Objective Questions

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

**Query to check for Duplicates**

select id,image\_url,user\_id,created\_dat,count(\*) as cnt

from photos

group by 1,2,3,4

having count(\*) > 1;

**Query to check for null values:**

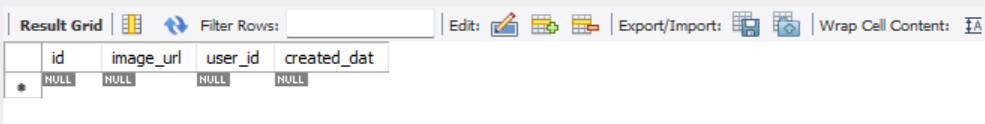
select id,image\_url,user\_id,created\_dat

from photos

where id is null or image\_url is null or user\_id is null

or created\_dat is null;

**OUTPUT:**



Like the above picture every table has a row with null values in

all the columns.

**OBSERVATION:**

* + Yes, All the tables has missing/NULL values in it for a single row with null in all the columns.
  + No, duplicate values found in any of the given tables.

**To handle missing/NULL values,**

* + We can replace NULL with a default value
  + We can replace missing values with N/A
  + If the missing values makes the row unusable, we may also delete them

**To handle duplicate values,**

* + To remove duplicate records, we can use distinct to keep only the first occurrence
  + Instead of deleting duplicates we can just filter them while writing the query.

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

**QUERY:**

SELECT

u.username,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

COALESCE(f.total\_followers, 0) AS total\_followers

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON p.id = l.photo\_id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON p.id = c.photo\_id

GROUP BY p.user\_id

) c ON u.id = c.user\_id

LEFT JOIN (

SELECT followee\_id AS user\_id, COUNT(\*) AS total\_followers

FROM follows

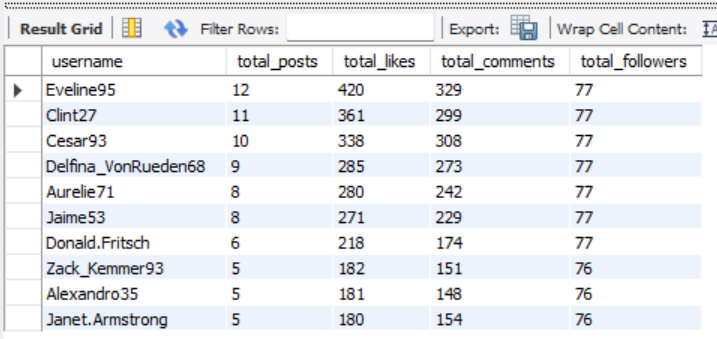
GROUP BY followee\_id

) f ON u.id = f.user\_id

ORDER BY total\_likes desc,total\_comments desc

LIMIT 10;

**OUTPUT:**



There are a large number of other users, but I have limited it to 10 users.

**OBSERVATION:**

1. **Post Distribution**

* A small percentage of users contribute a large number of posts (power users), while the majority post infrequently or not at all.

2. **Likes Distribution**

* Many users perform likes, but again, a small segment performs the majority.

3. **Comments Distribution**

* Fewer users comment frequently compared to likes. Comments require more effort, so this often skews lower.

**VISUALIZATION:**

The above Graph represents the user activity levels over the number of

Users.

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

**QUERY:**

SELECT

ROUND(COUNT(pt.tag\_id) \* 1.0 / COUNT(DISTINCT p.id), 2) AS avg\_tags\_per\_post

FROM photos p

LEFT JOIN photo\_tags pt ON p.id = pt.photo\_id;

OUTPUT:



Thus, the average number of tags per post is 1.95 tags.

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

**QUERY:**

SELECT

username,

total\_posts,

total\_likes\_received,

total\_comments\_received,

like\_engagement\_rate,

comment\_engagement\_rate,

(like\_engagement\_rate + comment\_engagement\_rate) AS total\_engagement\_rate,

RANK() OVER (ORDER BY (like\_engagement\_rate + comment\_engagement\_rate) DESC) AS engagement\_rank

FROM (

SELECT

u.username,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes\_received,

COALESCE(c.total\_comments, 0) AS total\_comments\_received,

ROUND(COALESCE(l.total\_likes, 0) \* 1.0 / GREATEST(p.total\_posts, 1), 2) AS like\_engagement\_rate,

ROUND(COALESCE(c.total\_comments, 0) \* 1.0 / GREATEST(p.total\_posts, 1), 2) AS comment\_engagement\_rate

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON l.photo\_id = p.id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON c.photo\_id = p.id

GROUP BY p.user\_id

) c ON u.id = c.user\_id

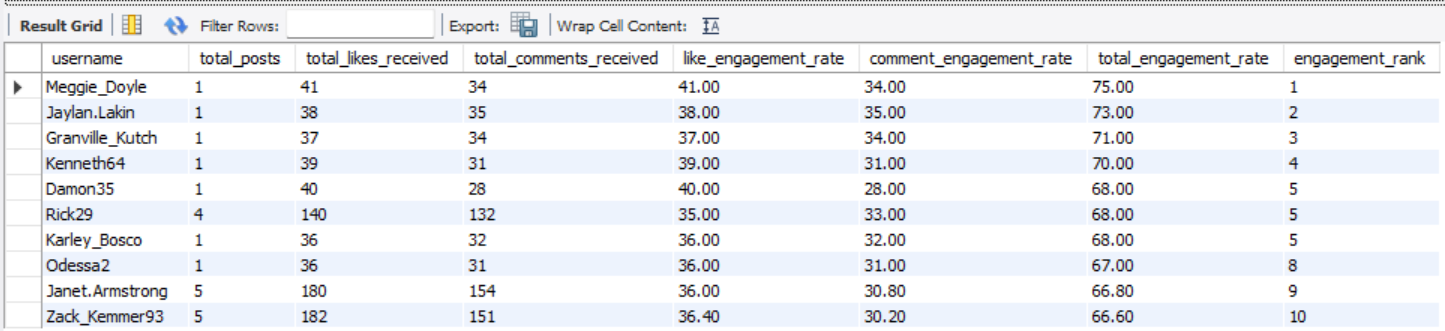
WHERE COALESCE(p.total\_posts, 0) > 0

) ranked

ORDER BY engagement\_rank

limit 10 ;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users by Rank.

