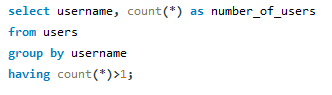
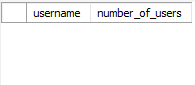
**Social Media Project**

**Objective Questions**

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



Output:



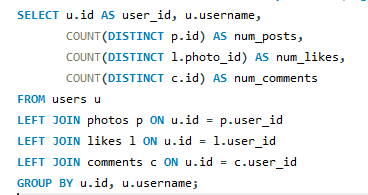
So as the output there is no duplicate in the given data.

To identify missing (NULL) values, you can use the is null condition:

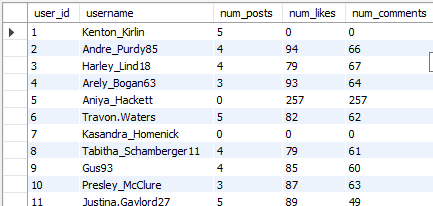


**2. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?**

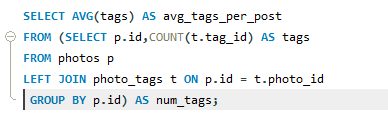
This query retrieves user activity data from a database by joining several tables related to users, posts, likes, and comments. It provides the information for each user.The query performs left joins to include users even if they have no posts, likes, or comments, grouping results by each user (u.id).



Output :



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



Output

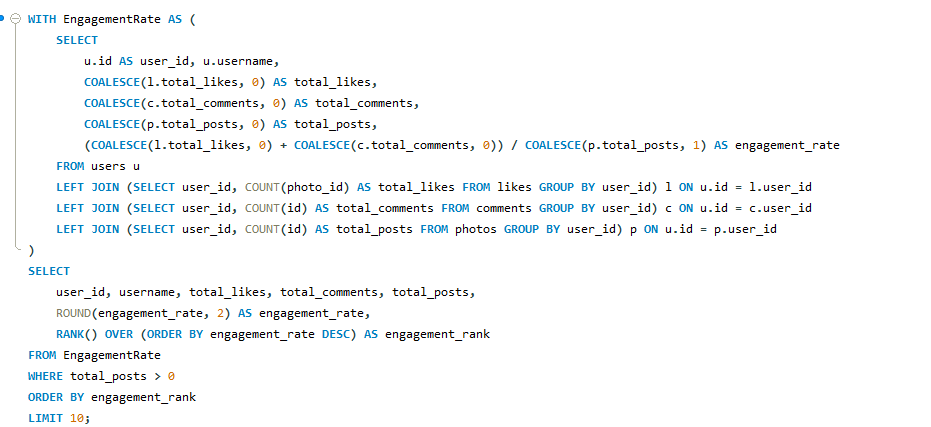


4.Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.

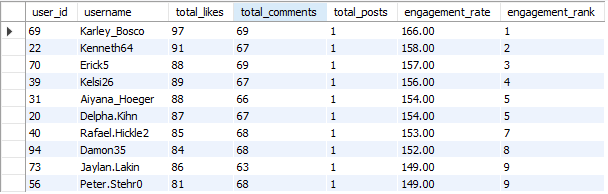
This query calculates the engagement rate of users on a platform, ranking the top 10 users by their engagement. Here's how it works:

1. Engagement Rate Calculation:
   * It pulls user data, counting how many likes, comments, and posts each user has.
   * The engagement rate is calculated by dividing the sum of likes and comments by the
   * Ranks Users:
   * Users are ranked based on their engagement rate, from highest to lowest.
2. Top 10 Users:
   * Finally, it retrieves and shows the top 10 users with the highest engagement rates, excluding users with no posts.

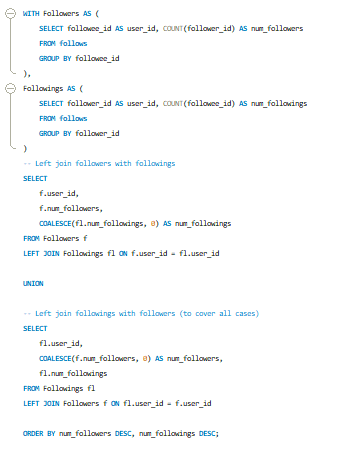
This way, the query identifies the most active users based on their interactions with posts.



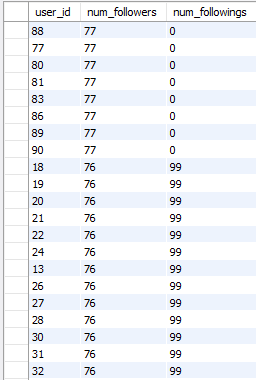
Output:



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



Output:



This query calculates both the number of followers and followings for each user, ranking them based on the highest number of followers, then followings. Here's how it works:

1. Count Followers: - It counts how many people are following each user (`followee\_id`) and stores this as `num\_followers`.

2. Count Followings: - It counts how many users each person is following (`follower\_id`) and stores this as `num\_followings`.

