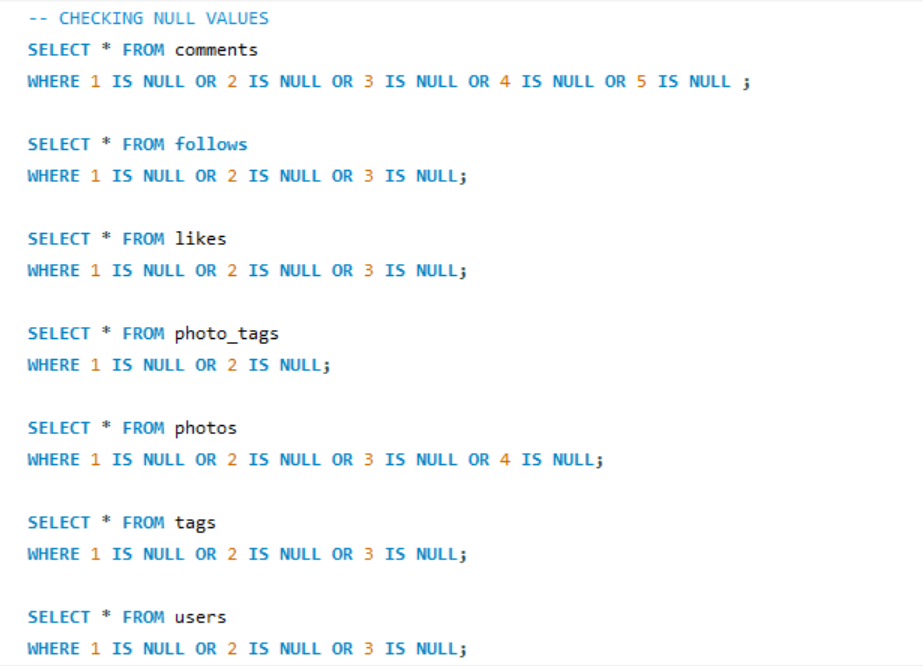
**SOCIAL MEDIA ANALYSIS**

Objective Questions

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

To check whether there are any NULL Values, I used below SQL queries .



RESULT: - NO NULL VALUES

1. DUPLICATE VALUES

To check whether there are any DUPLICATE Values, I used below SQL queries.



RESULT: - NO DUPLICATE VALUES

1. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?
2. **Photos\_posted**: This column represents the no. of posts/photo\_posted by each user

The number ofphotos posted shows how frequently a user contributes content, reflecting their overall activity level.

1. **Tags\_used:** This column represents the distinct count of tags used by each user

The distinct count of tags indicates how actively a user is engaging with different topics or communities, suggesting a broader involvement in the platform.

1. **Comments\_made:** This column represents the comments made by each user

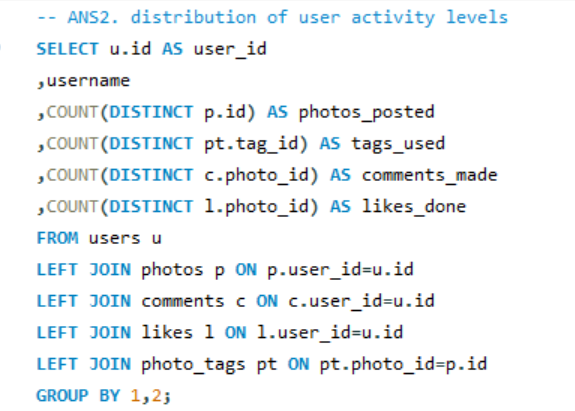
Comments demonstrate not just passive consumption but active participation in discussions, showing a user's willingness to engage with others.

1. **Likes\_done:** This column represents the no. of photos liked by each user

The number of likes given reflects a user's engagement with others' content, indicating how much they interact with the community.

**OVERALL ACTIVITY LEVEL:**

Together, these parameters provide a well-rounded view of user activity. They capture both content creation and interaction, helping to assess how involved a user is on the platform. A user who actively posts, tags, comments, and likes are likely to have a higher activity level, indicating a vibrant engagement with the community.



A screenshot of a computer

Description automatically generated

**VISULAIZATION:**

1. **COUNT OF PHOTOS POSTED BY USER**

**OBSERVATION:**

* **High Inactivity**: 26% of users haven’t posted any photos, highlighting a large inactive segment that presents a significant engagement opportunity.
* **Low Posting Activity**: 44% of users have posted between 1 to 3 photos, indicating low to moderate activity, where targeted campaigns could boost participation.
* **Moderate Posting Activity**: 40% of users have posted 4-5 photos, showing they are moderately engaged and could be motivated to post more with the right incentives.
* **High Posting Activity**: Only 7% of users have posted more than 5 photos, representing super-engagers who could be rewarded to maintain or increase their activity.

1. **USER COUNT BY COMMENTS MADE**

**OBSERVATION:**

* A significant number of users (23) have not made any comments, indicating a potential area for improvement in user engagement.
* The range of 51-60 comments has a count of 18, showing that while there is some activity, it is still below what might be expected for a highly engaged user base.
* Many users (36) fall within the 61-70 comments range, suggesting that this group is moderately engaged and could be targeted for initiatives to increase their activity levels.
* There are few users (1) in the higher ranges (71-80 and 81-90), indicating that very few users are actively participating in discussions or engaging with content at a high level.
* The category of 251+ comments have 13 users, indicating a small but dedicated group of super-engagers who actively participate in discussions and such HIGH ACTIVITY makes it suspicious to have called them as **‘BOT USERS’**

1. **USERS COUNT BY LIKES DONE**

**OBSERVATION:**

* A notable number of users (23) have not given any likes, suggesting a significant segment of the user base is disengaged and may need targeted strategies to encourage interaction with content.
* The range of 81-90 likes has the highest count (35 users), indicating a moderately engaged group that actively interacts with posts. This segment presents an opportunity for initiatives aimed at further increasing their participation.
* The 71-80 likes category has 14 users, showing some level of interaction, while the 91-100 range includes 13 users, indicating a drop in engagement as users move into higher liking frequencies.
* The category of 251-260 likes includes 13 users, indicating a dedicated group of super-engagers who frequently interact with content. These users can be pivotal in community-building efforts and may be influential in encouraging others to engage.

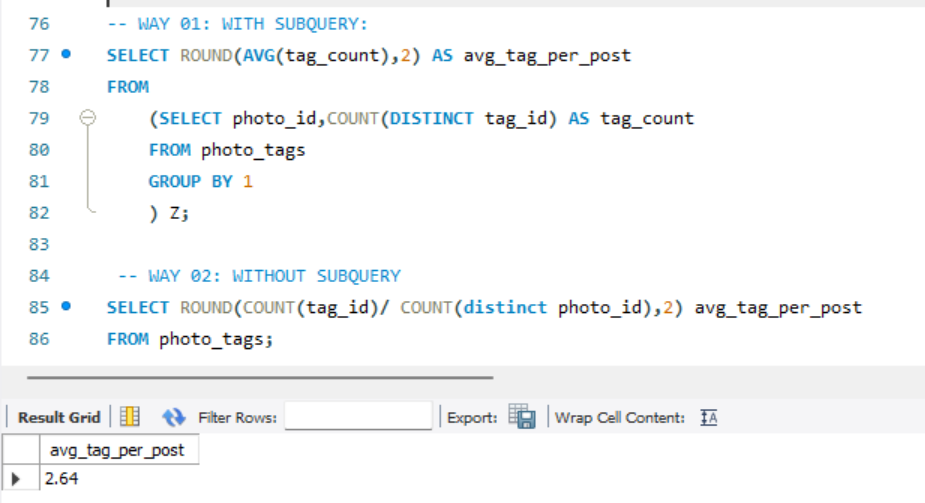
1. **USERS COUNT BY TAGS USED**

**OBSERVATION:**

* A significant portion of users (31) have not used any tags, indicating that a large segment of the user base is either unaware of the tagging feature or does not see its value in enhancing their posts' reach.
* Many users (6-9 users) fall within the range of using 1 to 7 tags. This suggests that most users are experimenting with tags but are not using them extensively, presenting an opportunity to educate users on the benefits of using more tags for discoverability.
* Users using 3 to 9 tags are consistent in number, with 4-9 users in each range, suggesting a small but consistent pattern of tag usage across this group.
* There are very few users (2-3) in the higher tag ranges (10-15), indicating that only a small number of users are actively utilizing many tags per post.

