, such as "beach", "foodie", and "stunning", have engagement rates closely clustered around the 34.5–34.9 range.
* These hashtags are consistent performers but lack significant differentiation.
* **Popular Content Themes:**
* Themes around lifestyle (e.g., "dreamy," "beauty," "style") and food (e.g., "delicious," "foodie") drive significant engagement.
* Hashtags with slightly lower engagement, like **"photography"** and **"hair,"** could be optimized with better-targeted content.
* **Recommendations for Content Strategy:**
* Prioritize creating content that aligns with high-engagement hashtags like "dreamy," "beauty," and "delicious."
* Combine top-performing hashtags to reach broader audiences (e.g., #dreamyFoodie, #BeachStyle).
* Develop visually stunning, aspirational content aligned with the themes of top hashtags.
* Use creative angles for medium performers like "hair" to test for improved engagement.
* **Recommendations for Ad Campaigns:**
* Build campaigns around the most engaging hashtags, such as "dreamy" or "beauty," to ensure alignment with audience interests.
* Use data-driven targeting to show ads to users who have interacted with these topics.
* Integrate hashtags like "beach" or "happy" into campaigns tied to holidays, vacations, or celebrations
* Create user-driven hashtag campaigns around top performers (e.g., #DreamyMoments) to encourage content creation and engagement.

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?**

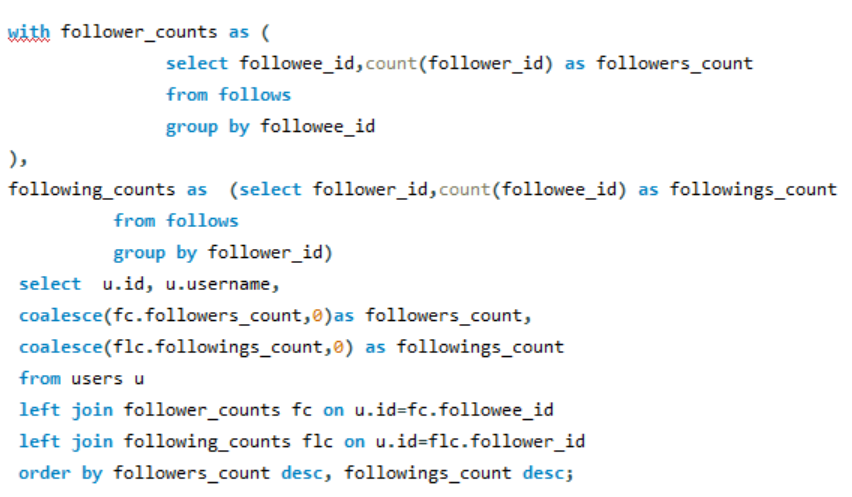
**Ans:**

* The provided schema lacks explicit demographic data (like age, location, or gender), insights can still be derived by combining available user activity data with external demographic information.
* **Insights to Inform Targeted Marketing Campaigns:**
* **Content Timing:**
* Schedule posts during peak engagement times, such as evenings or weekends, to maximize visibility.
* Align content with relevant time-based themes, such as food posts during lunch hours.
* **Demographic-Specific Campaigns:**
* Target younger audiences with visually engaging, trend-focused hashtags (#style, #beauty).
* Design region-specific campaigns leveraging popular location-based tags (#beach in coastal areas).
* **Aspirational and Lifestyle Content:**
* Create aspirational campaigns using high-performing hashtags like #dreamy and #stunning to appeal to users' desire for visually stunning and inspiring content.
* **Experimentation and A/B Testing:**
* Test different combinations of hashtags and posting times to identify the best-performing strategies for specific demographic groups.