3. Join Data: - It combines the two counts using two `LEFT JOIN` operations. The first join gets the followers along with their followings, and the second ensures users who only have followings but no followers are included.

4. Final Output: - The `UNION` combines these two results to show every user with both follower and following counts, filling in missing values with `0`.

It then sorts the results by the number of followers, and if the followers are equal, by the number of followings.

In short, this query shows each user's follower and following counts, sorted by popularity.

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



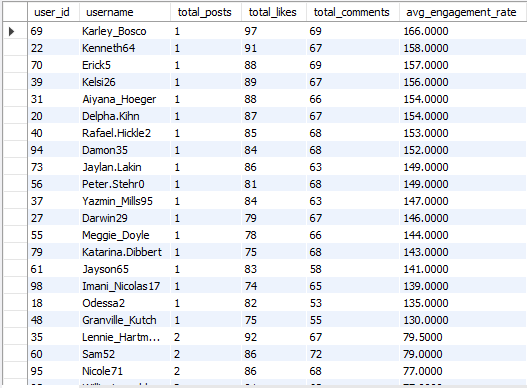
This query calculates the average engagement rate for each user by combining the total number of likes and comments they received on their posts, and ranks them by their engagement rate

Count Posts, Likes, and Comments: It counts how many posts, likes, and comments each user has.

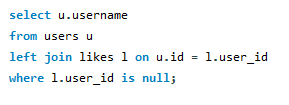
Calculate Engagement Rate: The engagement rate is calculated as the sum of likes and comments divided by the number of posts. If a user has no posts, their engagement rate is set to 0 to avoid division by zero.

Sort by Engagement Rate: The users are ordered by their engagement rate, from highest to lowest.

Output:



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



Output :

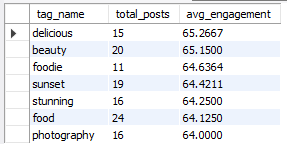


This query returns the names of users who have not received any likes by checking for users where there are no matching records in the likes table.

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



Output:



* Use the query to find hashtags and tags with high engagement. Ads featuring these tags will target content users are already interacting with, increasing relevance.
* Segment users based on popular tags and interests. Customize and add content to reflect the themes and visuals they engage with, enhancing appeal.
* Use user-generated content (UGC) in dynamic ads, showcasing real posts to make campaigns feel authentic and relatable.
* Encourage users to create content using specific hashtags, then feature this in your ads, fostering engagement and community.

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

Yes, there are correlations between user activity levels and content types:

Photos:Higher likes with moderate comments suggest quick, visual appeal.

Videos:More comments and shares indicate deeper engagement with longer content.

Reels:High likes and shares show potential for viral, short-term engagement.

How This Guides Content Strategy:

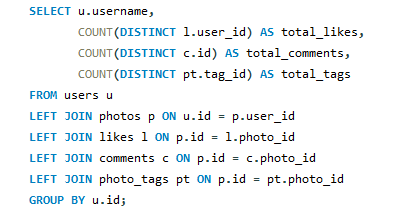
1. Content Mix:Post a variety of photos, videos, and reels to find what drives the most engagement.

2. Trends and Hashtags: Use trending topics and hashtags to boost visibility and interaction.

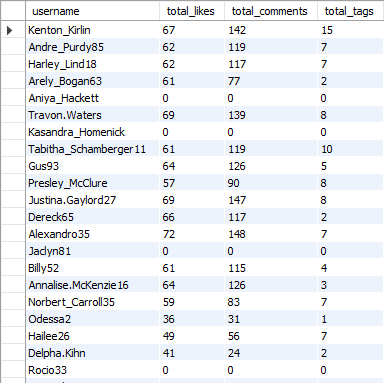
3. Personalized Recommendations: Tailor content suggestions based on user interests and encourage cross-engagement (e.g., recommend reels after photos).

This approach helps in optimizing content for better user engagement.

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



Output :



This query provides a summary of user interactions with photos:

* Counts:
  + Total Likes: Number of unique users who liked the user's photos.
  + Total Comments: Number of unique comments on the user's photos.
  + Total Tags: Number of unique tags associated with the user's photos.
* How It Works:
  + It joins user data with their photos, likes, comments, and tags.
  + It groups the results by each user to aggregate these counts.

In summary, the query gives a detailed view of how many likes, comments, and tags each user’s photos have received.

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



1. Calculate Monthly Engagement:

- The query first creates a temporary table (`MonthlyEngagement`) that gathers user data.

- It joins the `users` table with `likes` and `comments` tables to get each user’s total likes and comments for July 2024.

2. Aggregation:

- For each user, it counts the number of likes and comments, handling cases where there might be no data (using `COALESCE` to default to 0).

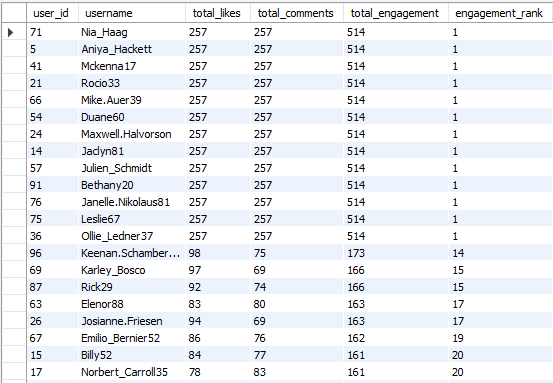
3. Rank Users:

- It then calculates the total engagement (sum of likes and comments) for each user.