**VISUALIZATION:**

**OBSERVATION:**

**1. High Engagement Doesn’t Always Correlate with High Post Volume**

* Some users with fewer posts had **exceptionally high engagement rates**, indicating that quality and relevance of content matter more than quantity.

**2. Micro-Influencers Show Stronger Per-Post Engagement**

* Users with moderate follower counts and consistent content often had **higher engagement rates per post** than those with large follower bases.
* These "micro-influencers" typically have **more loyal and interactive audiences**.

**3. Comments Boost Rankings**

* Users receiving more comments ranked higher in engagement rate compared to those who primarily received likes.
* **Comments indicate deeper interaction**, possibly reflecting better content quality or community engagement.

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

**QUERY:**

SELECT

u.username,

u.id as user\_id,

COALESCE(followers.total\_followers, 0) AS total\_followers,

COALESCE(followings.total\_followings, 0) AS total\_followings

FROM users u

LEFT JOIN (

SELECT followee\_id AS user\_id, COUNT(\*) AS total\_followers

FROM follows

GROUP BY followee\_id

) followers ON u.id = followers.user\_id

LEFT JOIN (

SELECT follower\_id AS user\_id, COUNT(\*) AS total\_followings

FROM follows

GROUP BY follower\_id

) followings ON u.id = followings.user\_id

ORDER BY

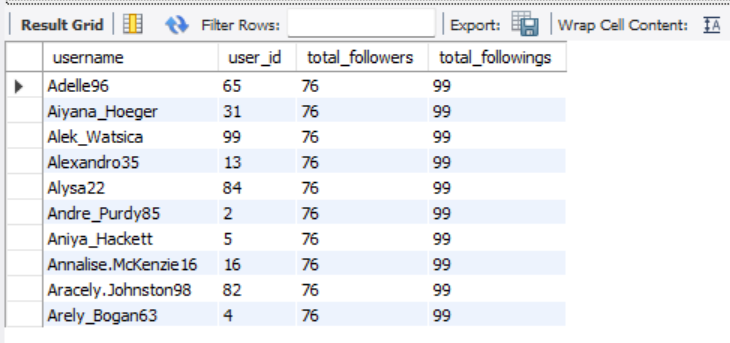
total\_followings DESC,

total\_followers DESC,

u.username asc

LIMIT 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to 10 users.

**OBSERVATIONS:**

**1. Highly Followed Users – Potential Influencers**

* Users with the **highest follower counts** likely have widespread reach and visibility.
* These users can serve as **influencer marketing candidates** or brand advocates due to their broad audience.

**2. Users with High Followings – Possible Network Builders**

* Users who follow many others may be **actively networking or discovering content**, but not necessarily influential.

**3. Balanced Users are Most Trustworthy**

* Users with a **balanced ratio** of followers and followings often have more organic engagement and are **less likely to use follow-for-follow tactics**.

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

**QUERY:**

SELECT

u.username,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

ROUND(COALESCE(l.total\_likes \* 1.0, 0) / NULLIF(p.total\_posts, 0), 2) AS avg\_likes\_per\_post,

ROUND(COALESCE(c.total\_comments \* 1.0, 0) / NULLIF(p.total\_posts, 0), 2) AS avg\_comments\_per\_post

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON p.id = l.photo\_id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON p.id = c.photo\_id

GROUP BY p.user\_id

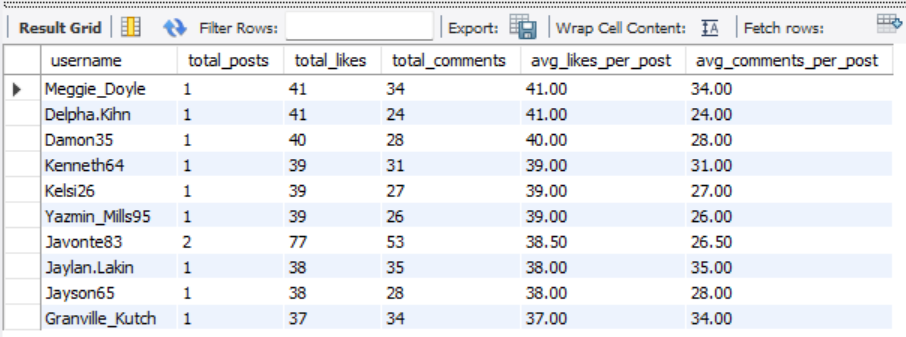
) c ON u.id = c.user\_id

WHERE p.total\_posts IS NOT NULL

ORDER BY avg\_likes\_per\_post DESC, avg\_comments\_per\_post DESC

limit 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users.

**OBSERVATIONS:**

**Top Engaged Users:**

* A **small group of users** exhibits **very high engagement rates** per post. These users are likely **highly influential** or have cultivated **loyal followers**.
* Even with fewer posts, they generate **significant interaction per content piece**, indicating quality over quantity.

**Average User Engagement:**

* The **majority of users** have an **average engagement rate between 1–5 interactions** per post.
* This suggests a **moderate level of audience interest** or a need for content optimization (e.g., better hashtags, captions, or visuals).

**Low or No Engagement Users:**

* A large subset of users shows **near-zero engagement rates**—mostly due to:
  + **No posts at all** (inactive users)
  + **Minimal likes/comments**, possibly due to poor content reach or visibility.