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

Used a subquery to first calculate the tag count per photo, then averages those counts.

****

RESULT

Avg\_tag\_per\_post = 2.64

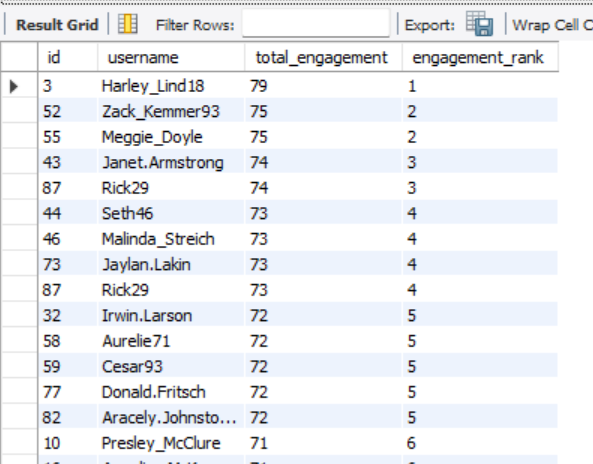
1. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.

**A screenshot of a computer program

Description automatically generated**

**RESULT:**

The final output identifies the top users with the highest engagement rates on their posts, along with their ranks, facilitating easy identification of the most engaged users.



**OBSERVATION:**

* 1. The top five users have total engagement scores ranging from 74 to 79, indicating a competitive environment at the top.
  2. Users such as Zack\_Kemmer93 and Meggie\_Doyle have the same total engagement score, resulting in tied rankings (2). This indicates that multiple users are successfully engaging their audiences to a similar extent.

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

**HIGHEST NO. OF FOLLOWERS**

* **Subquery for Follower Count:** It counts the number of followers for each user by joining the users table with the follows table on followee\_id, grouping by user ID and username, and calculates a dense rank based on the follower count.
* **Ranking:** It uses the DENSE\_RANK() function to assign ranks to users based on their follower counts in descending order, allowing for ties in follower numbers.
* **Final Selection:** The outer query filters to select only users with the highest follower count (ranked 1) and orders the results by user ID, providing a list of top users based on their followers.

A screen shot of a computer code

Description automatically generated

**RESULT: TOTAL 23 USERS**

A screenshot of a computer

Description automatically generated

**HIGHEST NO. OF FOLLOWINGS**

* **Subquery for Followings Count**: It counts the number of users each user is following by joining the users table with the follows table on follower\_id, grouping by user ID and username.
* **Ranking:** It applies the DENSE\_RANK() function to assign ranks based on the number of followings in descending order, allowing for ties in the followings count.
* **Final Selection:** The outer query filters to select only users with the highest followings count (ranked 1) and orders the results by user ID, providing a list of top users based on their followings.

A computer screen shot of a code

Description automatically generated

**RESULT: TOTAL 77 USERS**

**A screenshot of a computer

Description automatically generated**

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

A screenshot of a computer

Description automatically generated

**RESULT:**

A screenshot of a computer

Description automatically generated

**OBSERVATION:**

* **High Engagement Leaders**: The top user, Meggie\_Doyle, has a significantly higher average engagement rate (75) compared to many users, indicating a strong ability to connect with their audience.
* **Narrow Band of Engagement Rates**: The top 10 users have average engagement rates clustered closely between 75 and 66, suggesting that competitive differentiation in engagement may be minimal among this group.
* **Engagement Rate Decline**: There is a noticeable drop in average engagement rates toward the lower end of the spectrum, with users like Tomas.Beatty93 (53) exhibiting rates well below the top performers, highlighting a potential gap in audience engagement across the user base.

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

* Query retrieved the IDs and usernames of users from the users table.
* It counted the number of distinct posts (photo IDs) they liked, using a LEFT JOIN to ensure all users were included, even those with no likes.
* Then, filters for users with zero interactions in terms of likes, highlighting a segment of the user base that is inactive in this key engagement area

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**OBSERVATION:**

* **Inactive Users**: The result shows a list of 20 users who have never engaged with any posts through likes, indicating a lack of interaction with the platform.
* **Potential Engagement Issues**: This group may reflect potential engagement challenges within the user base, as they are not participating in a key interactive feature.

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

* **Tag Extraction**: The first CTE tag\_name retrieves user IDs along with the corresponding tag names and photo IDs for each liked photo. This is done by joining the

likes, photo\_tags, and tags tables.A computer code with text

Description automatically generated with medium confidence

* **Tag Categorization**: The second CTE tag\_category assigns categories to each tag name (e.g., 'Joy-Emotions', 'Aesthetics') using a CASE statement, facilitating easier analysis of user preferences.



* **Likes Count and Ranking**: The third CTE likes\_per\_category counts how many likes each user has given within each tag category and ranks these counts using DENSE\_RANK() to identify the top categories for each user.

****

* **Final Selection**: The main query selects users, their top tag categories, and the count of likes, filtering for those who fall within the top three ranked categories per user.

**A black text on a white background

Description automatically generated**

**RESULT: TOTAL 77 USERS WITH THEIR TOP 3 LIKED TAG CATEGORIES**

**A screenshot of a computer

Description automatically generated**

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

We don’t have any data regarding videos or reel. So, I am going to find the correlation between photo uploads and user activity/engagement

The query uses three CTEs (uploads, likes, comments) to calculate the number of photo uploads, total likes, and total comments for each user.

* **Uploads CTE:** Counts the number of photos uploaded by each user.

**A screenshot of a computer program

Description automatically generated**

* **Likes CTE:** Counts the total likes received on all photos uploaded by each user.

A screenshot of a computer

Description automatically generated

**Comments CTE**: Counts the total comments received on all photos uploaded by each user.

A screenshot of a computer program

Description automatically generated

* **Final Output 01**: The main query combines the results from the CTEs, producing a summary table with user details, including the number of uploads, likes, and comments.

A screenshot of a computer program

Description automatically generated

**RESULT 01:**

**A screenshot of a computer

Description automatically generated**

**OBSERVATION:**

1. **User Activity Levels:**

* **Photo Uploads:** The number of photos uploaded by users ranges significantly, with some users not uploading any photos at all, while others have uploaded up to 12 photos (e.g., Eveline95).
* **Total Likes and Comments:** These metrics are directly linked to user engagement, with higher counts typically correlating to more uploads and interactions.

1. **Engagement Metrics:**

* Users who upload more photos tend to receive more likes and comments. For instance, **Eveline95** uploaded 12 photos and received 420 likes and 329 comments, demonstrating a strong positive correlation between photo uploads and user engagement.
* Conversely, users with no uploads often have zero likes and comments (e.g., **Aniya\_Hackett**, **Kasandra\_Homenick**, and others).