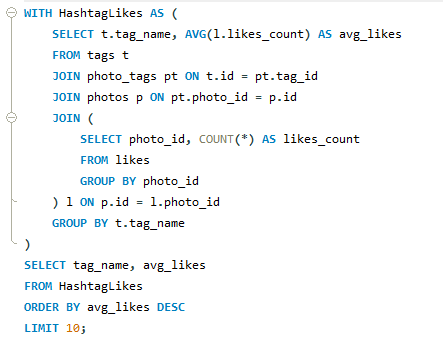
- Users are ranked based on their total engagement, with the highest engagement getting the top rank.

4. Final Output: - The final result shows each user's ID, username, total likes, total comments, total engagement, and their rank. The list is ordered by engagement rank.

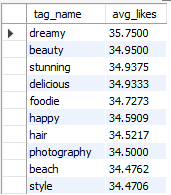
Output :



**12.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.**



Output :



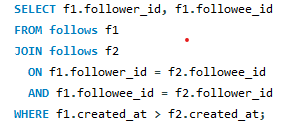
1. Calculate Average Likes by Hashtag:

The query first creates a temporary table (`HashtagLikes`) to compute the average number of likes for photos associated with each hashtag.

It joins the `tags`, `photo\_tags`, `photos`, and `likes` tables to aggregate the number of likes per photo and then average those likes for each hashtag.

2. Final Output: - It then selects the top 10 hashtags with the highest average likes, ordering them from most to least average likes.

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



Output :



This query finds pairs of users who mutually follow each other, where one user followed the other after they were followed back.

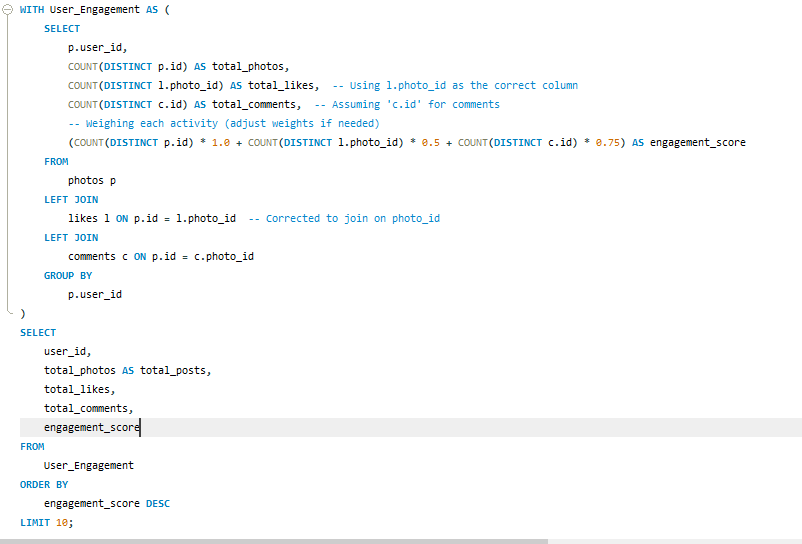
**Subjective Questions**

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?**

Based on the query, loyal or valuable users can be identified by their engagement score, which is calculated by weighting their activities:

* Posts (photos): Users who post regularly are highly active and contribute content to the platform, earning more weight.
* Likes: Users who consistently like content are interacting with others, but this is given slightly less weight.
* Comments: Comments signify deeper interaction with content, thus are weighted more heavily than likes.

The top 10 users, sorted by their engagement score, represent the most active and engaged members of the platform. These users frequently contribute content, like, and comment, making them highly loyal.



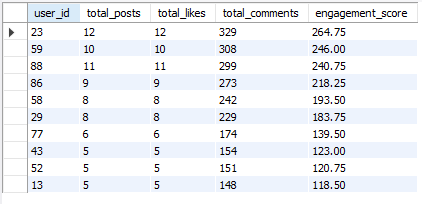
### 2. Rewarding and Incentivizing Loyal Users:

To maintain and boost the loyalty of these valuable users, here’s how you could reward and incentivize them:

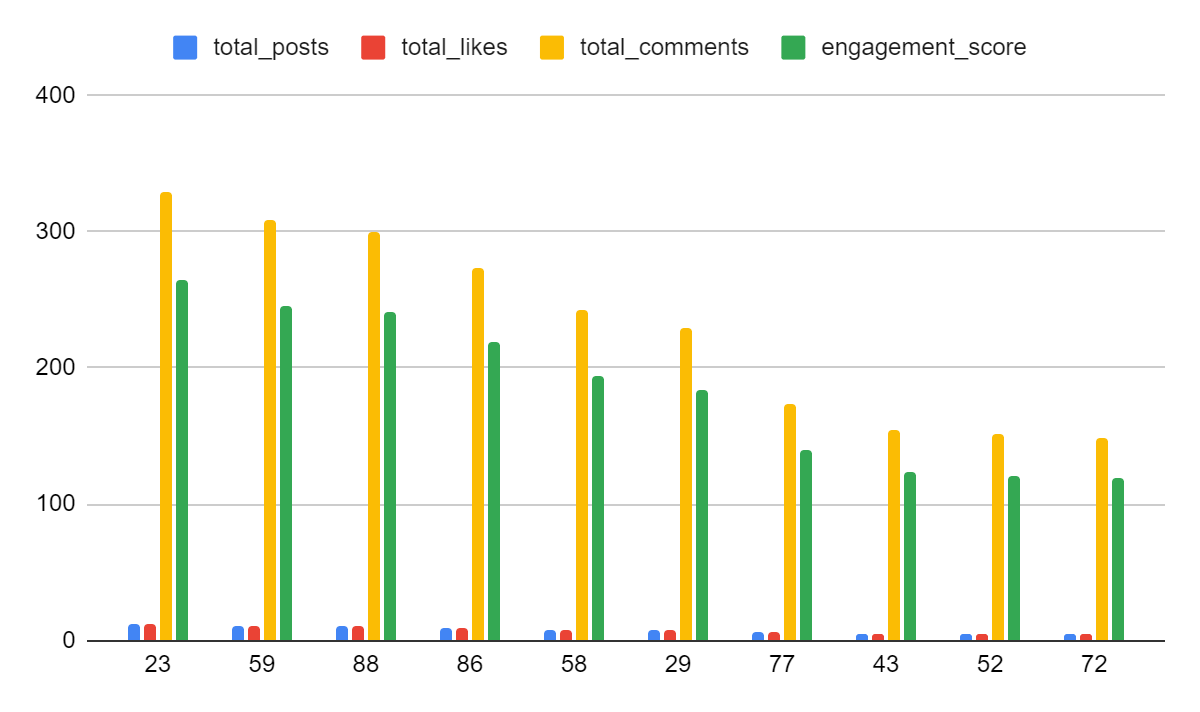
* Exclusive Content: Offer them early access to new features, events, or premium content.
* Recognition: Highlight their contributions on the platform through user spotlights or badges indicating their status as top contributors.
* Discounts or Freebies: Provide exclusive offers, discounts, or free merchandise to these users, acknowledging their loyalty.
* Challenges and Gamification: Introduce challenges or reward systems that further incentivize engagement, like milestone-based rewards (e.g., after 100 likes or 50 comments).
* Personalized Recommendations: Based on their behavior, offer personalized content or recommendations that align with their preferences.

Loyal users are those with high engagement scores, indicating they consistently contribute to and interact with the platform. Rewarding them with recognition, exclusive content, or incentives encourages continued loyalty and engagement.

Output:



Top 10 user\_id based on engagement\_score.



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

To re-engage inactive users and encourage them to start posting or engaging again, a combination of personalized, engaging, and incentive-driven strategies can be highly effective. Here are several strategies that can help:

### **Personalized Re-Engagement Campaigns**

* Email/Push Notifications: Send personalized emails or notifications to remind inactive users about what they’ve missed. Highlight new features, trending content, or updates from people they follow.
* Content Highlights: Show them personalized content, such as popular posts or new content from their connections, to spark interest.

### **2 .Targeted Ads or Messages**

* Social Media Re -targeting: Use ads or targeted messages on social media platforms to remind them about their account and show them what they’ve been missing.

### **3. Surveys and Feedback**

* Gather Feedback: Send surveys to understand why they stopped engaging and address any concerns or barriers they may have experienced.
* User Testimonials: Share stories of other users who became active again and highlight how they benefited from re-engaging

### **4. Community Building and Engagement**

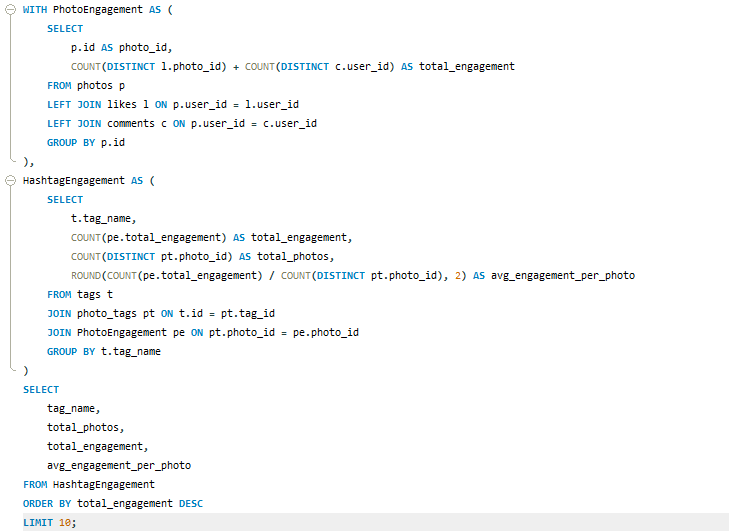
* Invite to Events or Groups: Encourage users to participate in community events, live streams, or specialized interest groups.
* Ask for Feedback: Reach out to inactive users and ask for feedback on why they stopped engaging. This can also make them feel valued and encourage them to return.