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?**

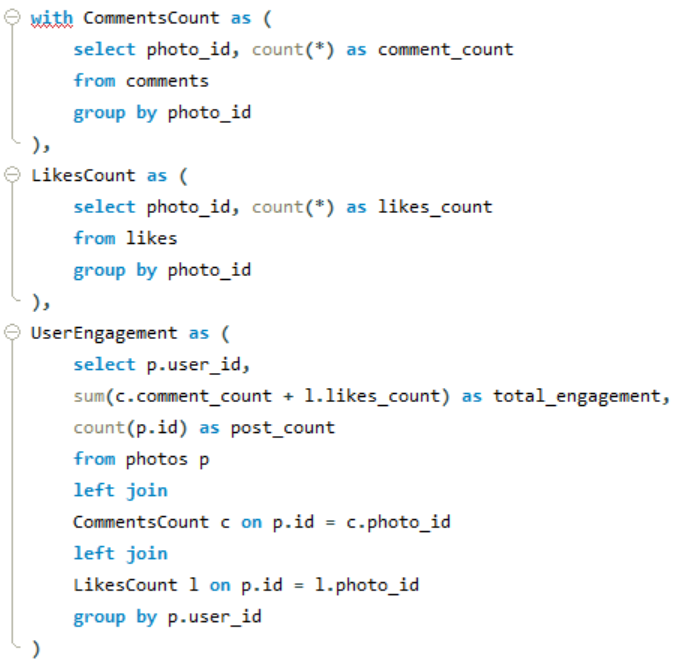
**Ans:**

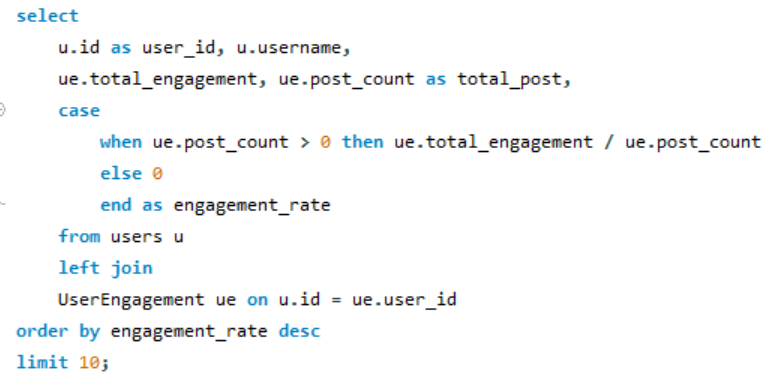
* **Follower Count:** Users with high follower counts (e.g., 77 followers) demonstrate strong potential to reach a wide audience.

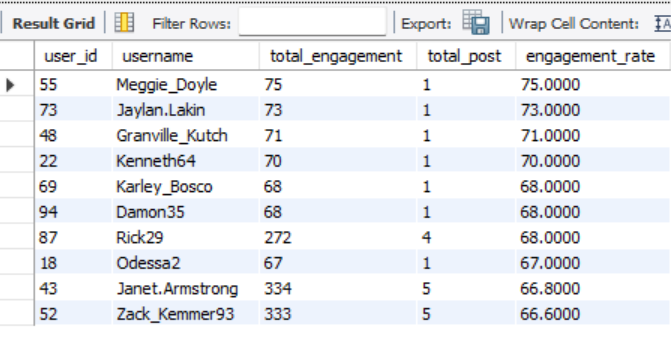
****

****

* **Engagement Rate:** Users with high engagement rates (e.g., above 65%) indicate active and engaged followers, a crucial metric for influencer marketing.