**CORRELATION BETWEEN UPLOADS AND ENGAGEMENT:**

**Final Output 02**: It calculates the average total engagement (likes + comments) per user, grouped by the number of photo uploads.

**A screenshot of a computer

Description automatically generated**

**RESULT 02:**

A screenshot of a computer

Description automatically generated

* There is a clear positive correlation between the number of photo uploads and total engagement. As the number of photo uploads increases, so does the average total engagement.
* Users with 12 photo uploads have the highest average total engagement (749), showing that users who post more frequently tend to engage more.
* Between 10 and 11 uploads, the engagement growth slows down, showing that merely uploading more content does not always guarantee significantly higher engagement. This shows that while uploads increase the chance of engagement, quality content is crucial.

**RECOMMENDATIONS FOR CONTENT CREATION AND CURATION STRATEGIES**

1. **User Engagement Enhancement**

* Initiate campaigns and contests that reward users for uploading photos and engaging with the community**.**
* Collaborate with highly engaged users to promote content initiatives and encourage participation from others.

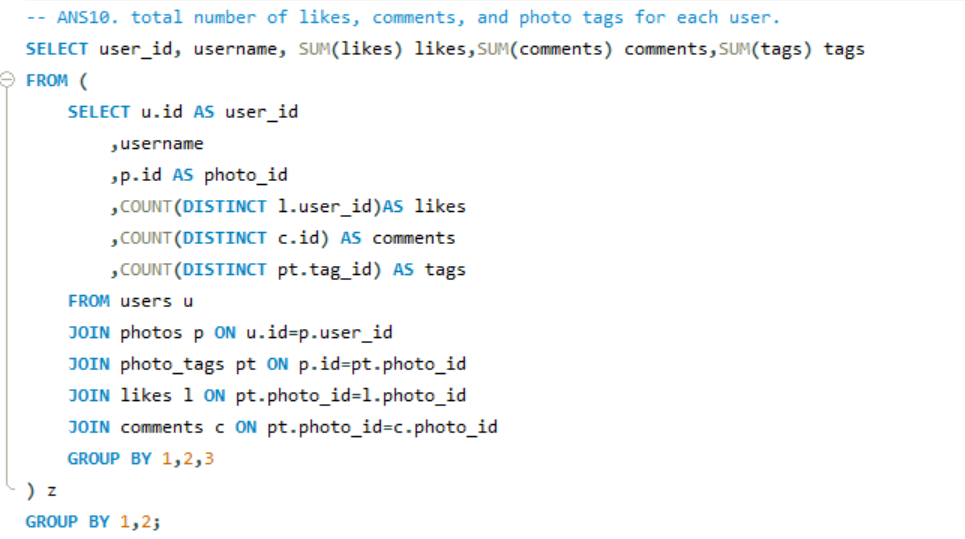
1. **Content Performance Analysis**

* Focus on identifying which types of photos or content formats generate the most likes and comments to inform future content creation.
* Schedule regular content prompts and themes based on user preferences and engagement patterns to maximize interaction.

1. **Personalized Communication Strategies**

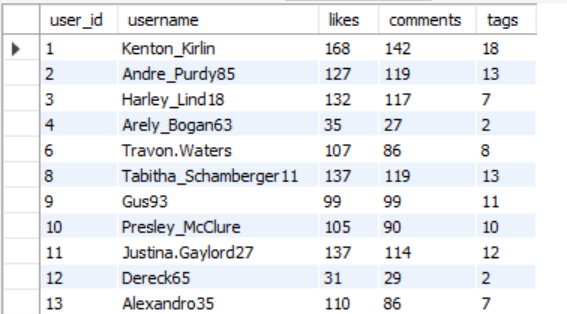
* Implement personalized messaging campaigns that encourage inactive users to participate by showcasing community highlights and success stories.
* Broaden the content strategy to include videos and reels, analysing engagement data to ensure alignment with user interests.

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



1. **Subquery Aggregation**: The subquery calculates **distinct counts** of likes, comments, and tags for each photo, grouped by user\_id, username, and photo\_id.
2. **Main Query Aggregation**: The outer query **sums** the total likes, comments, and tags per user across all their photos, providing user-level aggregates.
3. **Joins**: The query joins **users**, **photos**, **photo\_tags**, **likes**, and **comments** to link user activities (likes, comments, tags) with their respective photos.

**RESULT:**



**OBSERVATION:**

**1. High Engagement Users:**

* **Cesar93** (277 likes, 251 comments, 16 tags) and **Clint27** (264 likes, 212 comments, 26 tags) have the highest likes, comments, and tags, indicating high engagement with their content.
* **Eveline95** also stands out with **347 likes, 276 comments**, and **24 tags**, showcasing highly engaging content.

**2. Moderate Engagement Users:**

* Users like **Jaime53** (239 likes, 197 comments, 15 tags) and **Delfina\_VonRueden68** (197 likes, 174 comments, 17 tags) exhibit steady engagement across all metrics, indicating a strong but not top-performing presence.

**3. Low Engagement Users:**

* Users like **Aiyana\_Hoeger** (28 likes, 35 comments, 5 tags) and **Kenneth64** (39 likes, 31 comments, 1 tag) have low engagement across all categories, suggesting less impactful content.

**4. Tags and Engagement:**

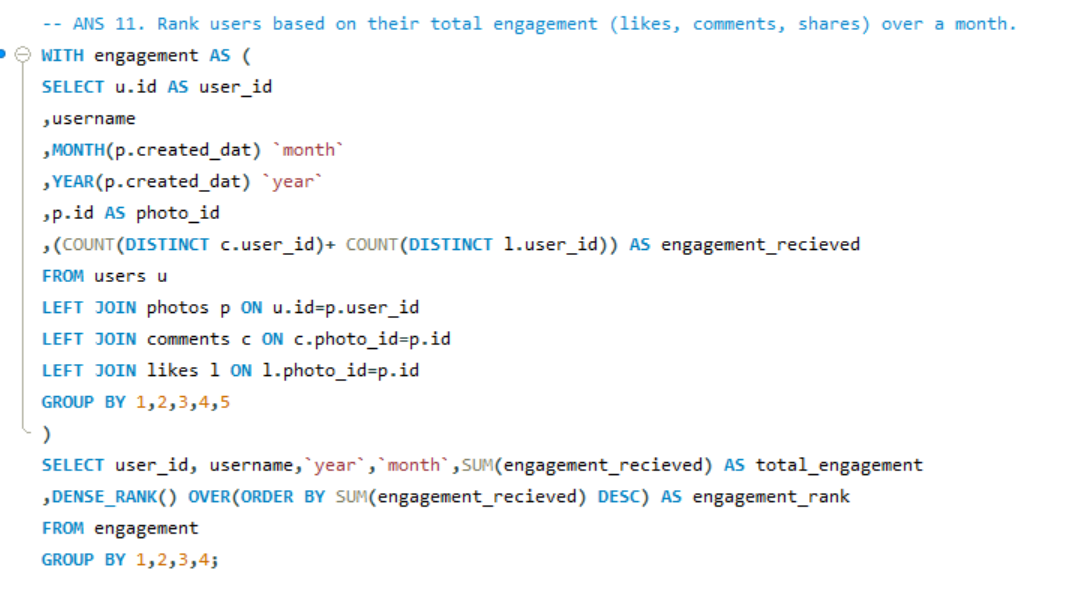
* Users with **more tags** (e.g., **Clint27**, **Cesar93**, **Eveline95**) generally show **higher engagement** (more likes and comments), implying that using more tags correlates with better visibility and interaction.