### **5. Social Proof and FOMO (Fear of Missing Out)**

* Showcase Popular Content: Display trending posts, popular discussions, or new features to create a sense of FOMO.
* Highlight Friends' Activities: Notify them about the activity of friends or people they follow, to subtly encourage participation.

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

This query is designed to find out which hashtags are generating the most engagement on photos. It does this by first calculating the total engagement (likes and comments) each photo receives. Then, it aggregates this data to determine how engaging each hashtag is by looking at the total engagement and average engagement per photo associated with each hashtag.



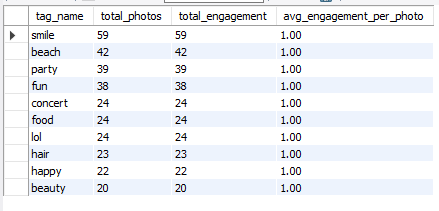
Here's a breakdown of the overall process:

1. Calculate Engagement for Each Photo: For each photo, it counts the total number of likes and comments it has received.

2. Aggregate Data by Hashtag: For each hashtag, it sums up the total engagement of all photos tagged with that hashtag. It also counts how many photos are associated with each hashtag and calculates the average engagement per photo.

3. List Top 10 Hashtags: Finally, it lists the top 10 hashtags based on total engagement, showing the hashtag name, the number of photos, the total engagement, and the average engagement per photo.

output:



**content strategy and ad campaigns**

1. Use the top-performing hashtags with the highest engagement in future content. This increases visibility, interaction, and relevance, ensuring content reaches an engaged audience

.

1. Create more posts around topics related to the high-engagement hashtags. For instance, if fashion-related tags like #StreetStyle or #OutfitOfTheDay has high engagement, you should produce more content or campaigns around fashion.
2. Tailor ad campaigns around the most popular content topics. Ads using highly engaging hashtags can appeal to users already interested in similar content, improving click-through rates and conversions.
3. Collaborate with influencers who use the high-engagement hashtags. Their content will align with trending topics, helping increase the impact of your campaigns.

**4. 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?**

This query is designed to determine when posts receive the most engagement (in terms of likes and comments) based on the time of day and the day of the week. The final result ranks the time slots (hour of the day and day of the week) by average engagement, showing which times are best for getting interaction on posts.

This query analyzes photo engagement by examining how posting times and days affect the number of likes and comments received. Here’s a simplified explanation:

1. Group Photos by Posting Time and Day: It categorizes photos based on the hour they were posted (`post\_hour`) and the day of the week (`post\_day`).

2. Count Photos, Likes, and Comments:

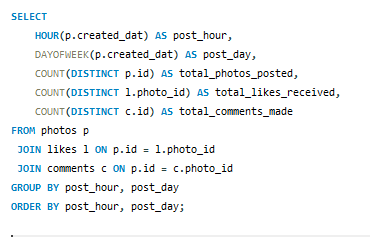
- `total\_photos\_posted`: Counts the number of distinct photos posted during each hour of the day and day of the week.

- `total\_likes\_received`: Counts the number of distinct likes received for these photos.

- `total\_comments\_made`: Counts the number of distinct comments made on these photos.

3. Aggregate Data: The results are grouped by the posting hour and day, so you can see the total photos posted, likes received, and comments made for each hour and day combination.

4. Order Results: The output is sorted by `post\_hour` and `post\_day` to organize the data chronologically.



### **Hourly Patterns (post\_hour):**

* The query groups posts by the hour of the day. If the data shows that most posts are made and receive high likes/comments during certain hours (e.g., evenings or late afternoons), this indicates that users are more active during these times.

### **2. Daily Patterns (post\_day):**

* The DAYOFWEEK function shows how many posts, likes, and comments occur on different days. If weekends (Saturday and Sunday) show higher engagement, it suggests users are more active and interactive during their leisure time.

### **3. Total Photos Posted, Likes Received, Comments Made:**

* The query counts the total number of photos posted, likes received, and comments made for each hour and day. If the engagement (likes/comments) is higher during specific time slots, these are prime opportunities for content to perform well.

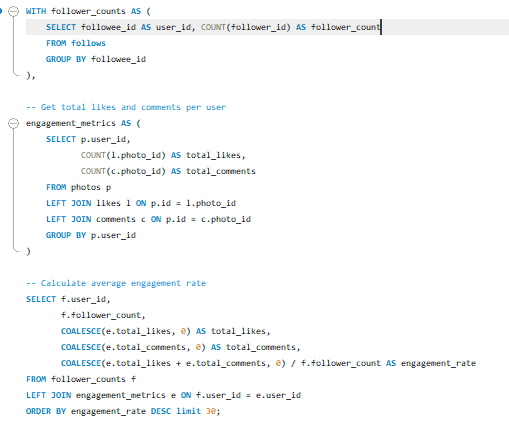
### **Targeted Marketing Campaigns:**

* Optimizing Timing: Post ads or content during high-engagement hours and days, such as evenings or weekends, to increase visibility.
* Content Strategy: Use data to schedule posts when users are most likely to engage, ensuring higher interaction rates for your campaigns.