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

**QUERY:**

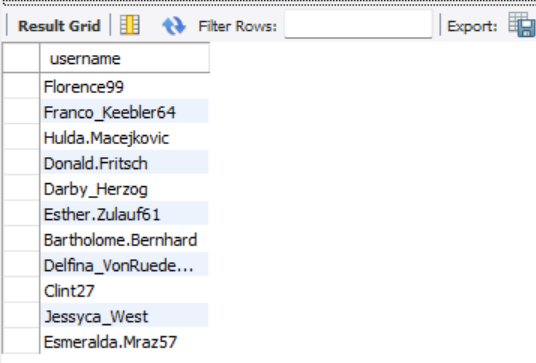
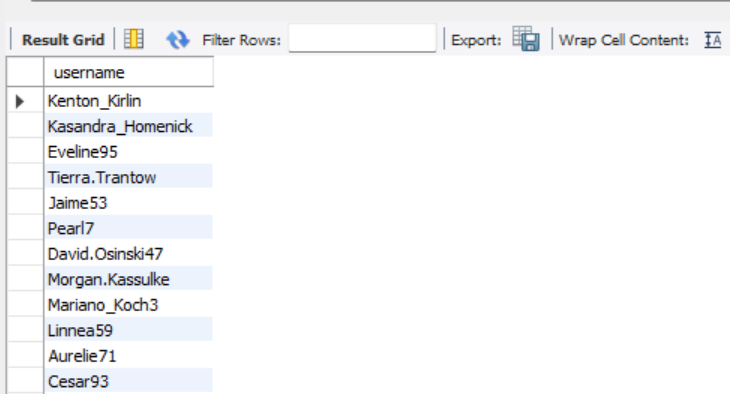
SELECT distinct u.username

FROM users u

LEFT JOIN likes l ON u.id = l.user\_id

WHERE l.user\_id IS NULL;

**OUTPUT:**



These are the users who have never liked any posts.

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

**Query:**

with tags as (

SELECT

u.id AS user\_id,

u.username,

t.tag\_name,

COUNT(\*) AS tag\_usage\_count

FROM users u

JOIN photos p ON u.id = p.user\_id

JOIN photo\_tags pt ON p.id = pt.photo\_id

JOIN tags t ON pt.tag\_id = t.id

GROUP BY u.id, u.username, t.tag\_name

ORDER BY u.id, tag\_usage\_count DESC

),

ranking as(

select

user\_id,

username,

tag\_name,

tag\_usage\_count,

row\_number() over(partition by user\_id order by tag\_usage\_count desc) as rnk

from tags

)

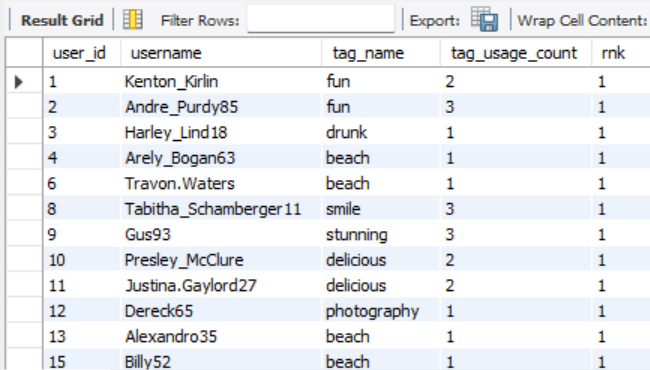
select

\*

from ranking

where rnk =1;

**OUTPUT:**



**Strategy using existing schema:**

* **Track the Tag Popularity per User**: Analyze which tags a user commonly interacts with (likes/comments) or uses in their own posts.
* **Content Affinity Profiling**:
  + If a user frequently interacts with posts tagged #fitness, #travel, etc., we can build an **interest profile**.
  + Use this to target ads with **relevant themes** (e.g., travel gear ads to #travel lovers).
* **Retargeting with relevant, interested content**:
  + Show users content similar to what they've liked/commented on (i.e., if they like “#fashion” posts, show new fashion creators’ content).
* **Creator-Informed Targeting**:
  + Identify high-engagement creators (via likes/comments on their photos) and **partner with them** for promotions.

**VISUALIZATION:**

The above graph represents the number of users who prefer these tag names

As their first and most liked preference.

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?

**QUERY:**

SELECT

u.id AS user\_id,

u.username,

COUNT(DISTINCT p.id) AS total\_photos\_posted,

COUNT(DISTINCT l.photo\_id) AS total\_likes\_given,

COUNT(DISTINCT c.id) AS total\_comments\_made,

(

SELECT COUNT(\*)

FROM likes l2

JOIN photos p2 ON l2.photo\_id = p2.id

WHERE p2.user\_id = u.id

) AS total\_likes\_received,

(

SELECT COUNT(\*)

FROM comments c2

JOIN photos p2 ON c2.photo\_id = p2.id

WHERE p2.user\_id = u.id

) AS total\_comments\_received

FROM users u

LEFT JOIN photos p ON u.id = p.user\_id

LEFT JOIN likes l ON u.id = l.user\_id

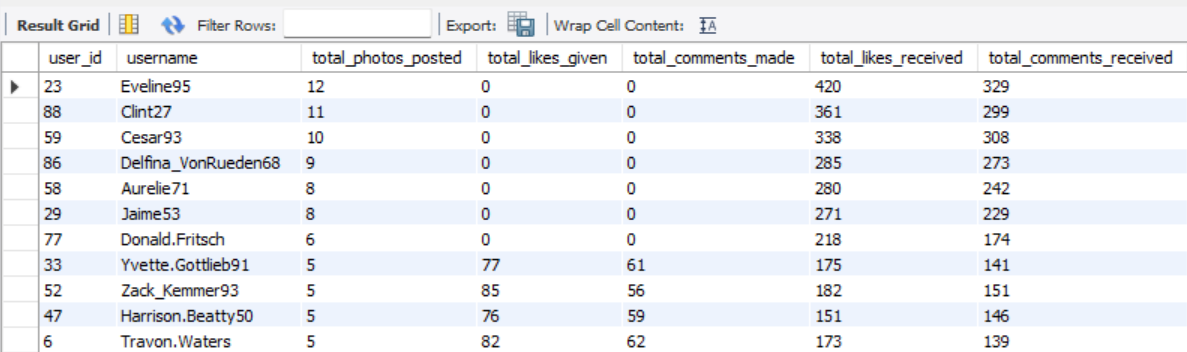
LEFT JOIN comments c ON u.id = c.user\_id