**5. Inconsistent Engagement:**

* Some users like **Kelsi26** (39 likes, 27 comments, 1 tag) and **Aurelie71** (210 likes, 185 comments, 20 tags) have a noticeable gap between tags and overall engagement, indicating possible missed opportunities for better interaction or content optimization.

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

* **Subquery (engagement)**: Calculates the engagement received for each user per photo by summing distinct likes and comments on their photos, grouping by user, photo, month, and year.
* **Engagement Calculation**: Engagement is the sum of unique users who liked or commented on each user's photo in a specific month and year.
* **Main Query**: Aggregates the total engagement per user, month, and year by summing the engagement received from the subquery.



RESULT

A screenshot of a computer

Description automatically generated

OBSERVATION

1. **Top Performers**: Eveline95 has the highest total engagement (749), followed by Clint27 (660) and Cesar93 (646), securing the top three spots in engagement rank.

2. **Moderate Engagement:** A cluster of users like Donald.Fritsch, Janet.Armstrong, and Zack\_Kemmer93 rank between 7th and 9th, with total engagements ranging from 333 to 392, showing moderate activity.

3. **Low Engagement Users**: Several users, including Meggie\_Doyle (75), Jaylan.Lakin (73), and Granville\_Kutch (71), have total engagements below 100, ranking towards the bottom of the list.

4. **Zero Engagement:** A significant number of users (e.g., Leslie67, Tierra.Trantow, Pearl7, and others) have zero total engagements, tied for the 63rd rank. This indicates inactive or non-participating users.

5. **Engagement Distribution:** Engagement scores range from 749 down to 0, with a noticeable drop-off in activity beyond the top 10. The top ranks have significantly higher engagement compared to the rest, indicating a few highly active users and many with minimal activity.

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

**CTE for Likes and Averages**:

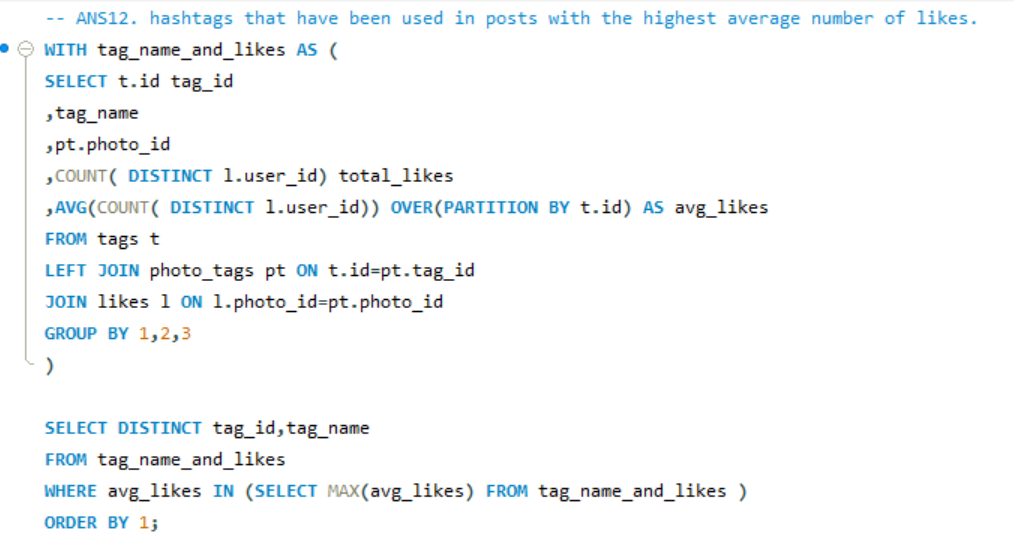
It total and average likes for each tag by counting distinct users who liked associated photos.

**Max Average Likes Filter**:

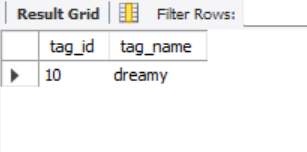
It uses a subquery to find tags that have the maximum average likes

**Distinct Tag Output**:

The final output returns distinct tag\_id and tag\_name for the tags with the highest average likes

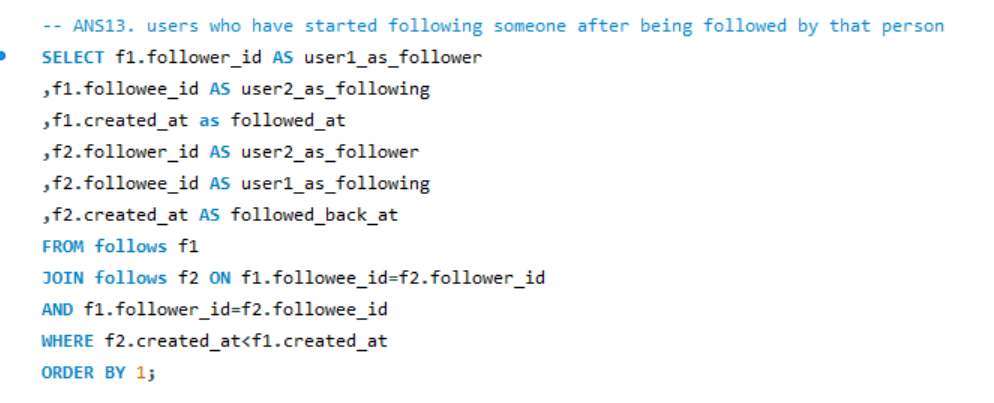
****

**RESULT:**

****

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

* The query identifies reciprocal follow relationships where one user (user1) follows another user (user2), and user2 follows user1 back.
* Used self-join on the follows table (aliased as f1 and f2) to match pairs of users by ensuring that the followee of f1 is the follower of f2, and the follower of f1 is the followee of f2.
* The WHERE clause filters results to include only those cases where user2 followed back (f2.created\_at) before user1 followed them (f1.created\_at).



RESULT ****

There is no user who started following someone after being followed by that person. As each follower followed each other at same time

**Subjective Questions**

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

To answer this question, I will find the top 10 users who are more active and making their audiences to engage with them.

With the help of **CTE-1, BOT USERS**, I first highlighted the ‘BOT USERS’ who like

and comment on every photo, which is generally rare for a real authentic user.

And many other reasons are:

* Bot users and inactive accounts are excluded from further analysis because their engagement is not indicative of genuine activity or loyalty, leading to misguided assessments of user loyalty and value.
* Analysing only real, engaged users helps allocate marketing and promotional resources efficiently, maximizing return on investment by targeting those who genuinely contribute to the platform's succeA screenshot of a computer program

  Description automatically generated

**CTE-2 engagement** – It calculates the total engagement received by each user from their photos by counting the unique users who commented or liked their photos.