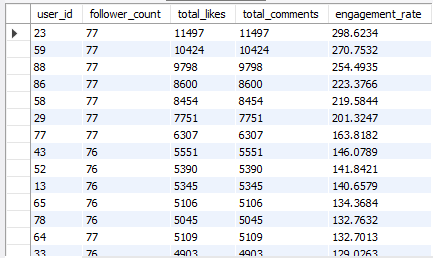
Output:



**5. 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?**



Output :



### Ideal Candidates for Influencer Marketing

**Based on the results of this query:**

* High Engagement Rate: Users with high engagement rates (i.e., a lot of likes and comments relative to their follower count) are likely more effective at engaging their audience. They are ideal candidates for influencer marketing campaigns because they can potentially drive more engagement for promotional content.

### **Approach and Collaborate with Influencers**

1. Identify Top Influencers: Use the query results to identify the top 30 users with the highest engagement rates.
2. Research and Evaluate:
   * Look at the content of these top influencers to ensure it aligns with your brand’s values and target audience.
   * Evaluate their previous collaborations and audience demographics to confirm they fit your marketing objectives.
3. Reach Out:
   * Personalized Proposal: Contact the influencers with a personalized proposal, highlighting why you believe a partnership would be mutually beneficial.
   * Offer Value: Offer incentives such as compensation, exclusive products, or other perks that might appeal to the influencer.
4. Collaborate Effectively:
   * Clear Objectives: Set clear goals and expectations for the collaboration.
   * Creative Freedom: Allow influencers some creative freedom to ensure the content feels authentic.
   * Monitor and Analyze: Track the performance of the campaign and analyze the results to gauge the effectiveness of the collaboration.

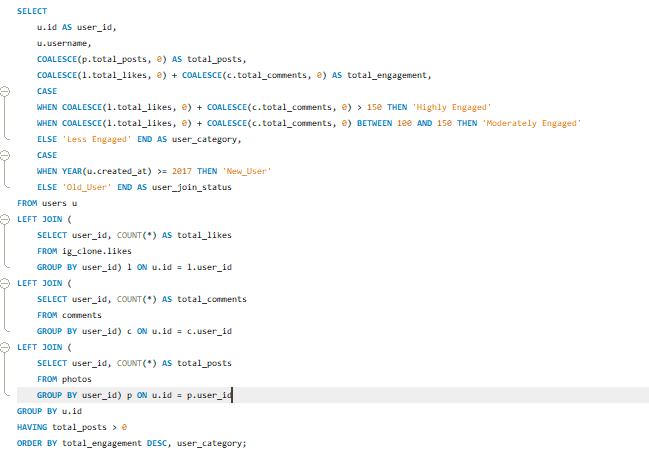
By focusing on influencers with high engagement rates, you increase the likelihood of reaching an engaged and responsive audience, making your marketing campaigns more effective.

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

To segment the user base for targeted marketing campaigns or personalized recommendations, we can derive meaningful insights from user behavior and engagement data based on the query provided. Here’s a structured breakdown of how this segmentation would work:

### **1. Segmentation Criteria:**

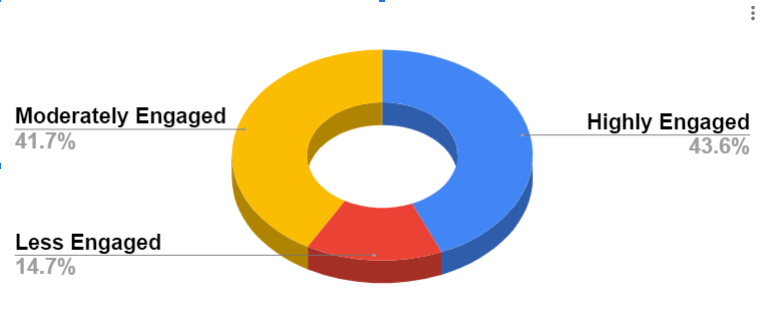
* Total Engagement: This combines the total number of likes and comments a user has made on the platform, which is a key measure of user activity.
* User Category: Based on total engagement, users are divided into three categories:
  + Highly Engaged: Users with more than 150 interactions (likes and comments).
  + Moderately Engaged: Users with 100-150 interactions.
  + Less Engaged: Users with fewer than 100 interactions.
* User Join Status: Users are categorized as either:
  + New\_User: Joined in or after 2017.
  + Old\_User: Joined before 2017.