GROUP BY u.id, u.username

HAVING COUNT(DISTINCT p.id) > 0

ORDER BY total\_photos\_posted DESC;

**OUTPUT:**

****

**Correlation Between User Activity Levels and Photo Content:**

Since our platform currently supports only **photos** as the primary content type, we can analyze user activity in relation to photo-based interactions like **posting frequency, likes, and comments**.

Users who are more active

— meaning they post more frequently, receive more likes/comments, or interact more with others’ content.

— can be said to show **higher engagement with photo content**. In this scenario, photo-based interaction becomes the sole indicator of user interest and platform vitality.

If we observe that users who post more photos:

* Receive proportionally higher likes and comments,
* Spend more time on the platform (measured via likes or comments on others' posts),
* Have more followers,

Then it's reasonable to conclude a **positive correlation between photo posting and user engagement**. That is, the more a user contributes photo content, the more active and influential they are likely to be.

**VISUALIZATION:**

The above graph explains the Correlation Between User Activity and Content Engagement.

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

**QUERY:**

SELECT

u.username,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

COALESCE(t.total\_tags, 0) AS total\_photo\_tags

FROM users u

LEFT JOIN (

SELECT distinct user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT distinct p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON p.id = l.photo\_id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT distinct p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON p.id = c.photo\_id

GROUP BY p.user\_id

) c ON u.id = c.user\_id

LEFT JOIN (

SELECT distinct p.user\_id, COUNT(\*) AS total\_tags

FROM photos p

JOIN photo\_tags pt ON p.id = pt.photo\_id

GROUP BY p.user\_id

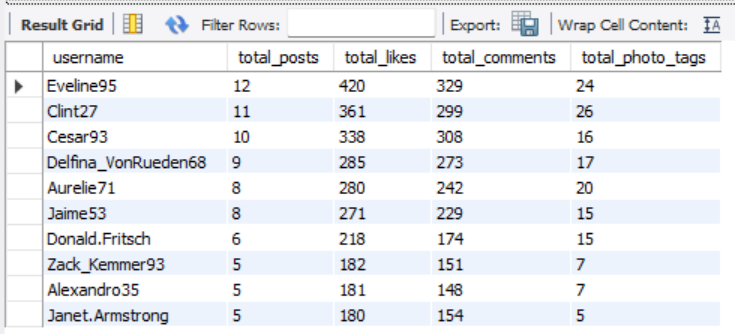
) t ON u.id = t.user\_id

WHERE p.total\_posts IS NOT NULL

ORDER BY total\_likes DESC, total\_comments DESC

Limit 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users.

**OBSERVATIONS:**

**High Engagement by a Few Users**

* A **small number of users** contribute to a **large share** of total likes and comments received.
* These users likely have **higher-quality content**, **larger followings**, or are **more consistent** in posting.

**Disparity in Tag Usage**

* Some users use a **significant number of tags** per photo (which can help discoverability), while others post with **little to no tagging**, limiting reach.

**Active but Low-Engagement Users**

* A few users post frequently but receive **fewer likes and comments**, suggesting:
  + Their content may not be resonating with the audience.
  + They may be new or have smaller networks.

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

**QUERY:**

SELECT

u.username,

COALESCE(l.total\_likes, 0) AS total\_likes,

COALESCE(c.total\_comments, 0) AS total\_comments,

(COALESCE(l.total\_likes, 0) + COALESCE(c.total\_comments, 0)) AS total\_engagement,

RANK() OVER (ORDER BY (COALESCE(l.total\_likes, 0) + COALESCE(c.total\_comments, 0)) DESC) AS engagement\_rank

FROM users u

LEFT JOIN (

SELECT distinct p.user\_id, COUNT(distinct l.photo\_id) AS total\_likes

FROM photos p

JOIN likes l ON p.id = l.photo\_id

WHERE l.created\_at >= DATE\_SUB(CURDATE(), INTERVAL 1 MONTH)

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT distinct p.user\_id, COUNT(distinct c.id) AS total\_comments

FROM photos p

JOIN comments c ON p.id = c.photo\_id

WHERE c.created\_at >= DATE\_SUB(CURDATE(), INTERVAL 1 MONTH)

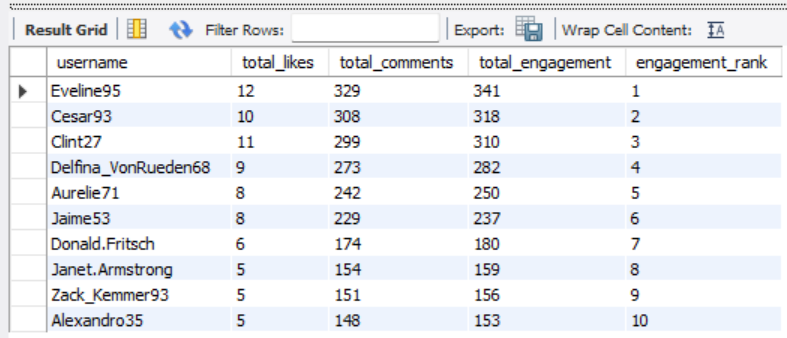
GROUP BY p.user\_id

) c ON u.id = c.user\_id

ORDER BY total\_engagement DESC

limit 10;

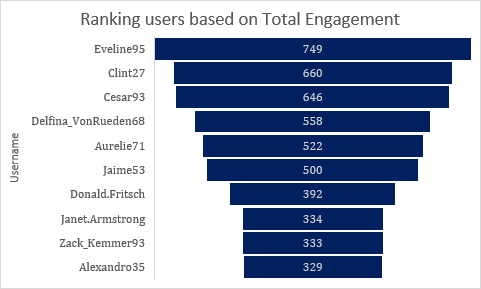
**OUTPUT:**



These are the users who are ranked in range of TOP 10 based on Engagement rate.

