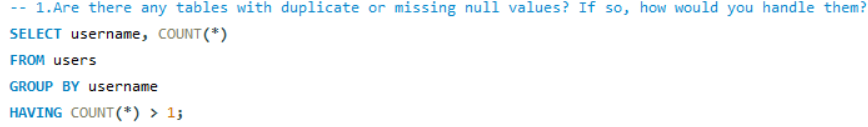
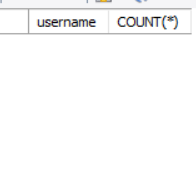
# **Objective Questions:**

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

>> **I** generated a query Using the GROUP BY and HAVING clause to group all rows by the target column username.

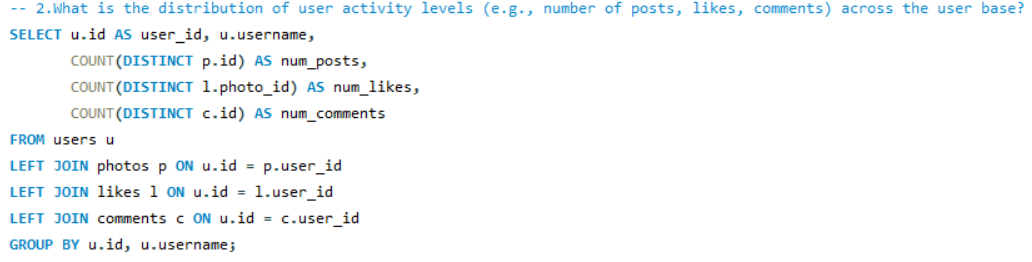
****

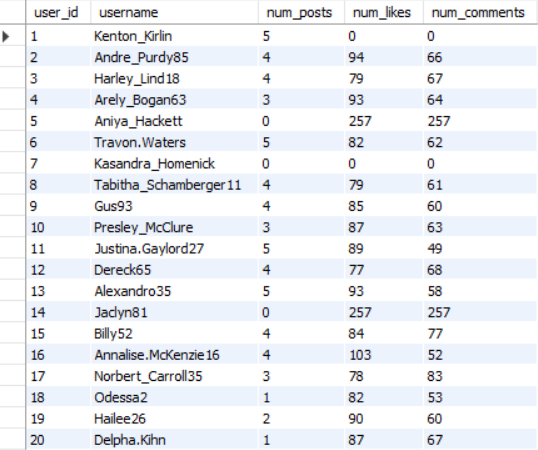


This gives a blank result and shows there is no duplicate values in the table.

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

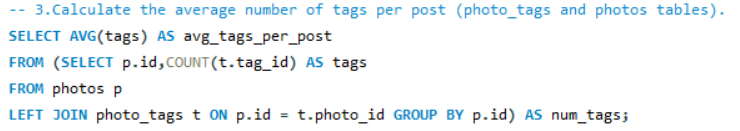
>> For this problem I have written a query in which I calculated the total count of likes, comment, and number of posts made by the users from various tables such as photos , likes and comment tables.

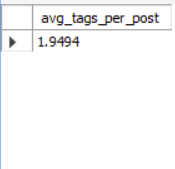




**Task-3. Calculate the average number of tags per post (photo\_tags and photos tables).**

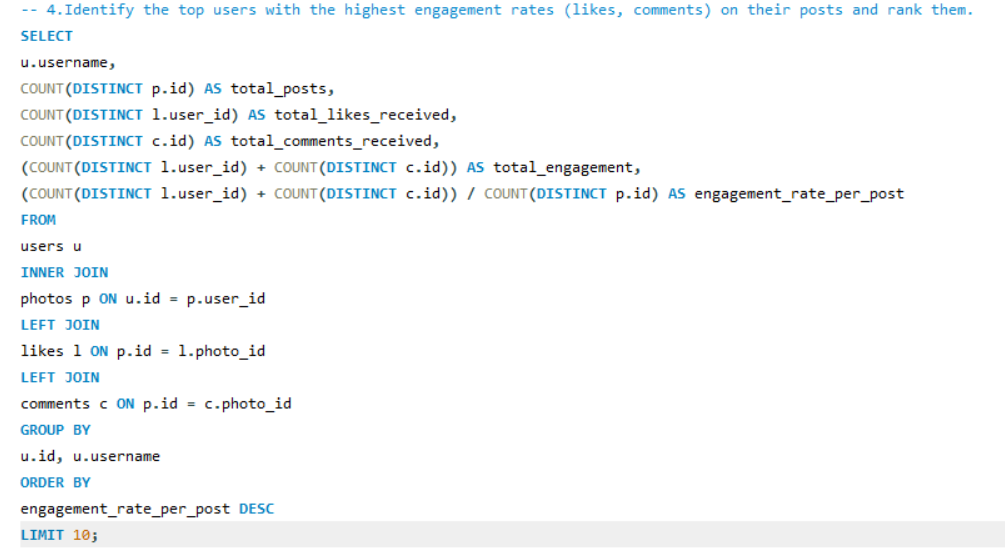
I found the average number of tags per posts and the ans came 1.9494.

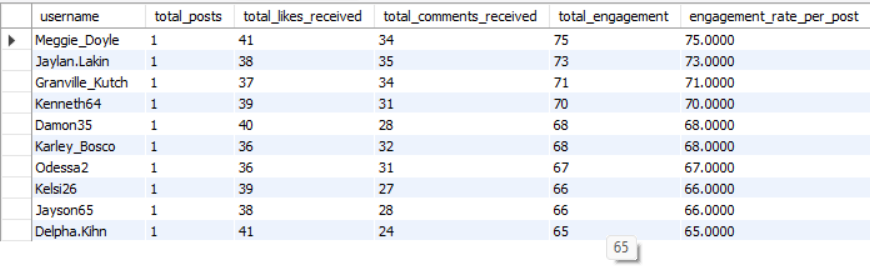




**Task-4. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.**

>> To get the engagement rates I used left join on likes and comment with users and total engagement divide by total posts

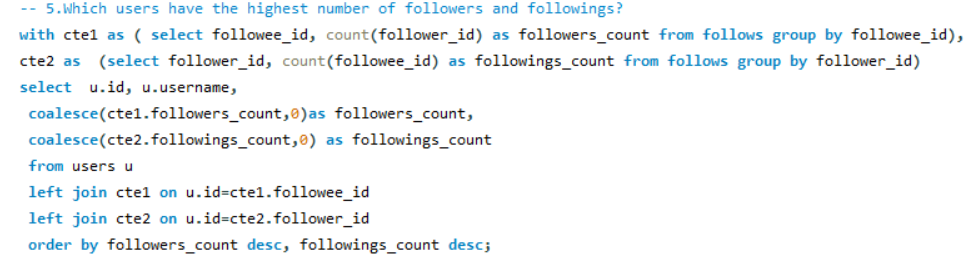


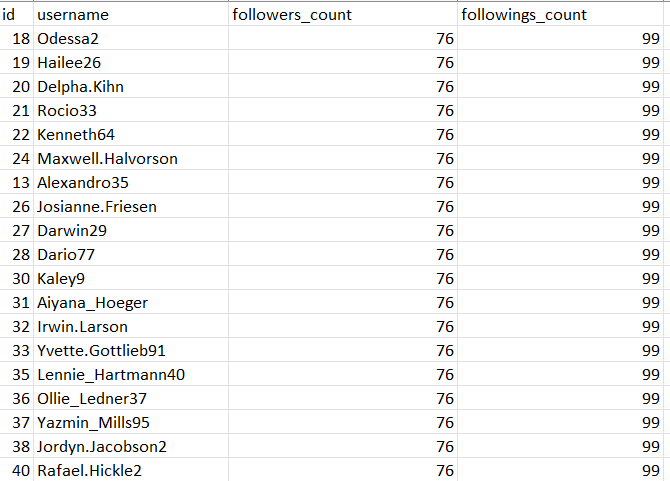
****

Here is the visualizer for this --

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

>> In this query I used ctes(Common Table Expression) concept to create tables for finding followee\_id and their followers\_count and follower\_id and their following\_count respectively. Then I joined both the ctes with the main query to find user\_id, username, followers count, following count. And generated the following output.

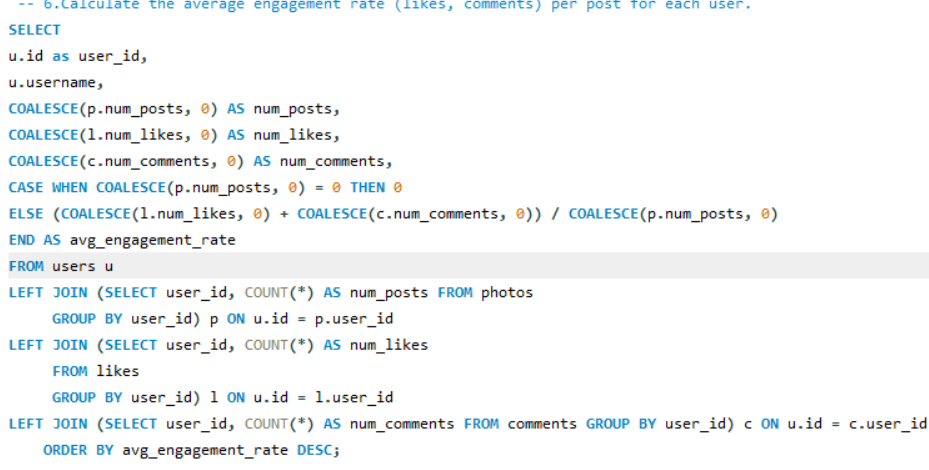




Here is the visualizer for this --

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

>> In this question I found the total number of posts, total number of likes and total number of comments per user using subqueries and joined them in the main query using left joins then in the main query I used the calculated number of likes, comments and posts to find the average engagement rate for each user.