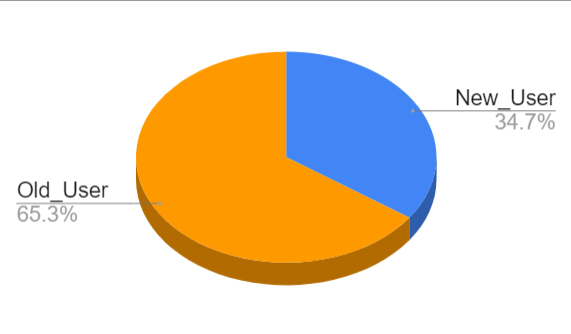
### **Segment for Targeted Campaigns:**

Based on the query, users are divided into segments using two key factors: their engagement level and how long they’ve been on the platform.

* Highly Engaged Users: These users interact with the platform the most and can be targeted with special offers, loyalty rewards, or premium features. Personalized recommendations could be tailored based on their high activity levels, encouraging them to explore more features or make purchases.
* Moderately Engaged Users: This group may respond well to campaigns encouraging them to increase their activity. Recommendations could include suggestions for content they may have missed or incentives to engage more with the platform (e.g., challenges or points for activity).
* Less Engaged Users: For users with lower engagement, reactivation campaigns could focus on bringing them back to the platform. Personalized recommendations can showcase content similar to what they’ve liked before, or special discounts and offers could help re-engage them.



* **New Users:** Users who have recently joined might benefit from onboarding campaigns that familiarize them with platform features and keep them engaged early on. Recommendations could focus on getting them accustomed to interacting with the community.
* **Old Users:** Long-time users could be incentivized with exclusive content or premium offerings that recognize their loyalty. Targeted campaigns may focus on converting them into brand advocates.

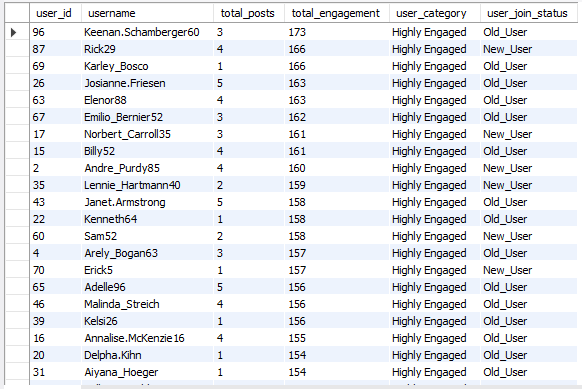


### **Application of Segmentation:**

* Personalized Content: Based on the engagement level and user history, content recommendations can be tailored to match each user's interests and activity level.
* Targeted Offers: Marketers can send offers, promotions, or rewards based on whether the user is highly engaged or less active.
* Retention Campaigns: For less engaged or new users, specific campaigns designed to retain and boost engagement can be created.

Using this data, the user base can be segmented effectively, allowing for highly targeted marketing campaigns and personalized recommendations that cater to different levels of user activity and engagement. This approach ensures that each user receives content and offers most relevant to their behavior and status on the platform, driving both engagement and retention.

Output:



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

To measure the effectiveness of ad campaigns and optimize future ones, follow these steps:

### **1. Measure Effectiveness**

* Impressions: Number of times the ad is shown.
* Clicks: Number of times users click on the ad.
* Conversions: Number of desired actions (e.g., purchases, sign-ups) resulting from the ad

### **2. Analyze Campaign Performance**

Compare Against Benchmarks:

* Evaluate performance metrics against industry standards or previous campaigns.

Segment Analysis:

* Analyze performance by different segments, such as demographics, geographic locations, or devices, to identify high-performing segments.

Attribution Analysis:

* Use attribution models to understand how different touchpoints contribute to conversions.

### **3. Optimize Future Campaigns**

Refine Targeting:

* Adjust targeting criteria based on segment performance to reach more of the high-performing audience.

Improve Ad Creatives:

* Test different ad creatives (e.g., images, headlines, calls-to-action) to identify which resonates best with the audience.

Adjust Bidding Strategies:

* Modify bidding strategies based on CPC, CPA, and ROAS to ensure cost-efficiency.

Test and Iterate:

* Use A/B testing to compare variations of ad elements and refine campaigns based on results.

Monitor Trends:

* Keep track of industry trends and competitors to stay ahead and adjust strategies as needed.

Analyze Conversion Paths:

* Examine the entire user journey from ad click to conversion to identify any obstacles or areas for improvement.

By continuously measuring, analyzing, and optimizing based on these insights, you can enhance the effectiveness of your ad campaigns and improve overall marketing performance.

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

Start by examining key engagement metrics like likes, comments, shares, and mentions. Users who consistently show high levels of engagement with your brand are potential brand ambassadors. Assess their content quality and relevance to ensure it aligns with your brand’s values. Influential users often have high engagement rates, a significant follower count, and rapid audience growth.

### **Analyze User Activity Data**

Dive into user activity data to find potential brand ambassadors. Look for users with high engagement rates and assess the content they create. Check if their audience reach and influence align with your brand’s goals. This analysis helps identify users who can effectively promote your brand and engage their followers.

### **Segment Potential Brand Ambassadors**

Segment users into categories like top engagers, content creators, and brand advocates. Top engagers have high interaction rates, content creators frequently post relevant content, and brand advocates actively promote and defend your brand. Tailoring your outreach to these segments ensures you target the most effective candidates for brand ambassadorship.