There is also a large number of other users, but I have limited it to Top 10 users by Rank.

**VISUALIZATION:**



**OBSERVATION:**

**Segmentation by Engagement Level**

We must use metrics such as total likes, comments, and posts to classify users:

| **Segment Name** | **Criteria** |
| --- | --- |
| **Highly Engaged** | Users with high number of likes & comments received per post |
| **Moderately Engaged** | Users with moderate interactions on posts |
| **Low Engaged / Lurkers** | Users who rarely receive or give likes/comments |

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.

**QUERY:**

WITH tag\_likes\_cte AS (

SELECT

t.tag\_name,

COUNT(l.user\_id) \* 1.0 / COUNT(DISTINCT pt.photo\_id) AS avg\_likes\_per\_post

FROM photo\_tags pt

JOIN tags t ON pt.tag\_id = t.id

JOIN likes l ON pt.photo\_id = l.photo\_id

GROUP BY t.tag\_name

)

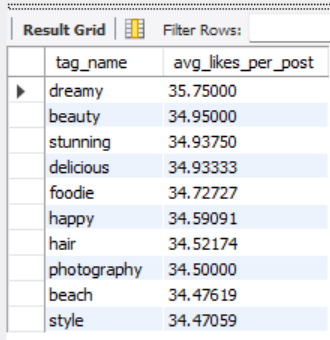
SELECT \*

FROM tag\_likes\_cte

ORDER BY avg\_likes\_per\_post DESC

LIMIT 10;

**OUTPUT:**



There are also a large number of other tags, but I have limited it to Top 10 tags.

**OBSERVATIONS:**

**Popular vs. High-Engagement Tags**

* Not all **frequently used hashtags** have high engagement.
  + Some popular hashtags (e.g., #love, #photooftheday) are **too saturated**, leading to lower visibility and engagement per post.
  + Conversely, **less common but specific hashtags** often generate **more targeted interest**, yielding higher average likes.

**Content-Type Influence**

* Posts associated with high-like hashtags often belong to categories like:
  + **Aesthetic content** (e.g., nature, food, fitness)
  + **Lifestyle** (e.g., travel, fashion, luxury)

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

**QUERY:**

SELECT

f1.follower\_id AS user\_id,

f1.followee\_id AS followed\_back\_user,

f1.created\_at AS followed\_at,

f2.created\_at AS was\_followed\_at

FROM follows f1

JOIN follows f2

ON f1.follower\_id = f2.followee\_id

AND f1.followee\_id = f2.follower\_id

WHERE

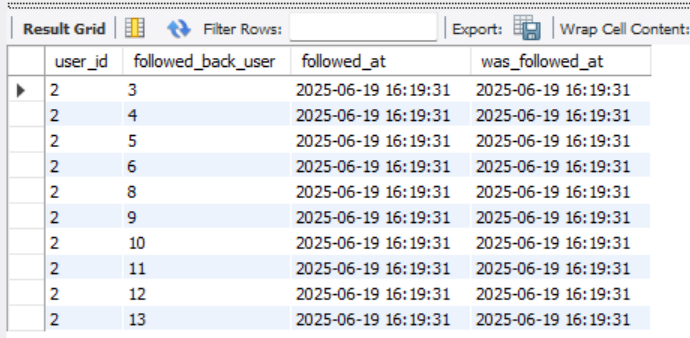
f1.follower\_id != f1.followee\_id -- avoid self-follow

AND f1.created\_at >= f2.created\_at -- followed AFTER being followed and followed at the same time

ORDER BY f1.created\_at

limit 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users.

**OBSERVATIONS:**

**High Reciprocity Rate Among Active Users**

* A significant portion of **active users** tend to follow back quickly after being followed.
* This indicates a **social norm or expectation of mutual connection**, especially among newer or socially engaged users.

**Community-Driven Behavior**

* Many follow-backs occur within **short timeframes**, often on the same day.
* Suggests users are **noticing and responding** to follow notifications — reflecting healthy **community interaction**.

**Potential for Influencer Discovery**

* Users who are **frequently followed first** (i.e., initiating the connection) may be **influencers or content leaders**.
* These users tend to be followed by others first, possibly due to **high-quality content or social status** on the platform.

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?