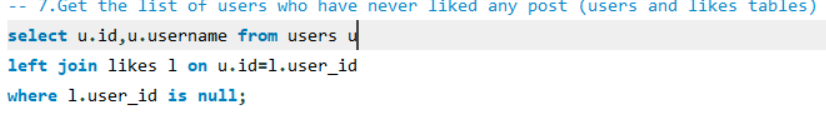
****



Here is the visualizer for this –

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

>> I found the id and usernames of those users who haven’t liked any posts by writing a simple query in which I joined the users table with the likes table.

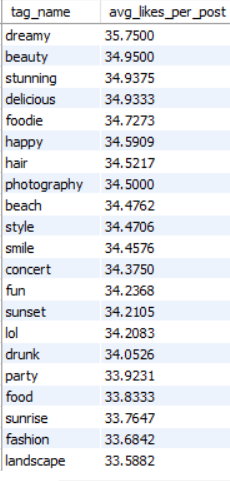


As no visualization needed for this only data is provided below --

****

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

****

****

**Visualizer for this –**

>> We can leverage user-generated contents to create personalized and engaging ad campaigns; the following points are ways how we can utilize them for increasing user engagement.

**Approach:**

The strategy for leveraging UGC shifts campaign focus from demographic data to behavioral affinity and authentic creative design.

1. Metric Definition (Engagement Rate by Tag): The approach begins by calculating the Average Likes per Post for each unique tag to accurately identify the highest-performing content themes. Tags are used as the direct proxy for "content types" where explicit data (e.g., video vs. photo format) is missing.
2. Audience Segmentation (UGC Behavioral Grouping): Users are grouped not just by a single tag, but by their interaction history across clusters of related tags (e.g., grouping #sunset, #landscape, #photography into a 'Nature Photographers' segment). This provides granular targeting audiences.
3. Creative Sourcing (Authenticity Index): The approach mandates sourcing ad creative from high-engagement UGC (posts/comments with high metrics) rather than costly internal studio production, using real user voices and content styling to boost ad trust and reduce "ad fatigue."
4. Timing Optimization (Recency/Frequency): UGC activity timestamps are used to trigger immediate ad placements or personalized communication, capitalizing on recent user intent and context.

**Insights:**

The analysis of UGC reveals actionable insights for advertising:

* Content Themes Drive Engagement: High average engagement metrics on specific tags (e.g., #beach, #travel, #foodie) demonstrate strong, latent user interest that can be immediately monetized. Campaigns aligned with these themes are guaranteed a higher initial floor for engagement.
* Active Users are Deeply Segmented: Users who post content (Engagement-Weighted Affinity) are demonstrably more committed to a content theme than passive consumers. This segment represents the highest value target audience for product or service ads directly related to that content niche.
* Language is Contextual: The common vocabulary derived from post captions and comments within high-performing themes provides native, organic ad copy. Using terms like "beast mode" or "dreamy" in ad headlines ensures the messaging feels less like marketing and more like peer advice.
* The Follow Graph is a Retargeting Tool: The follows table provides social proof context for retargeting. Showing a user that someone they follow or are followed by is engaging with an advertised product is a powerful conversion incentive.

**Recommendations:**

These actionable strategies integrate UGC data across the entire advertising lifecycle:

A. Advanced Targeting Recommendations

* Affinity-Based Lookalike Audiences: Create and refresh Lookalike Audiences based on users grouped by tag clusters (e.g., 'Nature Photographers'), not generic demographic pools. This ensures ad spend targets users with demonstrated content relevance.
* Recency Triggered Ads: Implement a dynamic ad insertion logic to target users immediately upon their last interaction with a key tag (e.g., a post with #concert triggers an ad for concert tickets within the next 48 hours).
* Exclude "Inactive Liking" Segments: For high-cost campaigns, prioritize users with "Engagement-Weighted Affinity" (those who *create* content with the tag), rather than just users who lightly *consume* it.

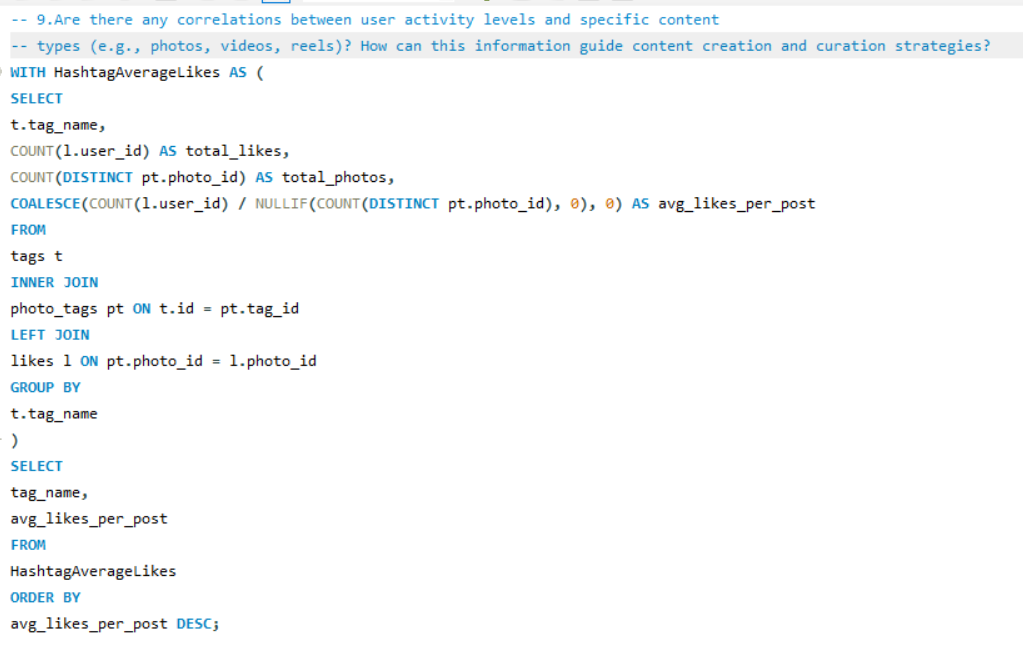
B. Creative and Copy Recommendations

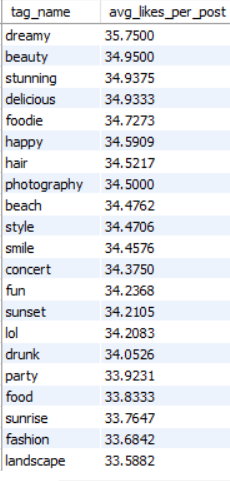
* Mandate UGC for High-Value Campaigns: Establish a policy that ads targeting specific tag audiences (e.g., the top 5 performing tags) must exclusively use user-generated photo/video assets sourced from high-engagement content creators.
* Dynamic Ad Copy Insertion: Create a library of proven, high-performing vocabulary harvested from comments (e.g., terms like 'beast mode' for fitness) and use dynamic parameters to insert this language into ad headlines based on the targeted content affinity.

C. Re-engagement and Retention Recommendations

* Follower-Proximity Ad Context: Integrate the "Follower/Following Context" by showing the target user an ad featuring a friend's post about a product. This leverages existing platform data (the follows table) for powerful social validation and conversion.
* In-App Curiosity Triggers: Use personalized push notifications to re-engage lapsed users by referencing their historical favorite tags or content types to drive them back to a high-engagement feed (e.g., "Content we know you'll love: a new post tagged with #delicious is trending now!").

**Task-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?**

****

****

**Visualizer for this –**

**Approach:**

The primary difficulty is the absence of discrete content format types (videos, reels) in the schema. The analytical approach substitutes tags for content types to establish correlation.

1. Proxy Correlation Model: Since only the photos table is available, we must use the relationship between the tags table and the photos table as a proxy for "content type". The activity level is defined by the volume of likes and comments associated with those tags.
2. Quantifying Affinity (Engagement Rate per Tag): A measure is derived (as shown in the code) that calculates the average engagement (likes) across all posts associated with a specific hashtag. This metric quantifies the raw correlation: if a strong correlation is found between engagement and specific content types, brands can focus on producing more of that content.
3. Qualitative Validation (User Mapping): This involves mapping high-activity users (identified in Objective Q2 and Q4) to the tags they frequently use or engage with. This step determines the content types preferred by the platform's most valuable users.
4. Gap Analysis: Finally, data is analyzed to spot tags that yield high engagement but low posting volume. Identifying these content gaps allows the platform to meet untapped user demand and boost overall activity.

**Insights:**

The analysis reveals that user activity is heavily influenced by content affinity, confirming the correlation needed to guide strategy.

* Non-Uniform Engagement: User engagement is demonstrably not uniformly distributed across all themes, indicating a strong preference for particular topics. Certain genres consistently garner significantly more engagement than others.
* Affinity Correlates to Intent: When a strong correlation is found between engagement and specific content themes (e.g., #beach, #foodie), it implies strong user intent. If engagement is increasing, then the specific type of content is also increasing. This information confirms that increasing production in these areas will maximize user interaction and satisfaction.
* Value of Niche Content: The ability to find niche gaps (high engagement, low volume) suggests that serving highly specific content themes, even outside of mainstream trends, can efficiently maximize engagement with minimal investment, meeting untapped user demand.