****

****



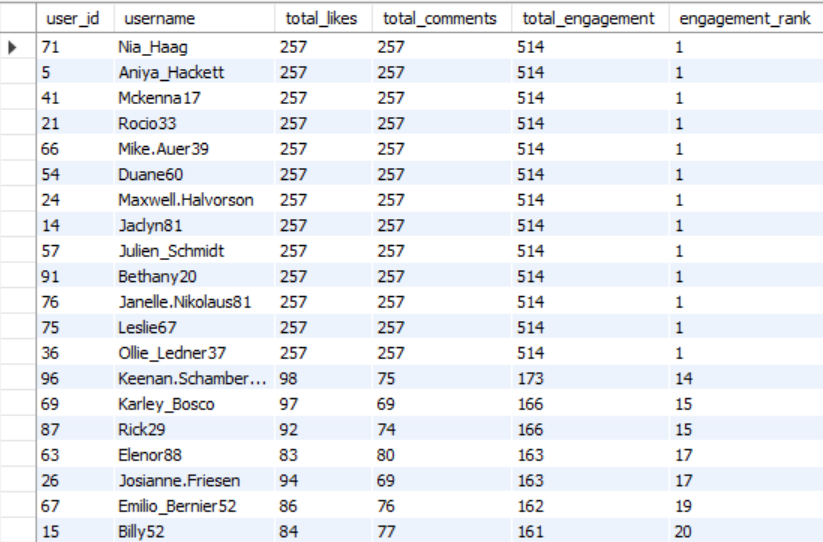
* **Insights:**
* **High Follower Count:**
* Candidates like *Kenton Kirlin* and *Eveline95* with 77 followers each can effectively amplify brand visibility due to their broad reach.
* **High Engagement Rates:**
* Users such as *Meggie Doyle (75%)* and *Jaylan Lakin (73%)* have a smaller following but excellent engagement rates, making them ideal for niche, high-impact campaigns.
* **Balanced Profiles:**
* Users like *Janet Armstrong (66.8%)* and *Rick29 (68%)* combine relatively high engagement rates with consistent activity and a moderate follower base, ideal for sustained campaigns.
* **Approaching and Collaborating with Influencers:**
* **Build Relationships:**
* Reach out with personalized messages acknowledging their influence and content quality.
* Offer exclusive perks (e.g., free products or services) to build goodwill.
* **Leverage Campaigns Based on Strengths**:
* High-follower influencers: Ideal for brand awareness campaigns.
* High-engagement influencers: Effective for driving conversions or product feedback.
* **Incentivize Performance:**
* Offer performance-based incentives (e.g., bonuses for meeting engagement targets) to encourage collaboration.
* **Co-Create Content**:
* Partner with influencers to co-create authentic and relatable content that resonates with their followers, such as product tutorials or reviews.

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

**Ans:**

* Segmenting the user base based on engagement allows for crafting highly personalized campaigns. Targeting power engagers for influence, steady contributors for consistent support, and emerging participants for growth ensures a balanced marketing strategy tailored to user behaviour.

****



* **High Total Engagement & Top Rank:**
* **Users:** *Nia\_Haag, Aniya\_Hackett, Mckenna17, etc.* (514 total engagement, ranked #1)
* **Behaviour:** These users consistently interact with posts, indicating they are highly active and engaged.
* **Marketing Strategy:**
* Use them as brand ambassadors for influencer campaigns.
* **Moderate Engagement:**
* **Users:** *Keenan.Schamberger60 (173 total engagement, rank #14)*
* **Behaviour:** Moderate interaction, likely to engage when content is relevant.
* **Marketing Strategy:**
* Offer personalized content based on past interactions.
* Send time-sensitive offers to increase activity levels.
* **Engagement Across Specific Themes:**
* **Users:** *Karley\_Bosco, Rick29, Josianne.Friesen* (166–163 total engagement, rank #15–17)
* **Behaviour:** Their engagement aligns with specific types of content.
* **Marketing Strategy:**
* Recommend niche products or services based on themes they engage with.
* Create campaigns around their favourite content categories.
* **Low Engagement:**
* **Users:** *Billy52, Emilio\_Bernier52 (161–162 total engagement, rank #19–20)*
* ** Behaviour:** Show some activity but are less engaged overall.
* **Marketing Strategy:**
* Offer onboarding campaigns to make content discovery easier.
* **Recommendations for Targeted Campaigns:**
* Use engagement data to recommend specific hashtags or content categories that resonate with each segment.
* Provide exclusive offers or discounts to steady contributors and emerging participants to increase their activity.
* Engage power engagers as micro-influencers or content co-creators to amplify your campaign reach.
* Use steady contributors and niche enthusiasts for time-sensitive promotions like flash sales or event-based marketing.