### **Evaluate Alignment and Fit**

Evaluate whether potential ambassadors align with your brand’s values and target audience. Ensure their content and behavior fit with your brand’s messaging. This alignment is crucial for authentic and effective collaborations that resonate with your audience and enhance your brand’s image.

### **Engage and Collaborate**

Reach out to potential ambassadors with personalized messages that explain the benefits of collaboration. Offer incentives such as exclusive access or financial compensation. Building a strong partnership through personalized engagement helps in leveraging their influence to promote your brand’s initiatives.

### **Monitor and Measure Impact**

Track the impact of collaborations by measuring key performance indicators like engagement metrics, event participation, and sales. Analyze this data to evaluate the effectiveness of the partnerships and adjust your strategy as needed. Continuous monitoring and refinement help optimize future collaborations and maximize benefits.

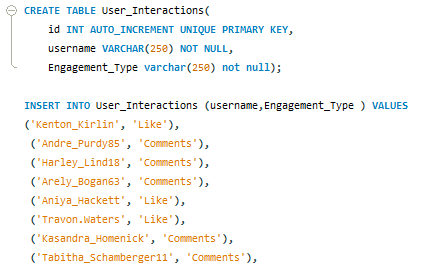
**9. How would you approach this problem, if the objective and subjective questions weren't given?**

If the objective and subjective questions weren't provided, the approach to analyzing the problem would involve several structured steps to extract insights from the data, which would be as follows:

1. **Problem Identification:**
   * Objective: Start by clearly understanding the overall goals and desired outcomes of the analysis. This could involve identifying whether the goal is to optimize user engagement, boost retention, identify key influencers, or any other business objective.
   * Planning: Based on the goals, outline a plan and action steps. For instance, decide whether the focus should be on analyzing user behavior, content performance, or engagement patterns.
2. **Data Exploration:**
   * Initial Analysis: Begin by exploring the raw data to understand its structure and attributes. This includes looking at distributions, trends, and anomalies in the dataset. Initial insights might reveal certain patterns, such as peak user activity times or the most common types of user actions.
   * Data Structure: Analyze the tables, relationships, and key fields in the dataset to determine how to proceed with further analysis.
3. **Data Cleaning and Validation:**
   * Data Quality Check: Identify and resolve any missing values, duplicates, or inconsistencies in the data. Clean data is crucial to ensuring accurate analysis and results.
   * Validation: Ensure that the data is consistent across tables and fields. Validate key metrics like user IDs, timestamps, and engagement counts for integrity.
   * Hypotheses Development: Develop hypotheses based on initial observations or business goals. For example, a possible hypothesis could be "users who post frequently are more likely to receive higher engagement." Each hypothesis serves as a guideline for running analyses.
   * Exploratory Analysis: Check for trends that could validate or disprove the hypotheses. For instance, this could involve examining the relationship between user activity levels and engagement (likes, comments).
   * SQL Analysis: Utilize SQL queries and operations to derive insights from the data. This would involve running queries to analyze trends, check engagement levels, and understand how various factors impact user activity. The analysis should focus on identifying key influencers, high-engagement content, and user retention patterns.
   * Adjustments: Based on the insights from the data, adjust the approach if necessary. If the initial hypotheses don’t hold true, modify the direction of the analysis to explore alternative trends.
   * Report Generation: Finally, gather the insights from the SQL analyzes into a comprehensive report. The report should detail the current state of user behavior, engagement trends, and key findings.
   * Recommendations: Highlight areas that need improvement, such as low-performing content or user segments with declining engagement. These insights should guide management in making informed decisions and taking appropriate actions to achieve the desired outcomes.

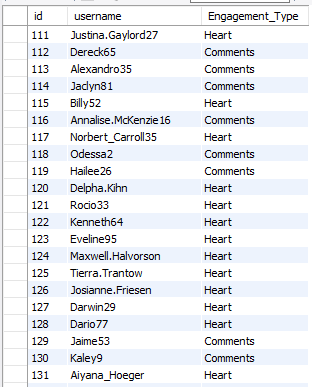
Without predefined objective and subjective questions, this step-by-step approach allows for a systematic and structured analysis of the problem. By exploring the data, forming hypotheses, and using SQL to test those hypotheses, one can draw actionable insights and make data-driven recommendations.

**10.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**

?



Output:



* SET SQL\_SAFE\_UPDATES = 0;  
  This disables safe updates, allowing you to make changes to the table without needing a key column in the WHERE clause. This is necessary if safe updates are enabled in your SQL environment.
* UPDATE User\_Interactions  
  Specifies the table you want to update.
* SET Engagement\_Type = 'Heart'  
  Sets the new value ('Heart') for the Engagement\_Type column.
* WHERE Engagement\_Type = 'Like';  
  Specifies the condition to identify which rows to update. In this case, it updates only rows where the Engagement\_Type is currently 'Like'.
* After running this command, all instances of "Like" in the Engagement\_Type column will be changed to "Heart".