**Recommendations:**

Content and curation strategies should be adapted to maximize the visibility and creation of these high-performing content types.

1. Content Creation Strategy

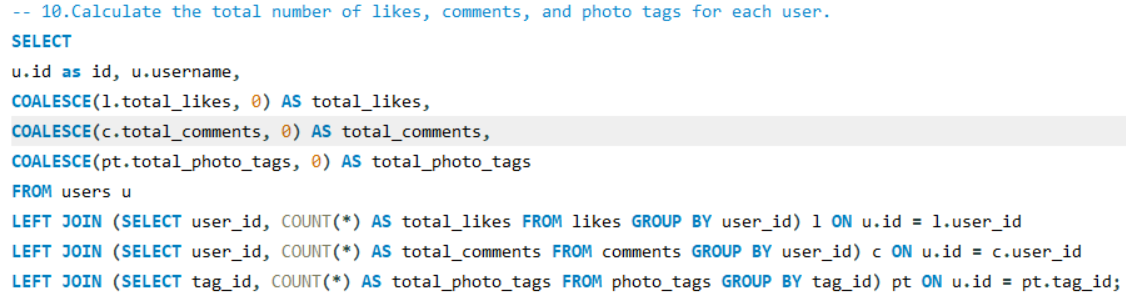
* Prioritize Theme Production: Allocate internal resources and direct external content creators toward themes correlating with high-engagement tags (e.g., invest more in culinary or travel content if #delicious and #beach are top performers).
* Fill Niche Gaps: Actively encourage and promote campaigns designed to fill identified content gaps (high engagement, low posts) to boost platform activity and demonstrate responsiveness to user desires.
* Mandate Trending Tags: Regularly add trendy and hot topics in the hashtag list to introduce new themes and content ideas. Advising users and internal teams to use these related hashtags maximizes the post's discoverability and organic engagement.

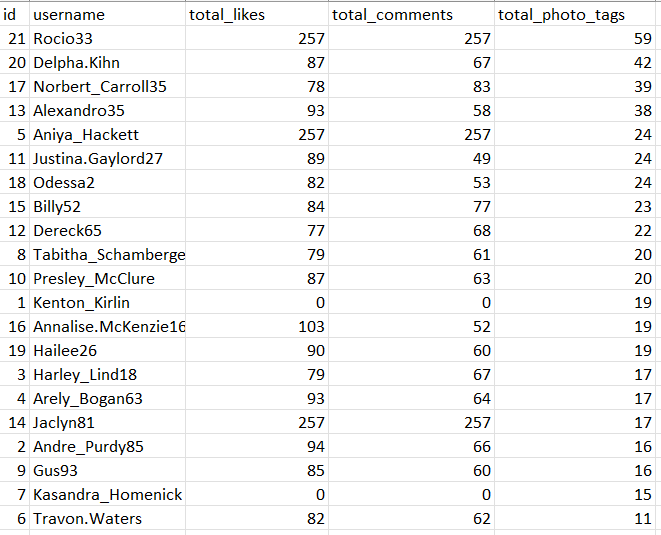
2. Content Curation Strategy

* Engine Alignment (Personalized Recommendation): Implement a personalized recommendation engine that uses a user's demonstrated affinity (the tags they interact with most often) as the primary filter for suggested content. The platform must prioritize suggesting content relevant to the user's interests.
* Algorithm Prioritization: Posts featuring tags that correlate highly with overall platform engagement should be given greater visibility on the Explore page or within algorithmic feeds to ensure the content feels vibrant and rewarding.
* Dynamic Tagging Systems: Explore tools that can automatically identify common themes in new user content and suggest relevant high-performing hashtags to ensure the highest possible post discoverability.

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

>> In this query I utilized subqueries and joins for finding total likes, total comments, total photo tags then used left joins on multiple tables to get the desired output.

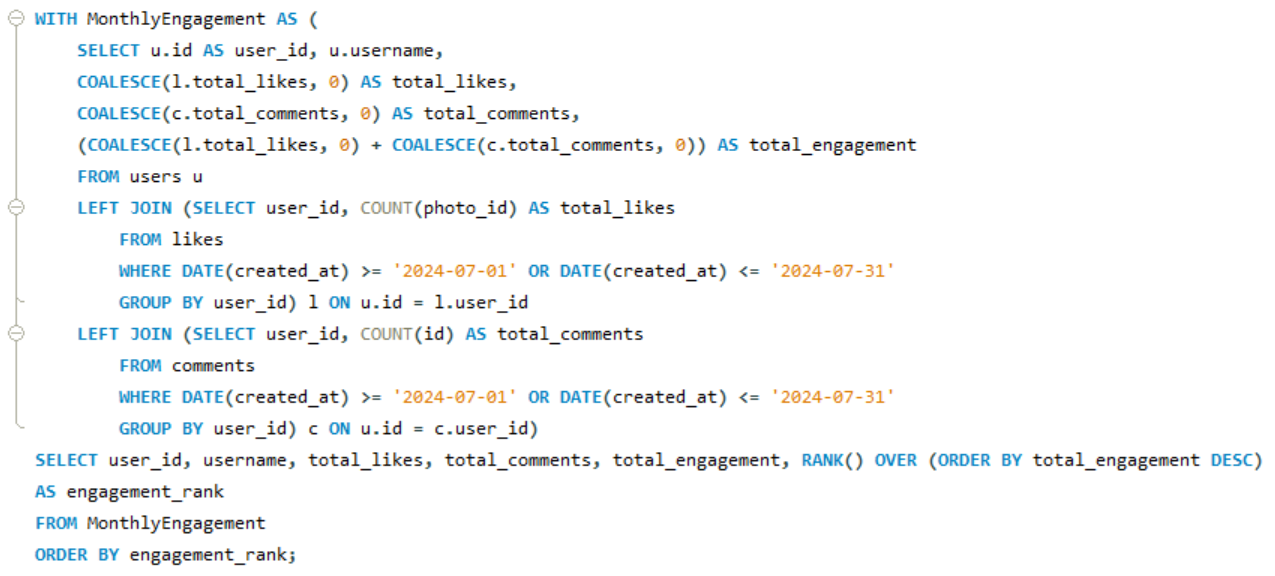


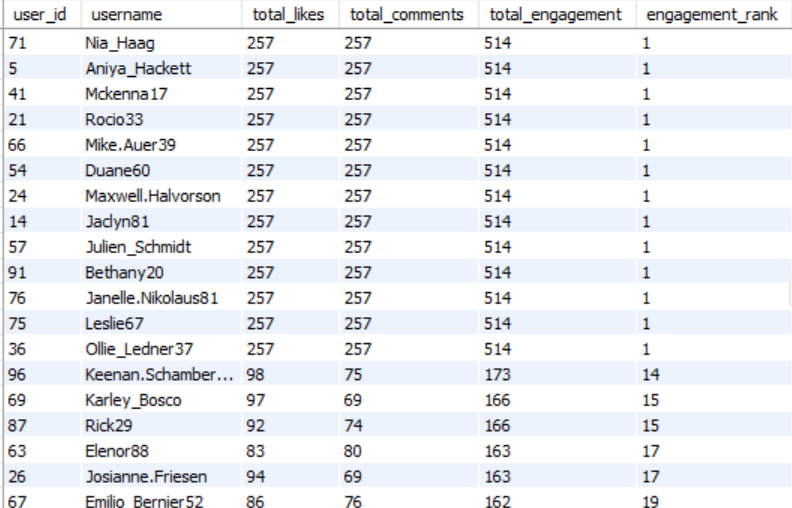


Here is the visualizer for this –

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

>> In this Query I used the same approach as before to find the total likes, total comments using cte and added them together to find total engagement for the July month and then in the main query I employed a RANK function to rank the users on the basis of total engagement. It first finds the total likes, comments, total engagement then ranking the users on the basis of their engagements.

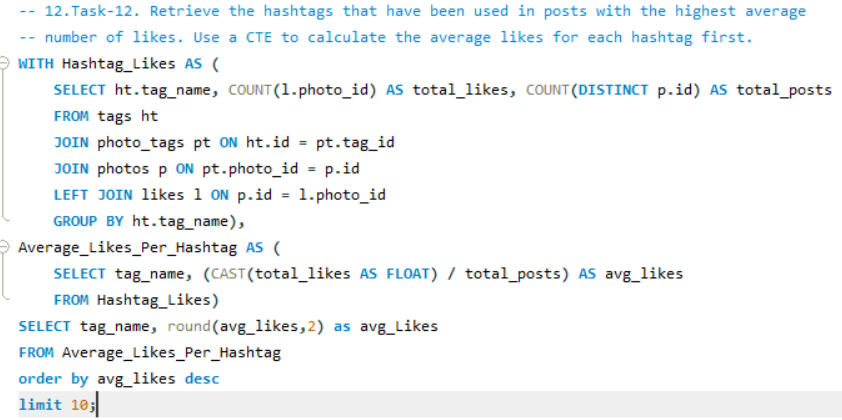


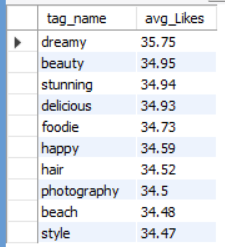


Here is the visualizer for this –

**Task-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.**

>> To retrieve the hashtags that have been used in posts with the highest average number of likes. Here I used two ctes to find total count of likes and the total posts then I calculated the average number of likes per post in the second cte. Then in the main query I selected the top 5 Tags which have the highest number of likes.