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

**Ans:**

* To measure the effectiveness of ad campaigns based on available data such as impressions, clicks, and conversions, you would typically track and analyse several key metrics. These metrics will allow you to assess the current performance and optimize future campaigns effectively.
* **Key Metrics:**
* **CTR (Click-Through Rate):** Measures ad engagement. Higher CTR = better creative/content.

**CTR =** Impressions / Clicks​ × 100

* **Conversion Rate:** Measures success of clicks leading to desired actions (e.g., purchases). Optimize landing pages if low.

**Conversion Rate =** Conversions​ / Clicks × 100

* **CPC (Cost Per Click):** Measures ad efficiency. Lower CPC = better targeting.

**CPC=** Total Spend​ / Total Clicks

* **CPA (Cost Per Conversion):** Measures cost-effectiveness in acquiring a conversion. Optimize targeting or content if high.

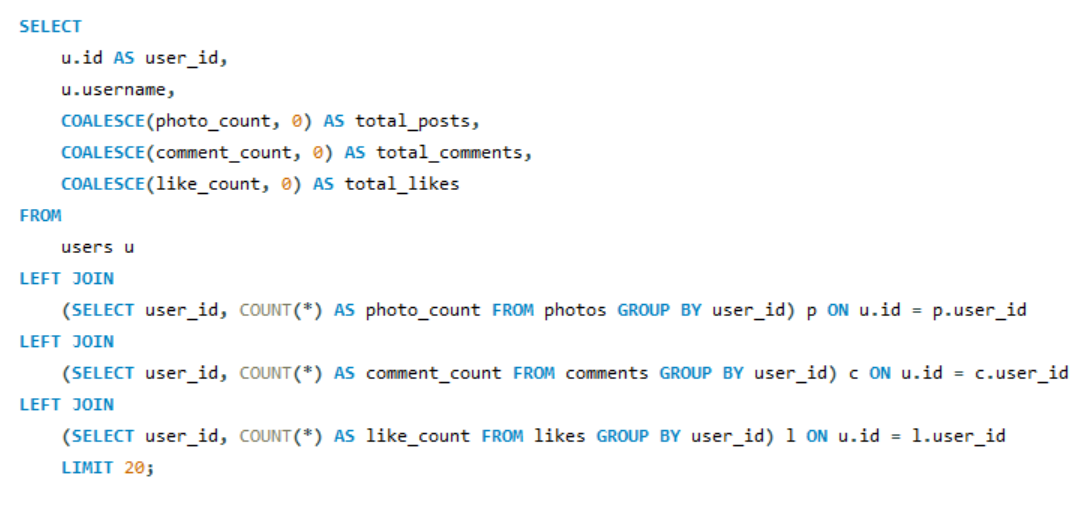
**CPA=** Total Spend​ / Total Conversions

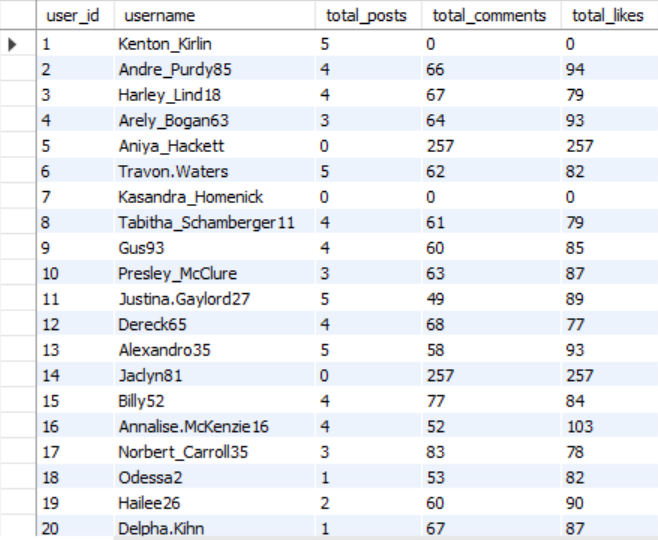
* **ROAS (Return on Ad Spend):** Measures revenue generated per dollar spent. Focus on campaigns with higher ROAS.
* **ROAS =** Revenue from Conversions​ / Total Ad Spend
* **Optimization:**
* **Targeting:** Segment users by demographics and behaviours.
* **A/B Testing:** Test ad visuals, copy, and CTAs to find the most effective combination.
* **Retargeting:** Engage users who clicked but didn’t convert.
* **Ad Timing:** Optimize based on performance by time and placement.
* **Content Optimization**:
* Monitor the engagement with posts that mention specific hashtags or content types (e.g., photos, videos). Leverage high-performing content for future campaigns and consider using the same hashtags or themes in future ads.