**A screenshot of a computer code

Description automatically generated**

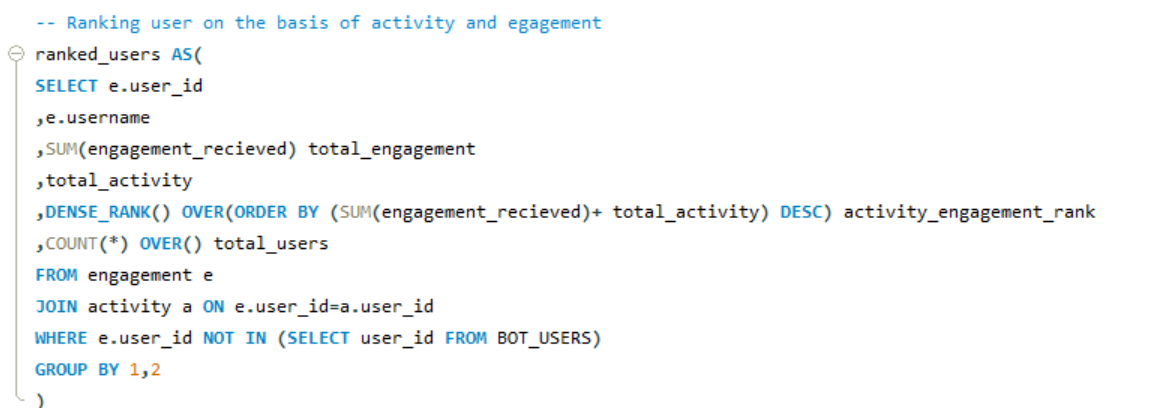
**CTE-3 activity** – It calculates the total activity of each user by counting the distinct photos they have commented on and liked.

**A computer screen shot of a code

Description automatically generated**

**CTE-3 Ranking users**

* The CTE is creating a dataset that ranks users by **engagement** (how much engagement they’ve received) and **activity** (the level of activity they’ve participated in).
* By doing this, the CTE prepares data that could later be used to select top users, analyze user behavior, or segment users for incentivization.

****

If I run this CTE as main query then the result will be as follows:

**A screenshot of a data

Description automatically generated**

**MAIN QUERY: Selecting users who are in top 15% of engagement level and activity level**

**A white background with blue text

Description automatically generated**

**A screenshot of a data

Description automatically generated**

**VISUALISATION:**

**OBSERVATION:**

**User Activity**

* **Low Activity (0-9):** There are **23 users** with a total activity between 0 and 9. This indicates that these users are very low in terms of their overall activity (posting, commenting, liking).
* **Moderate Activity (130-139):** The next largest group, **8 users**, has an activity between 130 and 139, indicating a notable but smaller cluster of users with a moderate activity level.
* **High Activity (170-179):** There is **only 1 user** in the 170-179 activity range, suggesting that very high activity levels are rare.

**User Engagement**

* **Low Engagement (0-49):** **13 users** fall in the 0-49 engagement range, which could imply low interaction or minimal content consumption.
* **Moderate Engagement (50-99):** A relatively large portion of users, **18 users**, have an engagement range between 50 and 99, showing that this group engages but not to a significant extent.
* **High Engagement (300-349):** There are **13 users** in the 300-349 range, which is a decent number indicating moderate to high engagement.
* **Very High Engagement (500+):** The highest engagement groups (500-549, 550-599, 600-649, and 700-749) have very **few users**—only **1-2 users** each in these ranges—indicating that extremely high engagement is rare among the dataset.

**TOP 10% USERS**

(\*Even with **zero activity**, users can still be considered valuable if they have **high influence**, **large follower base**, or are **potentially re-engageable**.)

**OBSERVATION**

**Users with Zero Activity but High Engagement (Receiving Engagement on Their Posts):**

* Eveline95 (749 total engagement, 0 activity), Clint27 (660 total engagement, 0 activity), and Cesar93 (646 total engagement, 0 activity)
* These users have successfully **attracted attention to their own posts** (high engagement) without interacting with others' content (zero activity). They may be **highly influential content creators** or **popular personalities** who others are eager to engage with. This indicates they could be **valuable for influencer marketing** strategies or **featured content promotion**

**Users with High Activity but Moderate Engagement (Interacting with Others’ Content but Receiving Moderate Engagement):**

* Janet.Armstrong (334 total engagement, 158 total activity), Alexandro35 (329 total engagement, 151 total activity), and Adelle96 (321 total engagement, 156 total activity)
* These users are **very active in the community**, engaging with others by liking, commenting, and sharing their content, but they don’t receive as much engagement on their own posts in return. This could mean they are **important community builders** or **networkers**

**Balanced Users with Both High Engagement and High Activity (Engaged Creators and Active Community Members):**

* Josianne.Friesen (309 total engagement, 163 total activity) and Colten.Harris76 (320 total engagement, 143 total activity)
* These users are **engaged creators** who not only receive attention on their own posts but also actively interact with others. They represent **high-value community members** who contribute to the platform in multiple ways: they create engaging content and foster community interaction.

**REWARD OR INCENTIVIZE:**

By leveraging the insights from the query results, the platform can effectively reward its most valuable users, enhancing overall user satisfaction and loyalty

**PERSONALIZED RECOMMENDATION:**

**1. High Engagement, Low Activity Users (e.g., Eveline95, Clint27, Cesar93):**

* Offer early access to new features and content to make them feel valued.
* Provide opportunities for brand partnerships and monetization.
* Award special badges or titles like "Top Creator" to recognize their content impact.

**2. High Activity, Moderate Engagement Users (e.g., Janet.Armstrong, Alexandro35, Adelle96):**

* Increase their post visibility with featured or boosted content.
* Offer community leader badges or VIP status for their contributions.
* Invite them to special events or workshops with influencers and brands.

**3. High Engagement, High Activity Users (e.g., Josianne.Friesen, Colten.Harris76):**

* Provide access to exclusive rewards, early features, and personalized perks.
* Offer cash bonuses, gift cards, or discounts as tangible incentives.
* Publicly recognize their contributions with badges or leaderboards.

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

While observing the data, I get to know there are total **26 inactive users** out of 100 users which gives a inactive user percentage =26%

A 26% inactive user percentage can be considered relatively high for a social media platform. Here’s how this can affect the platform's analysis and performance:

* **Distorted Engagement Metrics:** A high inactive user percentage skews engagement metrics, misleading perceptions of user activity and overall platform vitality.
* **Feedback Loop Challenges**: A high number of inactive users could create a feedback loop where the platform struggles to gather valuable insights from user behaviour, making it challenging to improve features and content that enhance user satisfaction.
* **Retention and Revenue Concerns:** Increased inactivity signals potential retention issues, impacting advertiser interest and revenue generation due to a less attractive user base.
* **User Retention Challenges:** High inactivity may indicate retention issues, suggesting that the platform might not be meeting user needs or expectations. This could lead to further drops in engagement if not addressed.

**QUERY & RESULT**

**CTE:** The CTE categorizes users based on whether they have posted any photos Using CASE statement.

The **CASE** statement assigns the label:

* 'Inactive User' if the user has not uploaded any photos (i.e., p.id IS NULL).
* 'Active User' if the user has uploaded at least one photo (i.e., p.id IS NOT NULL).

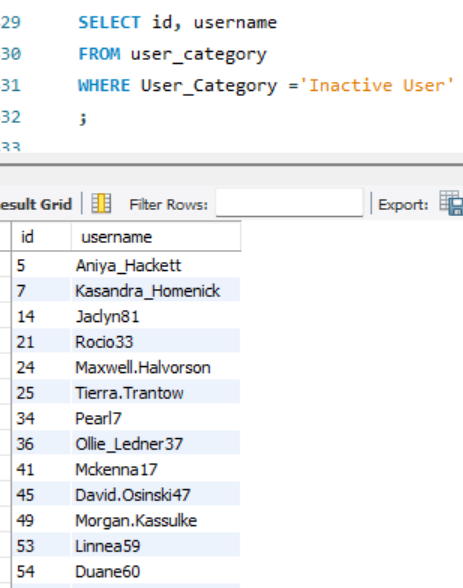
**A screen shot of a computer program

Description automatically generated**

**MAIN QUERY: I**t retrieves the id and username of all users who have not uploaded any photos, effectively identifying users who are inactive.