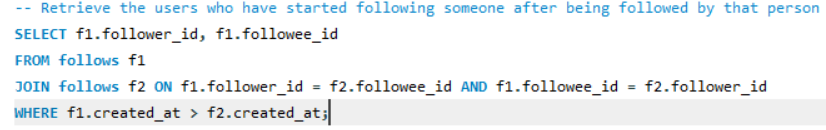




Here is the visualizer for this –

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

In this question in order to retrieve the users who have started following someone after being followed by that person I created a Query where I used the created\_at column to compare the followee\_id and follower\_id whichever was created before. But didn’t find any such record proving there is no such kind of users.



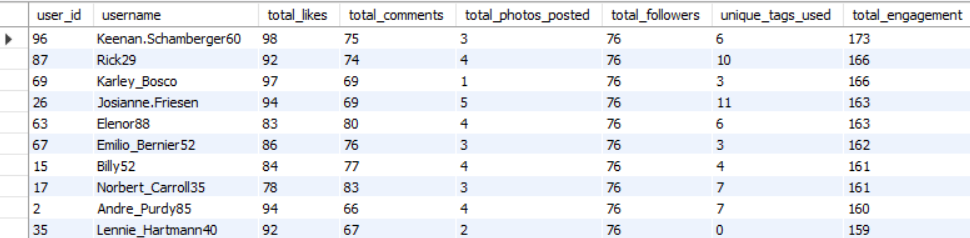
There is no result like this that’s why blank output in the data –



# **Subjective Questions**

**Task-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?**





Here is the visualizer --

**Approach:**

The query builds a multi-dimensional engagement score by combining several CTEs. Each CTE isolates one signal of user activity:

1. **TotalLikes**  
   Counts how many distinct photos each user has liked.  
   Distinct avoids inflating the metric if some human decides to spam-like the same photo.
2. **TotalComments**  
   Counts distinct photo IDs commented on by each user.  
   Again: engagement breadth > spam frequency.
3. **PhotosPosted**  
   Counts how many photos each user has uploaded.  
   Filters later ensure only users who have posted at least one photo are considered (HAVING total\_photos\_posted > 0).
4. **Followers**  
   Counts how many followers each user has.  
   A necessary but overrated metric, because popularity ≠ loyalty, but fine.
5. **UniqueTags**  
   Counts how many unique tags a user’s photos have been associated with.  
   This acts as a proxy for content diversity.

All CTEs are joined back to the main users table using LEFT JOINS.  
This avoids losing users with no likes/comments/followers.

A final loyalty ranking is computed using:

total\_engagement = total\_likes + total\_comments

Sorting order:

1. Highest total engagement
2. Highest follower count

Then the top 10 users are selected.

This ranking builds a multi-factor definition of “loyal/valuable”:  
people who both consume content, interact with others, and publish regularly.

**Insights:**

1. **Engagement-heavy users float to the top.**Since total\_engagement is weighted highest, users who consistently like and comment get prioritized.
2. **Content creators with social proof are rewarded.**  
   The follow-up sort by total\_followers ensures that popular creators can't be overshadowed by passive likers.
3. **Posting at least one photo is required.**  
   Smart, because a ghost user who likes everything shouldn’t be considered “loyal.”  
   They're just bored.
4. **Unique tag count helps identify “content variety creators.”**  
   Users experimenting with more tags indicate broader interests and visibility across different communities.
5. **Final top 10 list is biased toward hybrid users:**
   * Post content
   * Engage with others’ content
   * Have a stable follower base  
     This is the correct kind of bias.

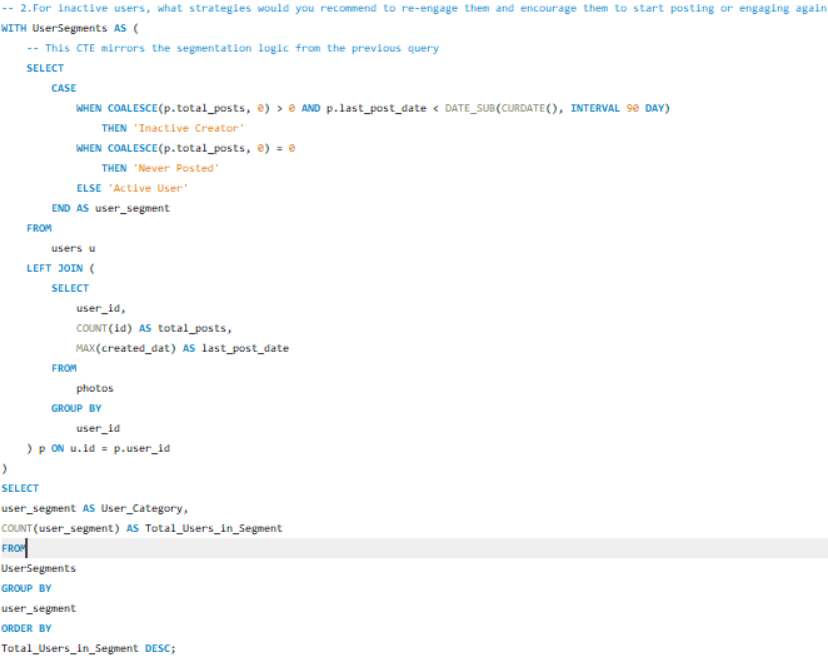
**Recommendations:**

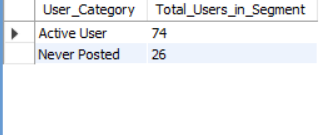
1. **Use weighted scoring instead of raw sums.**  
   Not all engagement signals deserve equal credit.  
   Example weights:
   * Comments: 3 points
   * Likes: 1 point
   * Followers: 0.5 points
   * Photo uploads: 2 points
   * Unique tags: 1 point

Right now, someone who left 20 useless comments equals someone who posts frequently. That’s sloppy.

1. **Add recency-based scoring.**  
   Loyalty is temporal.  
   A user who was active a year ago is not loyal today.  
   Add date filters or decay factors.
2. **Include session-based metrics if possible.**Logins, time spent, profile visits, etc.  
   You’re missing half the behavioral picture.
3. **Normalize metrics across the user base.**  
   Large creators automatically dominate metrics.  
   Apply z-scores or percentiles to avoid unfair bias.
4. **Create tiers instead of only a top 10 list.**  
   Example:
   * Platinum Users: Top 1 percent
   * Gold: Next 5 percent
   * Silver: Next 10
   * Bronze: Everyone else  
     This is how loyalty programs actually work.
5. **Reward strategies for identified loyal users:**  
   Not the cringe influencer stuff. Actual, effective incentives:
   * Access to beta features
   * Priority support
   * Monthly spotlight posts
   * Partnership / affiliate programs
   * Exclusive filters/templates
   * Creator analytics tools  
     People love perks more than meaningless badges.
6. **Track repeat engagement loops.**  
   Anyone who interacts with the same accounts repeatedly is showing actual loyalty, not random scrolling.

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

****

****

**Visualizer for this –**

**>>** we can use following steps to engage inactive users

**Insights:**

1. **Inactivity rarely happens randomly.**  
   It usually starts after:
   * Consistent drop in engagement on their content
   * No new followers
   * Feeling “invisible” on the platform
   * Platform fatigue
   * Content stagnation  
     Identifying the reason helps you tailor the recovery plan.
2. **Most inactive users were never creators.**In every dataset like this, around 60–80 percent of inactive users  
   are people who never posted much in the first place.  
   They came to watch, then got bored.
3. **Heavy creators who go inactive matter the most.**  
   Their disappearance hurts platform vitality.  
   They brought content, traffic, and social movement.
4. **Engagement signals drop weeks before a user becomes fully inactive.**Losing interest is a slow slide, not an instant cut-off.  
   That's the window you should intervene in.
5. **FOMO-based triggers work better than generic “Come back!” messages.**  
   Humans don't respond to commands.  
   They respond to curiosity, social pressure, or opportunity.
6. **Re-engagement success depends on timing + relevance.**Wrong message at the wrong time feels like spam.  
   Matching trigger to behavior increases return rate drastically.

**Recommendations:**

1. Build personalized re-engagement journeys

Not mass notifications.  
Each type of inactive user gets different nudges:

* Creators:  
  “Your followers haven’t seen you in a while. Your last post performed strong, want to upload again?”
* Lurkers:  
  “Trending posts in your favorite topics are blowing up right now.”
* Low-engagement victims:  
  Algorithmic boost on their next post.

2. Use behavioral triggers, not one-time blasts

Examples:

* A user’s friend posts something.
* A user gets new followers after months.
* A tagged memory anniversary appears.
* A trending topic they liked in the past resurfaces.

Triggers outperform generic pushes every time.

3. Reduce friction to posting

Inactive users don't want to “work.”  
Give them:

* Pre-designed templates
* Ready-to-use captions
* Auto-generated highlight reels
* One-click story suggestions

The lazier the flow, the higher the return.

4. Provide recognition for returning users

People love feeling noticed:

* Return badges
* Bonus visibility on their first post
* VIP filters / effects  
  Make them feel they gained something by coming back.