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

**Ans:**

* To identify potential brand ambassadors or advocates based on user activity data, we can focus on the following key criteria:

****



* **High Engagement**: Users with high likes, comments, and posts relative to their activity are likely influential.
* **Consistent Activity**: Frequent posting and engagement show a committed user.
* **High Followers Count**: Users with more followers can amplify brand reach.
* **Aligned Content**: Focus on users posting content related to your brand's niche.
* **Positive Interaction**: Comments and interactions that are supportive.
* **Insights:**
* **Target Active Users:** Focus on users with high engagement.
* **Build Relationships:** Reward and collaborate with them to promote campaigns.
* **High Engagement:** Users with strong engagement metrics are likely to be key candidates for brand ambassadorship.

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

**Ans:**

* Without predefined questions, the focus would be on understanding user behaviour and engagement patterns, identifying top-performing content, and optimizing strategies for both user retention and targeted marketing campaigns. By analysing user interactions, activity trends, and engagement, I’d generate insights to enhance content creation and marketing effectiveness.

1. **Understand the Goal:**

* The goal is to analyse user engagement and activity data to optimize content strategies, influencer marketing, and re-engagement efforts.

1. **Data Exploration:**

* Review tables (users, photos, likes, comments, etc.) to understand the structure and ensure there are no missing values.
* Focus on key metrics: total posts, likes, comments, and engagement rates per user.

1. **Metrics to Track:**

* **Engagement Rate:** Calculate likes/comments per post to gauge user interaction.
* Aggregate likes, comments, posts per user.
* **Active vs Inactive Users:** Identify users with high engagement and those with little or no activity.
* **Content Type Preferences:** Track popular hashtags and content types that generate the most engagement.

1. **Analyse Patterns:**

* **Influencers:** Identify users with high engagement rates and a significant following for influencer marketing.
* **Re-engagement:** Target inactive users by offering incentives or personalized content.
* **Content Strategy:** Focus on content types or hashtags with the highest engagement for future campaigns.

1. **Optimize Campaigns:**

* Use data to target ads or sponsored posts to high-engagement users and influencers.
* Create content that aligns with trending hashtags or successful content types.

1. **Sample Queries:**

* **High Engagement Users:** Find users with the most posts and interactions.
* **Trending Tags:** Identify tags with the highest average likes to inform content creation.

1. **Iterate and Improve:**

* Continuously monitor engagement to refine strategies, test new campaigns, and optimize based on insights.

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?**

**Ans:**

* To update the "Engagement\_Type" column in the "User\_Interactions" table, changing all instances of "Like" to "Heart," you can use the following SQL query:

UPDATE User\_Interactions

SET Engagement\_Type = 'Heart'

WHERE Engagement\_Type = 'Like';

* **Explanation:**
* **UPDATE User\_Interactions:** Specifies the table you're updating.
* **SET Engagement\_Type =** 'Heart': Updates the value in the "Engagement\_Type" column to "Heart."
* **WHERE Engagement\_Type = '**Like': Only updates the rows where the current value is "Like."

This query will ensure all instances of "Like" are replaced with "Heart" to reflect Instagram's terminology.