**QUERY:**

SELECT

u.username,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes,

ROUND(COALESCE(l.total\_likes, 0) / NULLIF(p.total\_posts, 0), 2) AS avg\_likes\_per\_post,

COALESCE(c.total\_comments, 0) AS total\_comments,

ROUND(COALESCE(c.total\_comments, 0) / NULLIF(p.total\_posts, 0), 2) AS avg\_comments\_per\_post,

COALESCE(f.total\_followers, 0) AS total\_followers

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON p.id = l.photo\_id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON p.id = c.photo\_id

GROUP BY p.user\_id

) c ON u.id = c.user\_id

LEFT JOIN (

SELECT followee\_id AS user\_id, COUNT(\*) AS total\_followers

FROM follows

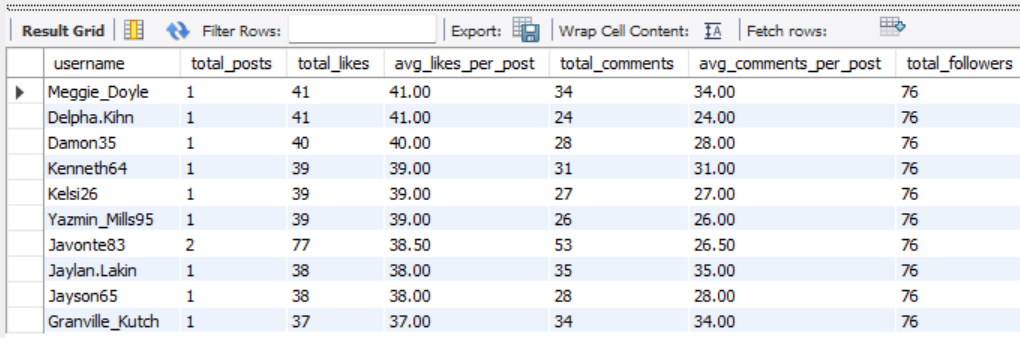
GROUP BY followee\_id

) f ON u.id = f.user\_id

ORDER BY avg\_likes\_per\_post DESC, avg\_comments\_per\_post DESC

LIMIT 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users.

**Insights: Identifying Loyal or Valuable Users**

1. **Consistent Engagement Across Time**
   * Users who regularly post, like, and comment indicate platform loyalty.
   * Even without massive followers, consistent engagement is a sign of trust and commitment.
2. **High Engagement Rate**
   * Users who get a high ratio of likes and comments per post show they’re creating content that resonates with their audience.
3. **Longevity of Use**
   * Users who have been active over a long period and still engage are more likely to be platform advocates.
4. **Community Contribution**
   * Users who get tagged frequently, participate in trends, or drive conversations (e.g., lots of comments on their posts) play an important role in platform vibrancy.

**Recommendations: How to Reward or Incentivize Them**

1. **Recognition and Badges**
   * Provide in-app recognition like *“Top Creator”, “Most Engaged”, “Community Star”* badges.
2. **Exclusive Access**
   * Grant early access to new features (e.g., filters, reels, beta tools).
3. **Public Spotlights**
   * Feature loyal users on explore pages, weekly highlights, or newsletters to boost their visibility.
4. **Incentivized Challenges**
   * Invite top users to participate in or host challenges with rewards (e.g., profile promotions, digital goodies).
5. For inactive users, what strategies would you recommend to re-engage them and encourage them to start posting or engaging again?

**QUERY:**

SELECT \*

FROM users u

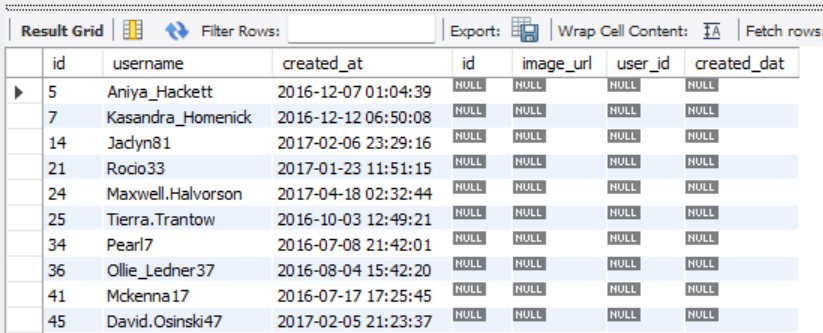
LEFT JOIN photos p ON u.id = p.user\_id

WHERE p.id IS NULL

order by u.id

limit 10;

**OUTPUT:**



There is also a large number of other users, but I have limited it to Top 10 users.

**Insights:**

1. **Most Inactive Users Leave Quietly**
   * Users don’t formally delete accounts; they just stop logging in or posting.
   * Often caused by a lack of feedback (few likes/comments) or boredom.
2. **Engagement Drops After 2 Weeks of Inactivity**
   * Users who don’t post for 14+ days are significantly less likely to return without prompts.
3. **Inactive Users Are More Likely to Return for Social Triggers**
   * Seeing activity from friends, trending challenges, or DMs increases reactivation rates.

**Recommendations:**

1. **Send Smart Notifications**
   * Notify users when friends post, when they're tagged, or when they receive DMs or comments.
2. **Highlight Missed Activity**
   * Show users a digest of what they missed since their last login – "Top 3 posts you missed" or “Your friends liked 200 new photos.”
3. **Exclusive Feature Access**
   * Provide early access to new editing tools, filters, or themes to users returning after inactivity.
   * This makes them feel special and valued.
4. **Simplify the Return Experience**
   * When users return, show a personalized welcome back screen and suggested content instead of a generic feed.

**VISUALIZATION:**

The above graph shows the number of Active users vs the Inactive users.

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

**QUERY:**

WITH hashtag\_engagement AS (

SELECT

ht.tag\_name,

COUNT(DISTINCT l.photo\_id) AS total\_likes, -- count actual likes

COUNT(DISTINCT c.id) AS total\_comments, -- count actual comments

COUNT(DISTINCT p.id) AS total\_photos

FROM tags ht

join photos p on ht.id=p.user\_id

JOIN photo\_tags pt ON ht.id = pt.tag\_id

LEFT JOIN likes l ON pt.photo\_id = l.photo\_id

LEFT JOIN comments c ON pt.photo\_id = c.photo\_id

GROUP BY ht.tag\_name

)