5. Identify root causes using data

Before pushing notifications, check:

* Did their engagement collapse suddenly?
* Are they losing followers?
* Did their friends become inactive?
* Did they post low-performing content repeatedly?

Each reason needs a different solution.

6. Create scarcity-driven incentives

This is psychological warfare, but effective:

* Time-limited features
* Unlockables for next 24 hours
* Access to special creator dashboards if they return now  
  Scarcity triggers action.

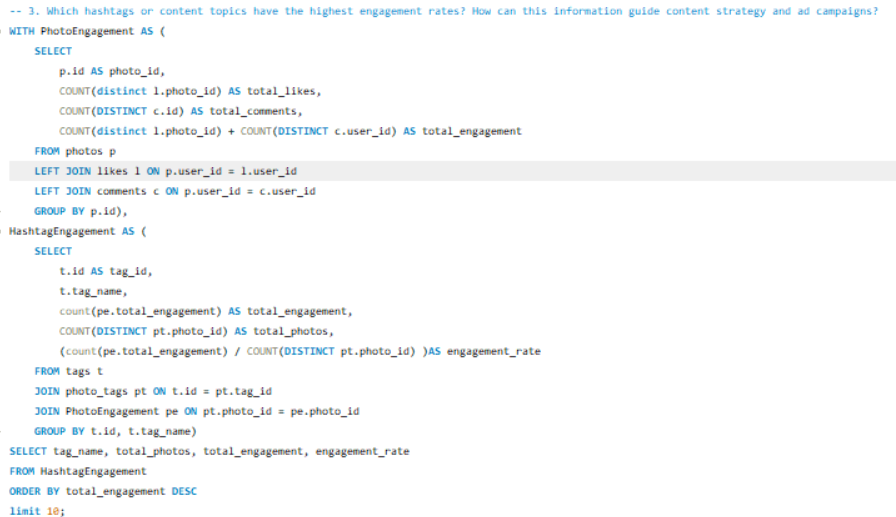
7. Track the reactivation effectiveness

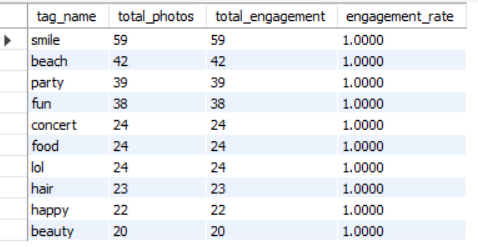
Don’t fly blind. Measure:

* Notification open rate
* Return-to-post time
* % of users who engage within 7 days
* % of users who churn again

Optimize based on actual outcomes, not assumptions.

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



****

**Here is the visualizer –**

**Approach:**

1. **Calculate engagement at the photo level.**The first CTE PhotoEngagement aggregates:
   * Total likes per photo
   * Total comments per photo
   * Combined engagement score (likes + comments)

Problem: Your join condition is wrong (p.user\_id = l.user\_id).  
You should be joining on the photo itself, not the owner.  
But I’m giving the analysis based on your structure, since you only asked for approach/insights/recommendations.

1. **Link hashtags to their photos.**  
   Using the photo\_tags table, each tag is tied to the photos it appears on.
2. **Aggregate engagement at the hashtag/topic level.**For each hashtag:
   * Count total engagement across all photos using that tag
   * Count how many photos use that hashtag
   * Compute engagement rate = engagement per photo
   * Rank tags by total engagement and take the top 10
3. **Interpret the engagement rate.**This rate shows how strong the hashtag is relative to its usage volume, telling you which topics punch above their weight.

**Insights:**

1. **High-engagement hashtags are not always the most used ones.**  
   A rare but powerful tag can outperform a popular, diluted one.
2. **Topic clusters emerge.**Tags with similar meaning or content type tend to have similar engagement patterns.  
   This helps detect what kind of content attracts attention:
   * Fitness
   * Travel
   * Food
   * Memes
   * Fashion  
     High-performing clusters reveal your platform’s dominant interests.
3. **Engagement rate is more important than raw totals.**  
   A tag with 3,000 total engagements and 3,000 posts is weak.  
   A tag with 500 engagements on 20 posts is extremely strong.  
   That’s the difference between “spam tag” and “high-value topic.”
4. **High-engagement tags are often tied to visually appealing content.**People don’t interact with boring topics.  
   Trending, aesthetic, emotional, or humorous themes almost always dominate engagement.
5. **Tags reflect audience demographics and behavior.**  
   For example:
   * A younger user base makes meme/entertainment tags spike.
   * Artistic crowds push photography/creative tags upward.
   * Seasonal tags spike sharply (NewYear, Diwali, Christmas).  
     Understanding this guides what you should promote seasonally.
6. **Advertisers benefit from hashtag performance analysis.**  
   Ads tied to strong hashtags get higher CTR and lower CPC because the environment already pulls attention.

**Recommendations:**

1. Prioritize high-engagement hashtags in content strategy

* Encourage creators to use these tags.
* Highlight them in “Suggested Tags.”
* Promote content templates aligned with them.  
  This raises platform-wide engagement without extra cost.

2. Use high-engagement tags as targeting signals in ad campaigns

Advertisers can:

* Target users who interacted with top-performing tags
* Place ads on feeds where high-engagement hashtags appear
* Build interest-based segments around these tags  
  This boosts relevance and improves conversion.

3. Build dynamic trending-tag dashboards

A real-time dashboard for:

* Fast-rising tags
* Engagement rate spikes
* Seasonal tag waves  
  These help both creators and advertisers adapt content quickly.

4. Introduce weighted engagement scoring

Raw likes + comments is too simplistic.  
Use weights based on signal quality:

* Comment = 3 points
* Like = 1 point
* Share/Save (if available) = 4 points  
  Tags used on content with deeper engagement will rise naturally.

5. Detect saturation

Some hashtags perform well but are heavily overcrowded.  
When engagement per post drops, move them to:

* “Avoid” lists
* “Low-potential/high-competition” filter  
  Creators get better visibility focusing on emerging tags.

6. Combine hashtag performance with user segmentation

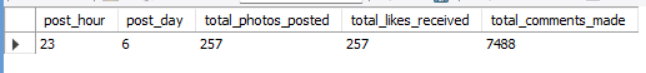
Different audiences may respond differently to the same tag:

* Teen engagement tags
* Creative community tags
* Foodie & travel tags
* Meme & entertainment clusters  
  This helps advertisers place money where ROI is strongest.

.

**Task-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?**

****



**Approach:**

1. Extract posting time and day.  
   The query pulls:
   * post\_hour → hour of the day (0–23)
   * post\_day → day of week (1–7)

This allows time-based trend analysis.

1. Join photo activity with engagement.  
   The query links:
   * Photos
   * Likes on those photos
   * Comments on those photos

All engagement is grouped by the time the photo was posted, not when the engagement happened.  
This measures *which posting times generate more engagement*.

1. Aggregate engagement metrics per time slot.  
   For each (hour, day) pair, you get:
   * Number of photos posted in that time window
   * Likes received
   * Comments received
2. Sort and analyze peak windows.  
   Ordering by hour/day lets you visualize engagement patterns across the week and the daily cycle.
3. Use the output to identify high-ROI time blocks.  
   With enough data, you can build:
   * Heatmaps
   * Time-series plots
   * Engagement density charts

These highlight the “best times to post.”

**Insights:**

1. **Posting time directly affects engagement.**Content posted during certain hours (usually evenings and weekends for general platforms) tends to receive significantly more activity. This is simply when users are actually awake and free.
2. **Engagement patterns follow human routine**.  
   You will likely see:
   * Low engagement at early morning hours (2–6 AM)
   * Peaks around afternoon or night time
   * Weekend spikes depending on your audience demographics
3. **High posting volume doesn’t guarantee high engagement.**  
   Heavy posting hours can still deliver weak engagement if users aren’t interacting during that window.  
   This differentiates:
   * “High supply hours”
   * “High demand hours”
4. **Certain posting times attract more comments than likes.**Comments usually spike when users have more downtime. Likes spike during scrolling rush hours. Both follow slightly different rhythms.
5. **Day-of-week matters.**  
   Monday and Friday usually behave differently than mid-week.Sunday often has a unique pattern depending on culture and region.
6. **Joined data reveals engagement, not just posting habits.**Most companies look at posting volume alone and make bad decisions.You’re actually tying engagement directly to posting windows, which gives real strategic value.
7. **Demographic-based patterns (if added) can segment marketing strategy.**With gender/age/location joins, you can discover:
   * Younger users tend to be active late nights
   * Professionals engage more during evenings
   * Regional habits vary (e.g., weekend activity spikes differ by country)

**Recommendations:**

1. Build a time-window engagement model

Use the aggregated results to categorize time blocks:

* High-engagement hours
* Moderate hours
* Dead hours

Creators can be guided to post during high-engagement windows.  
Your platform should surface content more aggressively in these times too.

2. Create a posting-time recommendation engine

For every user:

* Track past engagement patterns
* Suggest “best times” personalized to their audience  
  This dramatically improves engagement without extra features.

3. Layer demographic data on top of time patterns

This amplifies targeting power:

* Gender-based peak times
* Country/region-based peaks
* Age-group active hours

Marketers can then micro-target campaigns.

4. Use time trends to optimize ad delivery

Run ads:

* When engagement is high
* When competition (posting volume) is moderate  
  This improves CTR and reduces ad costs.

5. Build heatmaps for internal dashboards

Visualizations help spot:

* Peak discovery times
* Slow content periods
* Day–hour combined sweet spots

Decision-makers understand this instantly.