**A screenshot of a computer

Description automatically generated**

****

**VISUALIZATION:**

**RECOMMEND STRATEGIES TO RE-ENGAGE THEM AND ENCOURAGE THEM TO START POSTING OR ENGAGING AGAIN**

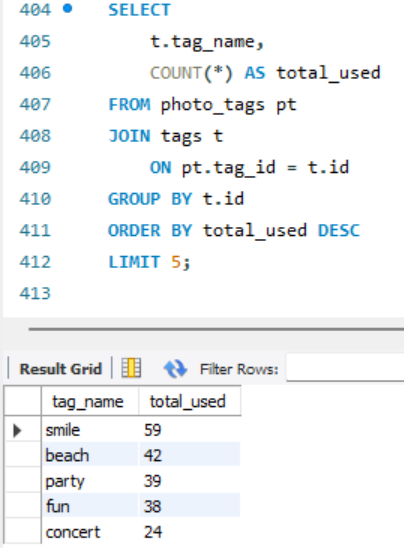
* **Personalized Outreach Campaigns:** Send personalized messages or emails highlighting new features, updates, or content they may have missed, based on their past activity, making them feel valued and invited back.
* **Exclusive Incentives and Rewards:** Offer limited-time promotions like special badges, profile boosts, or access to premium features to encourage them to post or engage with content again.
* **Gamification and Challenges:** Introduce interactive challenges, user milestones, or leaderboard systems to motivate users through friendly competition or achievements, making the platform more engaging and fun.
* **Content Recommendations and Reminders:** Use algorithms to recommend content they might like based on past interactions, along with gentle reminders about new posts or trending topics that fit their interests.
* **Community Engagement and Social Features:** Encourage them to join community groups or follow trending topics, where they can participate in discussions, comment on posts, or collaborate with others, fostering a sense of belonging and active involvement.

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

To know which hashtags are frequently used, I wrote a query that will give the 5 most used tags

* I joined the photo\_tags table (pt) and the tags table (t) based on the tag ID (t.id = pt.tag\_id) to get the tag names associated with each photo.
* Counted how many times each tag is used across all photos, grouping the results by the tag ID (t.id).
* The results are sorted in descending order by the number of times the tags have been used (total\_used), and only the top 5 tags are returned (LIMIT 5).

**QUERY AND RESULT:**

****

**A screenshot of a computer

Description automatically generated**

**VISUALIZATION:**

**INFORMATION GUIDE CONTENT STRATEGY AND AD CAMPAIGNS:**

* + 1. **Content Creation & Themed Engagement Campaigns**
* **Strategy:** Focus content on top-performing hashtags like **#smile, #beach, #party, #fun** by developing regular posts, challenges, or contests.
* **Example:** Create weekly content series (e.g., "Beach Fridays" or "Fun Challenges") that encourage user participation with these hashtags.
* **Benefit:** Increases organic engagement by leveraging themes that users are already interested in, keeping them active and involved.
  + 1. **Influencer Partnerships & Sponsored Ads**
* **Strategy:** Collaborate with influencers and launch ad campaigns centred around trending hashtags such as **#concert, #beach** to promote products or experiences.
* **Example:** Partner with travel influencers to promote beach vacations using **#beach** or run ads targeting concert-goers using **#concert**.
* **Benefit:** Amplifies reach through influencer audiences and maximizes ad effectiveness by tapping into already engaged communities.
  + 1. **Seasonal & Event-Based Marketing**
* **Strategy:** Tie high-performing hashtags like **#party** and **#fun** to seasonal events or campaigns (e.g., holiday parties or summer festivals).
* **Example:** During holidays, run a "Holiday Party" campaign using **#party** to showcase event-related products or experiences.
* **Benefit:** Aligns content with popular, timely themes to boost engagement during high-traffic periods, ensuring relevance and interaction.

1. Are there any patterns or trends in user engagement based on demographics (age, location, gender) or posting times? How can these insights inform targeted marketing campaigns?

Although IG\_CLONE does not contain traditional demographic data such as age, gender, or location, the two provided queries still answer the question by using proxy variables

**Query 1: Joining Year, Follower Count, and Engagement Patterns**

**What this query does:**

* **User Grouping by Year and Follower Count**: It segments users based on the year they joined and their follower count.
* **Metrics Calculated**:
  + **User Count**: The number of distinct users within each group (defined by joining\_year and follower\_count).
  + **Average Post Count**: The average number of posts made by users in each group.
  + **Average Engagement**: The average engagement (sum of unique likes and comments) received by users in each group.

**How it addresses the question:**

* **Proxy for Demographics**: The use of joining\_year as a demographic proxy helps observe patterns across different user groups based on how long they’ve been using the platform.
* **Engagement and Activity Analysis**: By analyzing follower\_count alongside engagement metrics (comments, likes), the query provides insights into how follower count influences engagement. For example, users with a higher follower count may receive more engagement, but this query helps confirm or refute such patterns.

**Key Observations**:

* **Established vs. New Users**: The query allows you to compare engagement patterns of older users (e.g., those who joined in 2016) with newer users (e.g., those who joined in 2017). Established users might show more consistent engagement.
* **Engagement Trends**: It helps understand whether users with higher follower counts generally receive higher engagement and if post frequency correlates with engagement levels.



RESULT

A screenshot of a computer

Description automatically generated

**Older users (2016)** who have higher engagement but fewer posts may be ideal for retention campaigns, as they already receive significant engagement despite lower activity. Encouraging them to post more could strengthen their connection to the platform. **Newer users (2017)**, who post more frequently and also receive high engagement, may be better suited for acquisition strategies, focusing on growing their follower base and increasing their long-term platform usage.

**VISUALIZATION:**

**Query 2: Tag Usage and Content Preferences**

**What this query does:**

* It looks at the **tags** that users attach to their posts and analyzes which tags are most popular among users based on their **joining year**.
* The query ranks tags based on how many users applied them, showing which types of content are trending among different user cohorts.

**How it addresses the question:**

* **Proxy for Content Demographics**: Without explicit demographic data, **tag usage** can reveal user preferences and content trends. The popularity of certain tags like “smile”, “party”, and “fun” tells us what kind of content resonates with different user groups.
  + For example, users from **2016** may prefer aspirational and social themes (e.g., “beach”, “party”, “beauty”), while **2017** users gravitate toward humor and social engagement (e.g., “fun”, “lol”, “happy”).
* **Content-Based Segmentation**: This gives us insight into how different content types (via tags) are consumed by users from different years, acting as a proxy for understanding what each group is interested in.

A computer screen shot of a computer code

Description automatically generated

RESULT

A screenshot of a computer

Description automatically generated

Users who favor **fun and humor** (2017) can be targeted with lighthearted campaigns, while **2016** users may respond better to lifestyle and social event-driven content.

VISUALIZATION

**RECOMMENDATION & STRATEGIES**

While **traditional demographics** are not available, these two queries provide a **behavioral and content-based analysis** that can act as a substitute:

* **Joining year** and **follower count** act as proxies for segmenting user groups.
* **Tag usage** offers insights into user preferences and content consumption patterns, allowing the marketing team to design campaigns based on what users engage with most.

**Recommendations for Targeted Marketing:**

* **For 2016 Users**:
  + Content should focus on **aspirational and lifestyle** themes such as travel, beauty, and social gatherings.
  + Since these users are more established, retention strategies like **loyalty rewards** or exclusive content can work well.
* **For 2017 Users**:
  + Focus on **fun, humor, and social engagement** as reflected in their tag usage.
  + Since these users may be newer and more active, acquisition strategies or **campaigns that encourage participation** (e.g., Instagram challenges) can help boost engagement.