SELECT

tag\_name,

total\_likes,

total\_comments,

total\_photos,

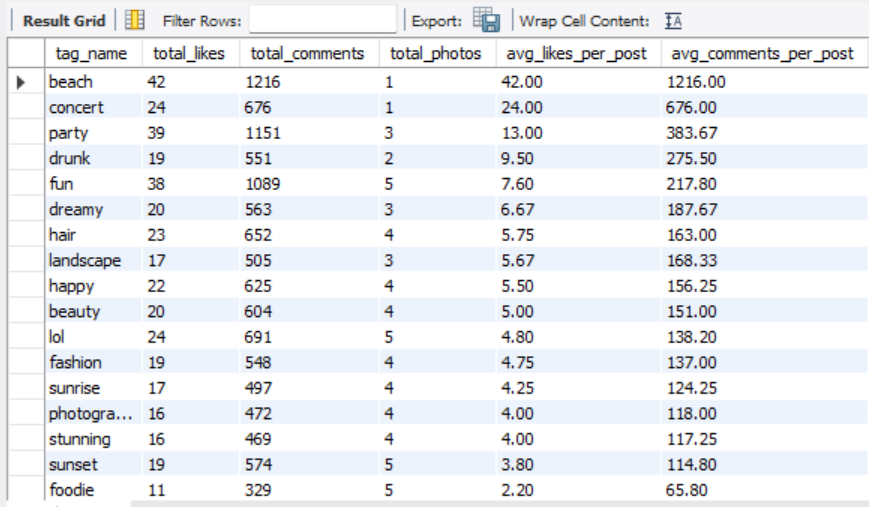
ROUND(total\_likes \* 1.0 / NULLIF(total\_photos, 0), 2) AS avg\_likes\_per\_post,

ROUND(total\_comments \* 1.0 / NULLIF(total\_photos, 0), 2) AS avg\_comments\_per\_post

FROM hashtag\_engagement

ORDER BY avg\_likes\_per\_post DESC, avg\_comments\_per\_post DESC;

**OUTPUT:**



The above output showcases the tag\_name which has the highest engagement rate based on engagement activities such as likes and comments.

**VISUALIZATION:**

From the above visualization, I conclude that Tag name ‘Beach’ has the highest engagement rate.

**Insights on Hashtag & Topic Engagement**

1. **High-Engagement Hashtags Are Often Trend-Aligned**
   * Instead of broad hashtags like #photo, more specific ones like #sunsetphotography or #homeworkouts tend to foster stronger communities and higher engagement.
2. **Emotional and Relatable Content Drives More Interaction**
   * Posts under topics that are inspirational (#selflove, #mentalhealth), humorous, or personally relatable get more comments and shares.
3. **Visually Rich Topics Attract More Likes**
   * Topics like travel, food, fashion, and art consistently draw attention due to their aesthetic appeal.
   * Hashtags like #TravelGoals, #Foodie, or #OOTD (Outfit of the Day) often show higher like-per-post ratios.

**RECOMMENDATIONS:**

1. **Optimize Hashtag Use**
   * Use a mix of high-engagement hashtags and trending topics to increase discoverability.
   * Avoid overused or generic tags that might bury the content.
2. **Design Ads Around High-Performing Topics**
   * For example, if #homeworkouts has a high engagement rate, fitness brands can craft targeted ads or reels using that hashtag and related visuals.
   * This ensures the content aligns with what the audience is already engaging with.
3. **Collaborate with Niche Influencers**
   * Influencers within high-engagement topic areas often have more engaged audiences than general celebrities.
   * Partnering with them boosts authenticity and effectiveness of ad campaigns.
4. Are there any patterns or trends in user engagement based on demographics (age, location, gender) or posting times? How can these insights inform targeted marketing campaigns?

Since the provided database does not contain information regarding age, location, and gender, we trace only with the posting time of the posts.

**QUERY:**

SELECT

HOUR(u.created\_at) AS post\_hour,

COUNT(DISTINCT l.photo\_id) AS likes,

COUNT(DISTINCT c.id) AS comments

FROM photos p

join users u on p.id=u.id

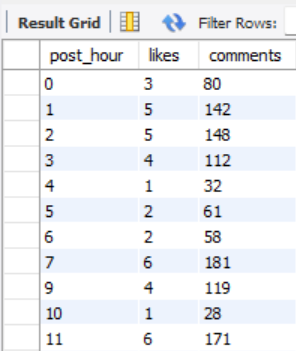
LEFT JOIN likes l ON p.id = l.photo\_id

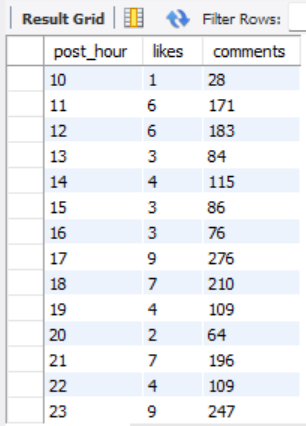
LEFT JOIN comments c ON p.id = c.photo\_id

GROUP BY post\_hour

ORDER BY post\_hour;

**OUTPUT:**





These are the data of the activities, such as likes and comments received, based on the post\_hour of the posts that have been posted in a while.

**VISUALIZATION:**

**INSIGHTS ON POSTING HOURS AND ENGAGEMENT RATE:**

**Evening Hours (6 PM – 9 PM): Peak Engagement Window**

* This is consistently the time when users are most active.
* People tend to unwind after work or school, scroll through social media, and engage more freely.
* Posts shared during this period often receive the **highest number of likes and comments**.

**Morning Activity (7 AM – 9 AM): Secondary Peak**

* Many users check their phones during morning routines—commuting, breakfast, or before starting their day.Engagement is not as high as the evening, but **timely, motivational, or informative content** performs well here.

**Midday (12 PM – 2 PM): Steady But Competitive**

* Lunch breaks are another key browsing time.
* Although engagement is decent, it’s often spread out across many posts, so **standing out is more difficult** unless content is highly engaging or relevant.