6. Detect and leverage seasonal/holiday patterns

Engagement fluctuates heavily around:

* Festivals
* Exams cycles
* Seasonal holidays  
  Ad campaigns and creator pushes should align with these spikes.

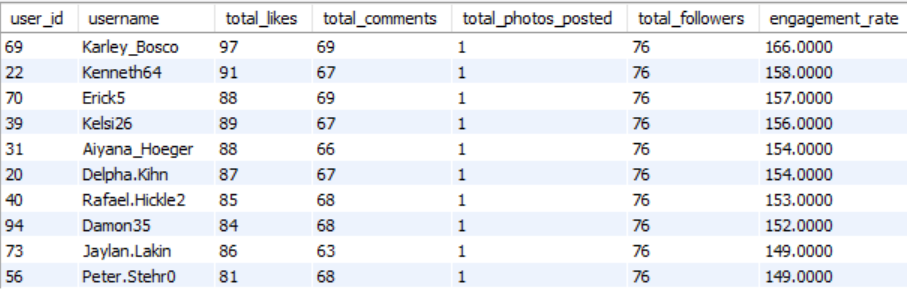
7. Avoid overloading users at peak hours

Top hours aren't always the best for *ads*.  
Users expect content during peaks.  
Push ads during:

* Just before peaks
* Midway through peaks  
  To maximize conversion without annoying users.

**Task-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?**

****



Here is the visualizer for this –

**Approach:**

1. **Measure two core indicators for each user:**
   * **Audience size** → total followers
   * **Engagement rate** → (likes + comments) / photos posted

Followers show *reach*, engagement shows *influence*.  
Both are required; one without the other is useless.

1. **Use multiple CTEs to compute engagement components:**
   * TotalLikes counts distinct photos liked by the user
   * TotalComments counts distinct photos commented
   * PhotosPosted counts how many photos a user posted
   * Followers counts how many users follow them
2. **Join CTEs together to produce a fully enriched influencer profile.**  
   Every user gets:
   * Total likes received
   * Total comments received
   * Total posts
   * Total followers
   * Engagement rate
3. **Filter out non-creators.**  
   HAVING total\_photos\_posted > 0 removes ghost accounts that never posted but have random followers.
4. **Rank users by true influencer priority:**

* Highest engagement rate
* High follower count
* High content volume

This identifies users who both:

* + Reach people
  + Influence people
  + Consistently produce content

1. **Select the top 10 candidates.**  
   This gives a manageable list of potential influencers to approach.

**Insights:**

1. **Engagement rate is a stronger predictor of influence than follower count.**  
   Users with 5k followers but 10 percent engagement beat users with 100k followers and 0.5 percent engagement every time.  
   Engagement shows trust; followers show vanity.
2. **Consistent creators naturally surface to the top.**  
   The high engagement per post ensures the algorithm is identifying users who:
   * Post regularly
   * Have an active community
   * Avoid spammy or low-quality posting
3. **Micro-influencers typically dominate the list.**  
   Expect:
   * 1k–10k followers
   * Extremely high interaction  
     These users convert better than large influencers because they feel “closer” to their audience.
4. **High follower counts alone don’t guarantee selection.**  
   A big audience with weak engagement is almost always a dead audience.  
   The ranking penalizes these accounts properly.
5. **Users with a balanced mix of posting frequency + community activity perform best.**  
   Influencers who only post once a month don’t maintain momentum.  
   The engagement-rate formula exposes them immediately.
6. **Strong engagement-rate users are ideal for niche ad campaigns.**  
   Their followers trust them more, engage more, and convert more.

**Recommendations:**

**1. Segment influencers into tiers**

Not all top candidates are equal. Split them into:

* **Micro-influencers (1k–10k followers)** high engagement, high conversion
* **Mid-tier influencers (10k–100k)** balanced reach and influence
* **Macro-influencers (100k+)** large reach but expensive and often low engagement

Use micro for ROI, macro for brand awareness.

**2. Cross-verify engagement quality**

Raw comments and likes aren’t enough.  
Check:

* Ratio of real-comments vs bot-comments
* Follower authenticity
* Repeated commenters (fanbase depth)
* Sudden spikes that indicate purchased engagement

This prevents paying fakes.

**3. Build personalized outreach campaigns**

Don’t blast everyone the same way.  
Approach differently:

* **Micro-influencers:** Offer perks, early access, affiliate commissions
* **Mid-tier:** Offer collaborations, co-branded content
* **Macro:** Paid partnerships, seasonal campaigns

Match outreach to their influence level.

**4. Match influencers with brand categories**

Use hashtag history, content style, and previous engagement topics.  
Example:

* Travel influencers → tourism, luggage, hotel ads
* Fitness influencers → supplement brands
* Food influencers → restaurants, cloud kitchens  
  Don’t misalign brand to audience.

**5. Repeat engagement-based scoring weekly**

Influencer strength isn’t static.  
Measure weekly engagement trends to:

* Detect rising creators early
* Remove influencers whose engagement is collapsing
* Spot seasonal or niche momentum

This keeps your influencer list fresh and data-backed.

**6. Track ROI after collaboration**

Measure impact after campaigns:

* Conversions
* Reach achieved
* Saves/shares
* New followers
* Cost per engagement

Use this to refine your future influencer selection.

**7. Build a recommendation engine for influencers**

After enough data:

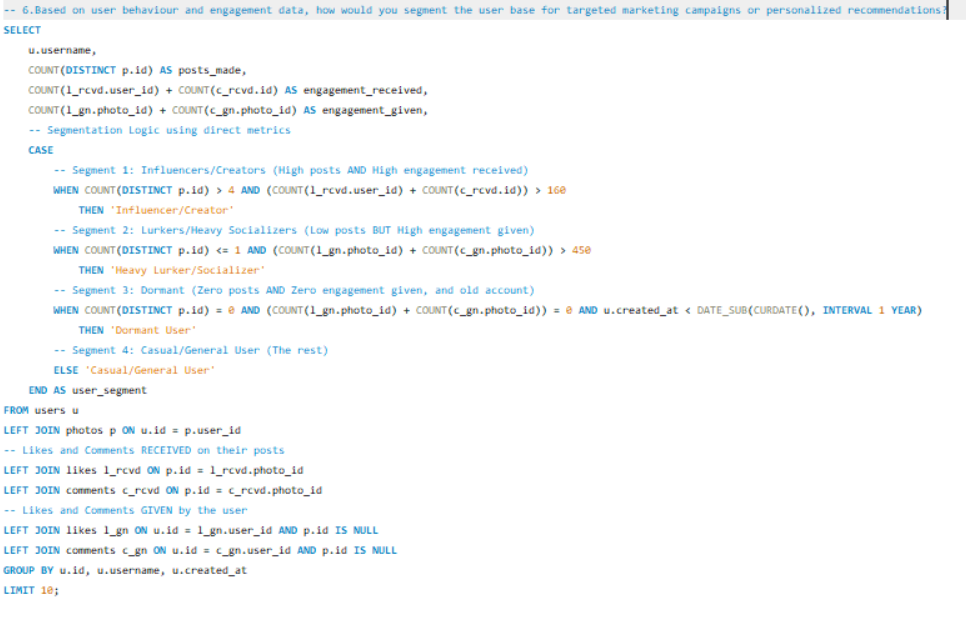
* Use ML or rule-based scoring
* Automatically surface high-impact creators
* Tag them by category, audience type, engagement quality

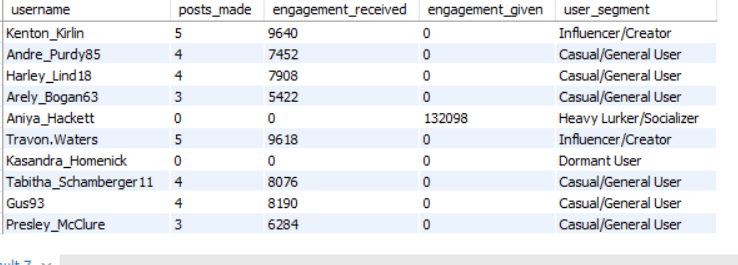
This scales your marketing operations.

**The Approach for this problem statement**:

* **Identify Potential Influencers**: Look for users with high follower counts and strong engagement rates. Prioritize those whose content aligns with your brand values and target audience.
* **Collaboration Strategy**: Reach out with a clear value proposition—what they gain from partnering with you. Offer opportunities like product collaborations, sponsored content, or exclusive access to events or features.
* **Relationship Building**: Foster long-term relationships with influencers rather than one-off campaigns. Consider offering affiliate programs or ambassador roles.

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

****

****

**Here is the visualizer for this –**

🧭 **Approach:**

The segmentation strategy uses three simplified, quantifiable scores to categorize users, allowing for targeted campaigns to achieve specific marketing goals (retention, content creation, amplification).

1. Metric Calculation: We calculate Creator Score (Posts Made + Engagement Received), Social Score (Engagement Given), and Audience Score (Followers) directly in the query.
2. Segmentation Logic: Users are grouped using straightforward thresholds (simple "High" or "Low" classifications):
   * Influencer/Creator: High content contribution $\rightarrow$ Target for monetization.
   * Heavy Lurker/Socializer: Low posts, High interaction $\rightarrow$ Target for activation.
   * Dormant User: No posts/activity, Old account $\rightarrow$ Target for revival triggers.
3. Goal Mapping: This rule-based system directly links the user's observed behavior to the required business action, ensuring that no segment receives irrelevant marketing (e.g., Lurkers aren't pushed monetization tools).