By focusing on these behavioural and content trends, Meta can effectively drive user engagement, retention, and acquisition, even without traditional demographic data.

1. Based on follower counts and engagement rates, which users would be ideal candidates for influencer marketing campaigns? How would you approach and collaborate with these influencers?

* **CTE-1 followers:** It calculates the total follower count for each user.A screenshot of a computer code

  Description automatically generated
* **CTE-2 engagement:** Itcalculates the total engagement received by each user based on their photos.A screenshot of a computer program

  Description automatically generated
* **CTE-3 influencers:** It identifies potential influencers by calculating their average engagement rate.
* **Main Quey:** The query effectively retrieves the top 5 potential influencers based on their follower count and average engagement rate, ensuring that the most influential users are prioritized for analysis or marketing campaigns.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

RESULT

**A screenshot of a computer

Description automatically generated**

**APPROACHING AND COLLABORATING WITH THE IDENTIFIED INFLUENCERS CAN BE STRUCTURED AS FOLLOWS:**

* **Personalized Outreach:**

Crafting personalized messages for each influencer, highlighting their unique engagement rates and follower counts. Acknowledge their content style and audience to build rapport and demonstrate that we've done our research. This approach fosters a genuine connection and sets a positive tone for collaboration.

* **Tailored Campaigns**:

Developing tailored marketing campaigns that align with each influencer's niche and audience. For instance, if an influencer specializes in lifestyle content, we could create campaigns that seamlessly integrate our brand products into their lifestyle posts. This ensures that the partnership feels authentic to their followers and enhances the campaign's effectiveness.

* **Incentives and Benefits**:

Offer attractive incentives for collaboration, such as competitive compensation, free products, or exclusive access to events. Additionally, we can highlight potential growth opportunities for the influencer, such as co-branded content or exposure to our broader audience through our marketing channels. This motivates influencers to engage actively in promoting our brand.

* **Performance Metrics and Feedback**:

Establish clear performance metrics to measure the success of the influencer partnership. This includes tracking engagement rates, conversions, and overall reach. Regular feedback sessions will help us optimize future campaigns and maintain strong relationships with the influencers.

* **Long-term Partnerships**:

Consider developing long-term partnerships with the most effective influencers. This not only ensures consistent brand representation but also builds trust with their audience over time, making our collaborations more impactful and effective in driving brand awareness and loyalty.

1. 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, I categorised User into two categories

**Category 1: Activity Status**

* **No Activity**: total\_activity = 0
* **High Activity**: total\_activity >= 250
* **Low Activity**: total\_activity < 150
* **Moderate Activity**: total\_activity BETWEEN 150 AND 250

**Category: Engagement Status**

* **No Engagement:** engagement\_received = 0
* **High Engagement:** engagement\_received >= 300
* **Low Engagement:** engagement\_received < 200
* **Moderate Engagement:** engagement\_received BETWEEN 200 AND 300

A computer code with text

Description automatically generated with medium confidence

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A close-up of a text

Description automatically generated**

RESULT

**A screenshot of a data

Description automatically generated**

VISUALIZATION

**OBSERVATION AND STRATEGY:**

1. **No Activity and High Engagement**
   * Example User: Eveline95 (engagement\_received: 749, total\_activity: 0)
   * Strategy:
     + Engage through personalized emails or notifications to encourage participation.
     + Highlight community benefits and exclusive content to motivate them to take action.
2. **Moderate Activity and Moderate Engagement**
   * Example User: Billy52 (engagement\_received: 244, total\_activity: 161)
   * Strategy:
     + Nurture with regular updates and promotions tailored to their interests.
     + Introduce loyalty programs or exclusive content to enhance engagement.
3. **Low Activity and Moderate Engagement**
   * Example User: Dario77 (engagement\_received: 252, total\_activity: 136)
   * Strategy:
     + Use gamification and challenges to stimulate activity.
     + Communicate the benefits of increased participation through targeted messaging.
4. **Moderate Activity and Low Engagement**
   * Example User: Presley\_McClure (engagement\_received: 195, total\_activity: 150)
   * Strategy:
     + Identify barriers to engagement and tailor content to their interests.
     + Create feedback loops to understand their preferences better.
5. **High Activity and No Engagement**
   * Example User: Aniya\_Hackett (engagement\_received: 0, total\_activity: 514)
   * Strategy:
     + Analyze activity patterns and implement feedback mechanisms to understand their needs.
     + Motivate engagement through targeted campaigns or relevant content suggestions.
6. **Low Activity And High Engagement**
   * Example User: Travon. Waters (engagement\_received: 312, total\_activity: 144)
   * Strategy:
     + Encourage more frequent interactions with relevant content or curated experiences.
     + Highlight benefits of increased activity to enhance their engagement.
7. **No Activity and No Engagement**
   * Example User: Kasandra\_Homenick (engagement\_received: 0, total\_activity: 0)
   * Strategy:
     + Develop campaigns that offer incentives, personalized messages, or surveys to understand their lack of participation.
     + Use onboarding strategies to educate them about platform benefits.
8. **Moderate Activity and High Engagement**
   * Example User: Yvette.Gottlieb91 (engagement\_received: 316, total\_activity: 138)
   * Strategy:
     + Recognize loyalty with rewards and create community-building opportunities.
     + Offer exclusive events or content tailored to their interests.
9. **Low Activity and Low Engagement**
   * Example User: Yazmin\_Mills95 (engagement\_received: 65, total\_activity: 147)
   * Strategy:
     + Create re-engagement strategies that include educational content about platform benefits.
     + Implement onboarding campaigns that clearly communicate value**.**
10. If data on ad campaigns (impressions, clicks, conversions) is available, how would you measure their effectiveness and optimize future campaigns?

Assuming, there’s a **Table campaign metrics** including campaign name, impressions, clicks, conversions, and dates.

**Querying Campaign Effectiveness**

To measure the effectiveness, we would focus on calculating key metrics such as Click-Through Rate (CTR) and Conversion Rate.

**A. CALCULATING CTR**

SELECT

campaign\_name,

SUM(impressions) AS total\_impressions,

SUM(clicks) AS total\_clicks,

(SUM(clicks) / NULLIF(SUM(impressions), 0)) \* 100 AS ctr

FROM

campaign\_metrics

GROUP BY

campaign\_name;

**B. CALCULATING CONVERSION RATE**

SELECT

campaign\_name,

SUM(clicks) AS total\_clicks,

SUM(conversions) AS total\_conversions,

(SUM(conversions) / NULLIF(SUM(clicks), 0)) \* 100 AS conversion\_rate

FROM

campaign\_metrics

GROUP BY

campaign\_name;

**Combining Metrics for Comprehensive Analysis**

SELECT

cm.campaign\_name,

SUM(cm.impressions) AS total\_impressions,

SUM(cm.clicks) AS total\_clicks,

(SUM(cm.clicks) / NULLIF(SUM(cm.impressions), 0)) \* 100 AS ctr,

SUM(cm.conversions) AS total\_conversions,

(SUM(cm.conversions) / NULLIF(SUM(cm.clicks), 0)) \* 100 AS conversion\_rate

FROM

campaign\_metrics cm

GROUP BY

cm.campaign\_name;

**OPTIMIZATION STRATEGIES**

After calculating these metrics, the next step is to analyze the data and determine where improvements can be made. This can involve:

* **Identifying High-Performing Campaigns**: Analyzing the CTR and conversion rate to identify campaigns that performed well and may warrant additional budget allocation.
* **Finding Underperforming Campaigns**: Investigating campaigns with low CTR or conversion rates to understand potential reasons (e.g., targeting issues, creative quality).