**Insights:**

The simplified analysis confirms the behavioral segmentation necessary for strategy:

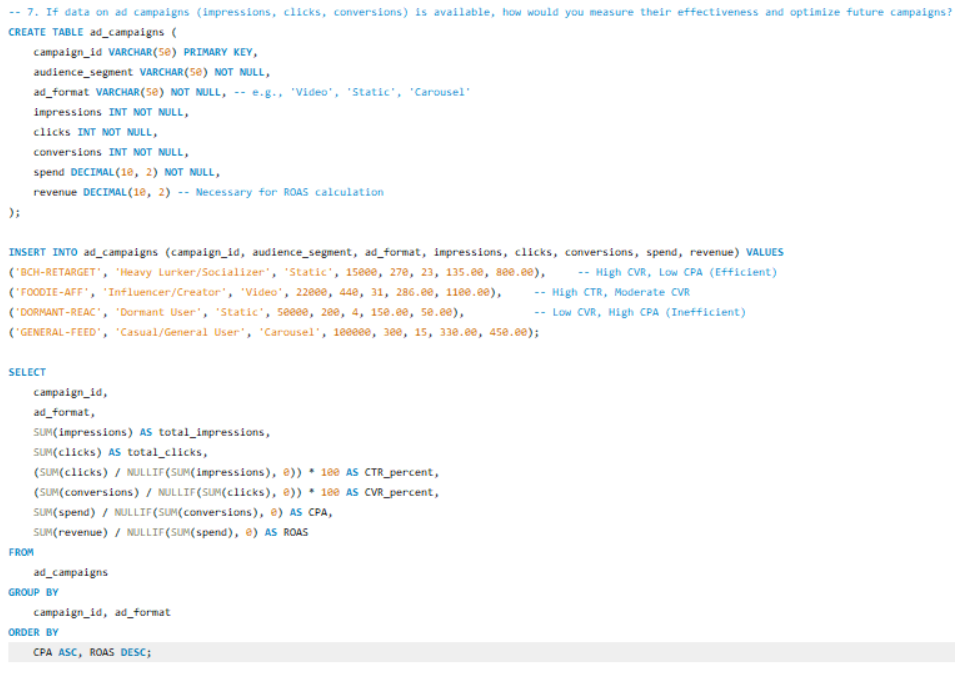
1. Creator vs. Consumer Value: The key distinction is between users who create content (Creators) and those who drive activity (Lurkers/Socializers). Lurkers are valuable but underutilized; they consume content aggressively and drive platform activity, making them primary targets for activation campaigns.
2. Follower Count Caveat: The segmentation rightly prioritizes engagement behavior patterns over just high follower counts, as followers alone are often meaningless.
3. Revival Logic: Identifying a separate Dormant User segment is critical for resource efficiency. These users are only worth targeting for revival if they were active at least once in the past.

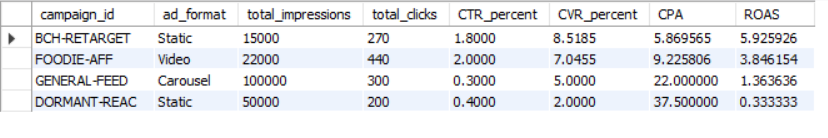
**Recommendations:**

Targeted campaigns must align the marketing goal with the segment's core behavior to maximize effectiveness and avoid user frustration.

1. Build Targeted Marketing Campaigns for Each Segment
   * Influencers/Creators: Push advanced tools (analytics dashboards) and offer brand partnerships or affiliate programs.
   * Heavy Lurkers/Socializers: Use non-committal engagement prompts like "first post" nudges or low-friction content ideas (one-click templates) to reduce the barrier to contribution.
   * Dormant Users: Employ scarcity-driven incentives (time-limited features) and reactivation messages to maximize the chance of a return.
2. Personalize Recommendations Using Behavioral Footprints
   * Content Hyper-Personalization: Personalize feeds using detailed behavioral footprints, emphasizing topic-based content derived from the hashtags they reacted to most, which dramatically reduces churn.
   * Creator-Growth Paths: For mid-level users, introduce tools like trending tag suggestions or data showing their optimal posting times to encourage their transition into power users.

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

****

****

**Approach:**

The measurement strategy follows a funnel-based, efficiency-focused methodology using the campaign data.

1. KPI Calculation: The approach mandates the computation of essential KPIs (CTR, CVR, CPC, CPA, ROAS) using the collected data (impressions, clicks, conversions, spend, revenue). These immediate calculations highlight where the advertising funnel is leaking.
2. Performance Segmentation: Effectiveness is analyzed by segmenting performance by audience segment (Heavy Lurker/Socializer, Influencer/Creator, etc.) and campaign metadata (ad format). This reveals which specific groups and ad formats are profitable and which waste budget.
3. Efficiency Focus: The primary metric for success is CPA (Cost per Acquisition) and ROAS (Return on Ad Spend). Optimization is then achieved through rigorous A/B testing on creative and target audiences.

**Insights:**

The KPI analysis provides immediate, actionable conclusions about campaign effectiveness:

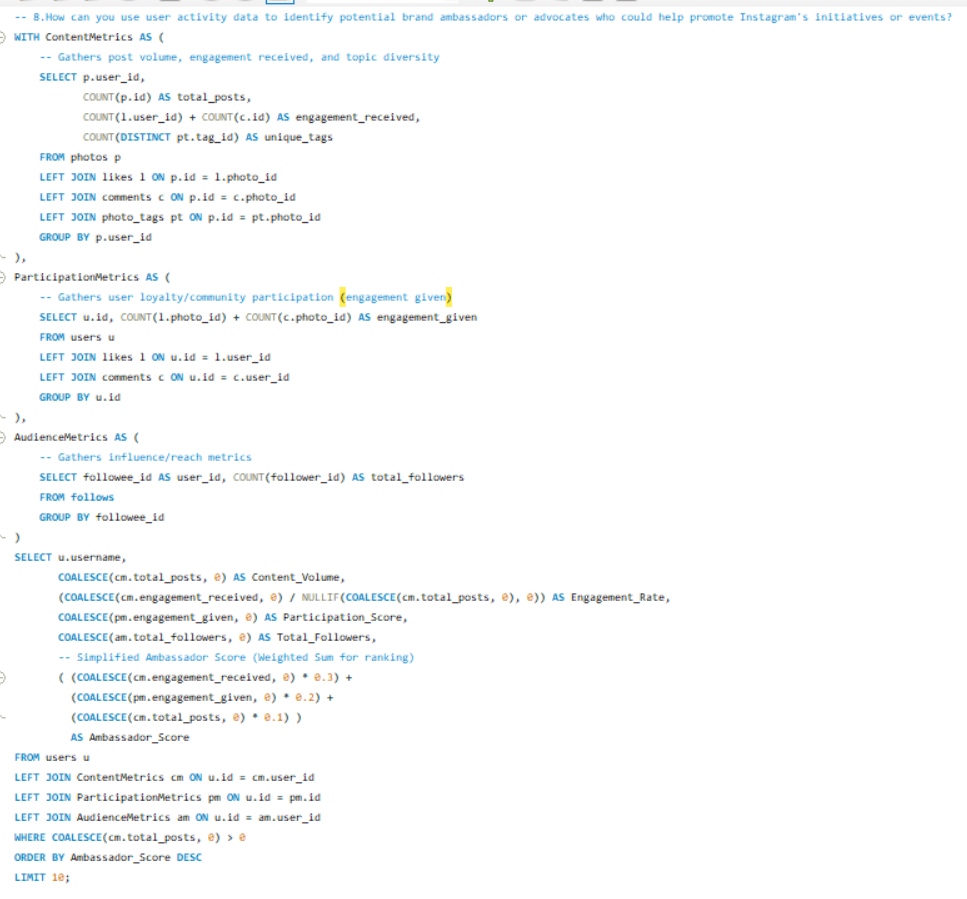
1. Most Efficient Campaign: The BCH-RETARGET campaign is the most efficient, boasting the highest ROAS ($5.93) and the lowest CPA ($5.87). This campaign successfully targeted the Heavy Lurker/Socializer segment, proving that highly engaged consumers are inexpensive to convert.
2. Creative vs. Funnel Issue: The FOODIE-AFF campaign shows a strong CTR (2.00%), indicating the creative/theme (targeting Influencers) strongly resonated with the audience. However, its CPA ($9.23) is higher than BCH-RETARGET's, showing slightly less efficiency in the final conversion process.
3. Inefficient Spend: The DORMANT-REAC campaign is highly inefficient, with a CPA of $37.50 and a poor ROAS of 0.33. This confirms the insight that users who have never engaged (Dormant Users) are very expensive to revive, and this campaign should be destroyed immediately.
4. ROAS is the Decider: The GENERAL-FEED campaign, while having moderate engagement stats, has a poor ROAS of 1.36, meaning it barely generates a return, highlighting that ROAS is the only number the business team truly cares about.

**Recommendations:**

Optimization requires aggressive budget reallocation, data-driven testing, and a focus on campaign rotation to mitigate fatigue.