**RECOMMENDATIONS FOR OPTIMIZING FUTURE CAMPAIGNS**

**1. Evaluate Click-Through Rate (CTR):**

* **CTR Analysis**: If certain campaigns have significantly higher CTRs, it indicates that the ad creatives or targeting resonates well with the audience. Conversely, low CTRs might suggest that the ads aren't compelling enough or that the targeting needs adjustment.
* **Recommendation**: For campaigns with low CTR, experiment with different ad creatives (e.g., changing headlines, visuals, or CTAs). Additionally, review and refine the audience segmentation to better target users interested in the product or service.

**2. Analyze Conversion Rate:**

* **Conversion Rate Analysis**: A high conversion rate indicates that the landing pages and user experience are effective. Low conversion rates suggest there might be friction points in the user journey post-click (e.g., complex forms, slow-loading pages).
* **Recommendation**: For campaigns with a low conversion rate, review the user flow on landing pages. A/B testing different elements on the page (like headlines, forms, or visuals) and ensuring fast load times can improve conversion rates.

**3. Identify Underperforming Campaigns:**

* **Underperforming Campaigns**: Some campaigns might show low impressions, CTR, and conversions across the board.
* **Recommendation**: For campaigns consistently underperforming, consider reallocating the budget to higher-performing ones or refining the campaign targeting by using lookalike audiences, retargeting previous site visitors, or optimizing the delivery schedule.

**4. Optimize for High Performers:**

* **Top-Performing Campaigns**: Analyze which campaigns have the best balance of high impressions, high CTR, and high conversions. These campaigns are likely aligned well in terms of creative, targeting, and timing.
* **Recommendation**: Invest more in the top-performing campaigns by increasing the budget or expanding the reach to similar audiences (e.g., lookalike audiences). Replicate these tactics across other campaigns.

**5. Segmentation of Campaign Performance:**

* **Segmentation**: Break down campaign performance by specific audience demographics, regions, or devices. This helps identify if certain groups are responding better to certain campaigns.
* **Recommendation**: Tailor future campaigns to the segments showing the highest engagement and conversions. Create more personalized messaging based on these insights to optimize the impact.

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

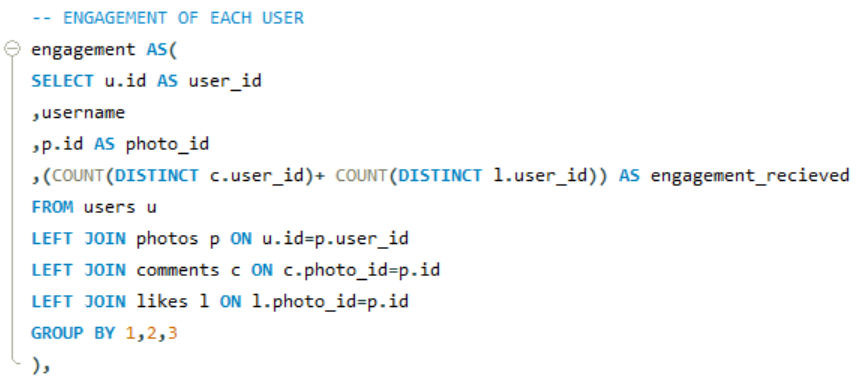
**APPROACH FOR ANALYSIS:**

1. **High Follower Count and High Engagement**:
   * Users with a high number of followers and consistently high numbers of likes and comments are prime candidates.
   * These users not only have a large reach but also have an engaged audience.
2. **Consistency in Engagement**:
   * Users who consistently receive high numbers of likes and comments across multiple posts may indicate genuine interest and influence, as opposed to one-time viral activity.
3. **Identify Users with Balanced Engagement**:
   * We should focus on users who both generate and receive engagement, as these users are more likely to influence their network and drive engagement for Instagram initiatives.

* **CTE-1 followers:** It calculates the total follower count for each user.

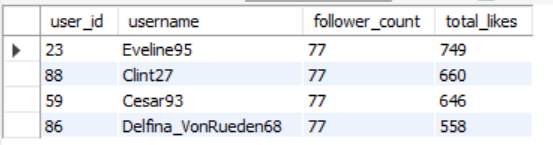
****

* **CTE-2 engagement:** Itcalculates the total engagement received by each user based on their photos.

****

* **Main Quey:** the query identifies the most engaged users from those with maximum followers, highlighting potential influencers who have both reach and engagement

**RESULT:**

****

**STRATEGY FOR ENGAGEMENT:**

* **Leverage Their Audience**: Instagram can collaborate with these users for product launches, event promotions, and community initiatives, leveraging their large, engaged audiences to increase visibility and participation.
* **Personalized Campaigns**: Tailor campaigns to these users' specific interests or the types of content they typically engage with to ensure authentic promotion and maintain high engagement levels.

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

To address the objectives of increasing user engagement, retention, and acquisition, I would follow a systematic approach consisting of several steps

**Step 1: Analyze User Engagement**

**Objective**: Identify which users are most active in terms of likes, comments, and photo uploads.

* **Query 1**: Calculate total likes and comments per user.

**Step 2: Identify User Retention**

**Objective**: Determine how long users have been active on the platform and identify trends in retention.

* **Query 2**: Calculate the age of each user account based on the created\_at timestamp.

**Step 3: Explore User Activity Patterns**

**Objective**: Examine user activity patterns over time to identify trends in engagement.

* **Query 3**: Analyze engagement metrics (likes and comments) over time

**Step 4: Identify Potential Brand Ambassadors**

**Objective**: Identify users who are highly engaged and have a significant follower count, making them potential brand advocates.

* **Query 4**: Combine follower counts with engagement metrics to identify advocates

**Step 5: Analyze Tag Usage**

**Objective**: Understand which tags are most frequently used and how they correlate with user engagement.

* **Query 5**: Identify popular tags and their engagement levels

Based on the insights gathered, I would like to suggest actionable recommendations to enhance user engagement, improve retention, and drive acquisition:

* Leverage the identified potential brand ambassadors (users with high engagement and follower counts) for targeted marketing campaigns. Collaborate with these influencers to promote new features, events, or brand initiatives.
* Analyze the most liked and commented photos to identify trends in user preferences. Use this data to inform content creation strategies, ensuring that marketing campaigns align with user interests.
* Use insights from the tag usage analysis to enhance discoverability. Identify which tags are trending and encourage users to utilize these tags in their posts.
* Establish feedback mechanisms to understand why certain users disengage. Analyzing user retention metrics alongside qualitative feedback can provide insights into areas for improvement.
* Implement gamification strategies to enhance user interaction. Reward users for activities such as posting, commenting, liking, and tagging.
* Continuously analyze engagement, retention, and acquisition data to identify emerging trends and adjust strategies accordingly.

1. Assuming there's a "User\_Interactions" table tracking user engagements, how can you update the "Engagement\_Type" column to change all instances of "Like" to "Heart" to align with Instagram's terminology?

To perform this Task, I will use DML- DATA MANIPULATION LANGUAGE (UPDATE statement)

UPDATE User\_Interactions

SET Engagement\_Type = 'Heart'

WHERE Engagement\_Type = 'Like';

**UPDATE User\_Interactions**: This specifies that you want to update the User\_Interactions table.

**SET Engagement\_Type = 'Heart'**: This sets the Engagement\_Type column to "Heart."

**WHERE Engagement\_Type = 'Like'**: This condition ensures that only rows where Engagement\_Type is currently "Like" are updated to "Heart."