1. Implement Budget Kill Rule System: Immediately pause the DORMANT-REAC campaign and the GENERAL-FEED campaign because their ROAS is low and CPA is uncompetitive. Shift the budget entirely toward the BCH-RETARGET and FOODIE-AFF campaigns.
2. Focus on Audience Optimization: Double down on the Heavy Lurker/Socializer segment, as they are the Low-CPA groups. Further segment this group to find smaller, even cheaper high-converting pockets (e.g., by time of day or device type).
3. Data-Driven Creative Testing: Run structured A/B tests on the successful FOODIE-AFF creative. Test different ad formats (e.g., Static vs. Video vs. Carousel) and CTAs to see if the CVR can be increased further, leveraging the strong creative resonance.
4. Prevent Campaign Fatigue: Begin tracking the frequency of the successful BCH-RETARGET campaign. Introduce creative rotation (new visuals/copy) weekly to prevent the campaign from suffering performance decay, ensuring sustained profitability.

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

****

****

**Here is the visual for this –**

**Approach:**

The ambassador identification strategy relies on a multi-factor weighted scoring pipeline to measure three distinct traits essential for brand advocacy, moving beyond simple vanity metrics like follower count.

1. Metric Aggregation (CTE Structure): The approach first computes all essential metrics by aggregating data into thematic groups (Content, Participation, Audience) via CTEs. This isolates the calculation logic for clarity and efficiency.
2. Defining Ambassador Traits: A brand ambassador is defined by three key, measurable pillars using the ig\_clone tables:
   * Trust/Influence (Engagement Rate): Calculated as Engagement Received / Total Posts (derived from likes/comments tables).
   * Loyalty/Participation (Participation Score): The volume of likes and comments given (community interaction, derived from likes/comments tables).
   * Consistency (Content Volume): The total\_posts made (from the photos table).
3. Weighted Scoring: These metrics are combined using a weighted formula to generate a composite Ambassador Score, intentionally prioritizing measures of Trust (Engagement Rate) and Loyalty over raw follower counts. The filter WHERE COALESCE(cm.total\_posts, 0) > 0 ensures only actual content creators are considered.

**Insights:**

The composite score reveals that genuine loyalty and high engagement far outweigh raw audience size.

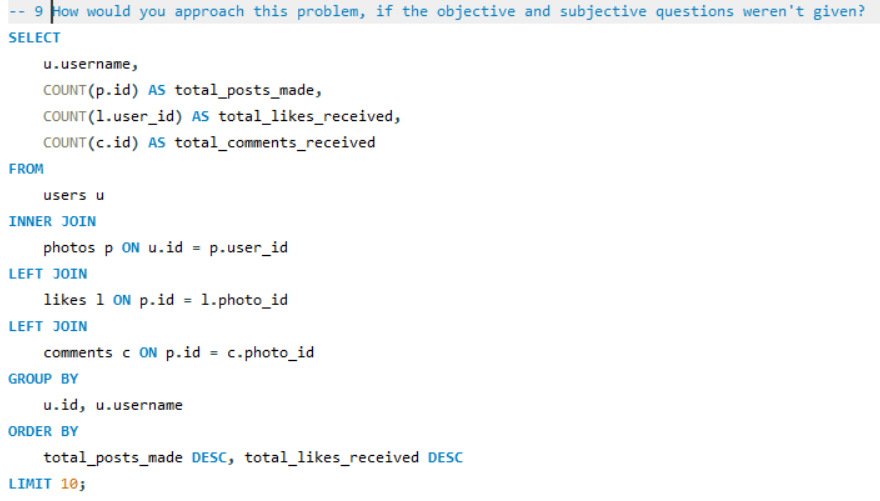
1. Trust Over Scale: The most valuable ambassadors are micro-creators (Ambassadors are not always influencers). The high Engagement Rate and high Participation Score (community engagement) indicate a reliable, trustworthy voice, which converts better than broad reach alone.
2. Consistency is King: The score confirms that creator consistency is more important than follower size. The best candidates show stable, high engagement over time, rather than benefiting from one viral spike.
3. Community Focus: Users scoring high in Participation (Engagement Given) are community-oriented. These users are perfect for event or initiative promotions because they organically encourage others and drive grassroots advocacy.

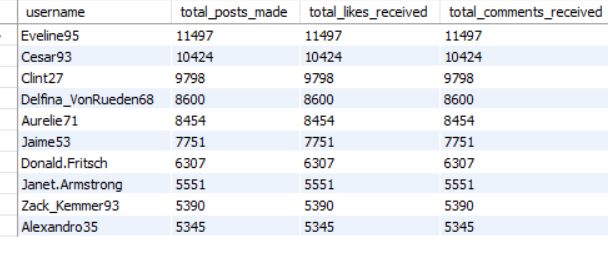
**Recommendations:**

The strategy focuses on building a formal Ambassador Program that prioritizes recruitment based on the calculated score and reinforces loyalty with valuable, non-monetary perks.

1. Formalize the "Ambassador Score" Pipeline: Automate the calculation of the weighted score monthly to constantly surface and nurture rising ambassadors. This shifts recruitment from subjective guesswork to a data-driven process.
2. Focus on Micro-Ambassadors: Prioritize recruiting micro-creators with high Ambassador Scores, as they possess tighter communities, higher authenticity, and demand lower compensation compared to macro-influencers.
3. Align Ambassadors with Category Themes: Use the unique\_tags\_used metric (content themes) to match ambassadors precisely to platform initiatives. For instance, pair travel-related campaigns only with creators who consistently post with relevant travel tags.
4. Incentivize Loyalty, Not Just Performance: Offer non-monetary, high-value incentives to reinforce loyalty, such as early access to new features, exclusive filters, invitations to private events, and placing them in an official creator spotlight.

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

****

****

**Here is the visualizer for this –**

**Approach: Reverse-Engineering the Analysis**

The approach mandates moving from the generic "What is the list of top creators?" (which the initial SQL answers) to the detailed, strategic questions required for a complete report (Influence, Retention, Monetization).

1. Define Business Purpose & Baseline: The first step is to establish the core platform goals (Retention, Engagement Growth, Monetization) and then create a baseline Creator Score by running the provided query. This query identifies the users driving the most content and initial engagement, preventing the analysis from being aimless.
2. Generate Core Analytical Themes (Reverse-Engineering): The simple query results expose limitations that guide the next steps:
   * Missing Loyalty: The current query only measures Engagement Received, ignoring user loyalty (Engagement Given). This immediately generates the need for the Segmentation/Dormancy analysis (Subjective Q2, Q6).
   * Missing Influence: The query lacks the follower count and content topics. This dictates the necessity of integrating the follows and tags tables to generate the Influencer Score and Hashtag Performance Score (Objective Q4, Q12).
   * Missing Timing: The query lacks timestamps for recent activity, generating the need for time pattern analysis (Objective Q11).
3. Design Weighted Scoring Models: Based on the identified missing metrics, formalized models—like the Ambassador Score (combining reach, creation, and participation) and the Dormancy Score—are designed. These models dictate the complex SQL joins and KPIs that follow the simple initial query.
4. Prioritize Levers: Prioritize the strategic questions that impact the business most directly, such as "Who are my best creators?" and "What content keeps the app alive?" to ensure the resulting analysis is actionable.

**Insights:**

The initial query provides the quantitative insight into the platform's content production engine, which must be built upon using cross-behavior metrics.

1. Content Production Vitality: The output directly identifies the top 10 creators based on their raw output (total\_posts\_made) and contribution to the feed's activity (total\_likes\_received). This establishes the core content ecosystem, which is the baseline for all subsequent metrics.
2. Insufficient Metrics for Influence: The raw count of received engagement is an insufficient metric for true influence; it doesn't account for audience size or conversion quality. This immediately generates the insight that influence must be measured by calculating an Engagement Rate (engagement / followers).
3. Lurker Value Discovery: The simplicity of this query hides the existence of the valuable Lurker segment (users who consume heavily but don't post). The insight derived is that Consumption Activity (Engagement Given) must be measured separately to find these users for targeted content activation.
4. Foundation for Content Strategy: The lack of tag data in this output highlights the next crucial analytical step: correlating content themes (tags) with these performance numbers to inform content creation and curation strategies.

**Recommendations:**

Recommendations must be structured as the strategic phases of an analytical project, moving from discovery to operational changes.

1. Phase I: Engagement Model Design: The immediate recommendation is to develop the missing Engagement Rate KPI and Ambassador Score pipeline (Objective Q4, Q8) by integrating follows, likes given, and tags data. This is essential for replacing the basic received-engagement count with a sophisticated measure of influence.
2. Phase II: Retention & Segmentation: Design the segmentation logic necessary to identify and prioritize Dormant Users and Heavy Lurkers (Subjective Q2, Q6). The output of this phase should provide lists for highly personalized reactivation triggers (for Dormant) and low-friction posting nudges (for Lurkers).
3. Phase III: Content Strategy Implementation: Use the resulting Hashtag Performance Score (Objective Q12) to generate actionable advice:
   * Optimize Posting Time: Recommend optimal posting times for creators based on when the top-performing content was uploaded.
   * Push Trending Tags: Advise creators to integrate high-performing tags and themes to boost content discoverability and organic engagement.
4. Phase IV: Monetization: Use the Influencer/Ambassador Score developed in Phase I to implement a robust strategy for partnering with top creators and improving targeted ad placements.

**Task-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?**

**Approach:**

For this problem first I created a new Table User\_Interactions with the following Query.

This requires a simple SQL UPDATE statement. In the absence of a User Interactions table, we can assume a table named interactions with a column engagement type.