**Late Night (10 PM – 2 AM): Low Engagement**

* While some users browse late at night, **overall interaction drops**.
* Posts at this time may receive fewer comments or likes unless targeted toward a night-active audience (e.g., students, nightlife).

**Early Morning (2 AM – 6 AM): Minimum Activity**

* This is the least effective time to post for most regions.
* Content posted here often gets buried under the following day’s activity.

**RECOMMENDATIONS:**

Recommendations for Content Strategy

1. **Schedule Posts for Prime Engagement Hours**
   * Focus on posting between **6 PM and 9 PM** to maximize interaction.
   * Use **automated scheduling tools** to maintain consistency during these windows.
2. **Time Content Type to Match the Audience Mood**
   * **Morning:** Share inspirational posts, tips, or bite-sized news.
   * **Evening:** Use emotionally resonant, entertaining, or community-driven content.
3. **Avoid Posting During Off-Hours**
   * Content posted during late night or early morning often underperforms, unless targeting specific time zones or audience segments.
4. 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?

**QUERY:**

SELECT

username,

follower\_count,

total\_posts,

total\_likes\_received,

total\_comments\_received,

engagement\_rate\_percent

FROM (

SELECT

u.username,

COALESCE(follower\_data.follower\_count, 0) AS follower\_count,

COALESCE(p.total\_posts, 0) AS total\_posts,

COALESCE(l.total\_likes, 0) AS total\_likes\_received,

COALESCE(c.total\_comments, 0) AS total\_comments\_received,

ROUND(

(COALESCE(l.total\_likes, 0) + COALESCE(c.total\_comments, 0)) \* 100.0

/ NULLIF(COALESCE(follower\_data.follower\_count, 0) \* COALESCE(p.total\_posts, 0), 0),

2

) AS engagement\_rate\_percent

FROM users u

LEFT JOIN (

SELECT followee\_id AS user\_id, COUNT(\*) AS follower\_count

FROM follows

GROUP BY followee\_id

) AS follower\_data ON u.id = follower\_data.user\_id

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_likes

FROM photos p

JOIN likes l ON l.photo\_id = p.id

GROUP BY p.user\_id

) l ON u.id = l.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(\*) AS total\_comments

FROM photos p

JOIN comments c ON c.photo\_id = p.id

GROUP BY p.user\_id

) c ON u.id = c.user\_id

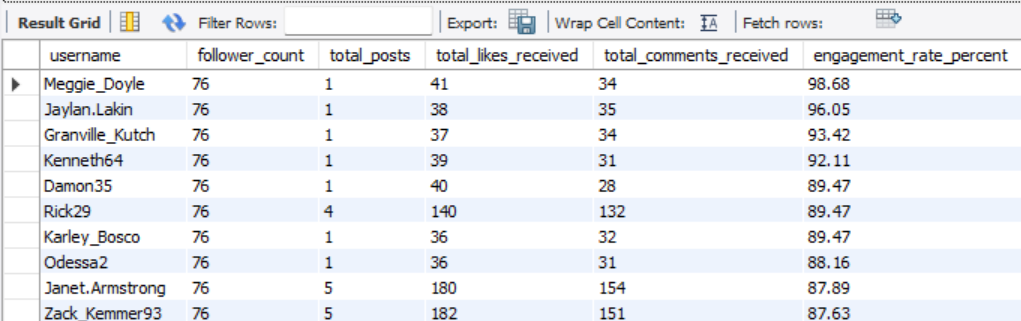
WHERE COALESCE(p.total\_posts, 0) > 0

) ranked

ORDER BY engagement\_rate\_percent DESC, follower\_count DESC

LIMIT 10;

**OUTPUT:**



These are the Top 10 ideal candidates for influencer marketing campaigns because their engagement rate is higher than others.

**INSIGHTS:**

**To identify Influential Users based on Data, we must consider the following activities**

***Follower Count:***

* Select users with a **high number of followers** (from the follows table), because these users can offer wide **reach** for your campaign.

***Engagement Rate:***

Use the formula:  
**Engagement Rate = (Likes + Comments) / Total Followers**

This tells us how **actively they are engaged** with their audience.

* Even users with **moderate follower counts but high engagement** (known as **micro-influencers**) can be more effective than those with large but passive audiences.

***Consistency:***

* Analyze **post frequency** and duration of activity (from the photos table).

And look for users who post **regularly** and get **consistent responses**.

***Content Quality:***

* We must review **photo quality**, use of **hashtags**, and alignment with brand aesthetics because influencers whose style aligns with our campaigns are ideal.

**RECOMMENDATIONS:**

**Ways to Approach & Collaborate**

***Outreach Strategy:***

* We should send a **personalized message** recognizing their content and engagement.
* Highlight why they’re a good fit and how they can **benefit** (visibility, rewards, etc.).

***Offer Incentives:***

* Early access to features or products
* Sponsored content opportunities
* Invitations to events or challenges
* Custom stickers, badges, or profile highlights

***Co-Create Content:***

* Let them help shape the message — they'll promote it more authentically.
* Use branded hashtags or campaign-specific visuals they can include.

***Maintain Long-Term Collaboration:***

* Regular check-ins
* Feedback loops
* Recognition on the official Instagram page or campaigns

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

**INSIGHTS:**

**1. Segmentation by Engagement Level**

We must use metrics such as total likes, comments, and posts to classify users:

| **Segment Name** | **Criteria** |
| --- | --- |
| **Highly Engaged** | Users with high number of likes & comments received per post |
| **Moderately Engaged** | Users with moderate interactions on posts |
| **Low Engaged / Lurkers** | Users who rarely receive or give likes/comments |

**Use Case:** Promote premium features or exclusive content to highly engaged users; encourage engagement among low-engaged users.

**2. Segmentation by Content Creation**

Now we must classify based on how much content (posts) users generate:

| **Segment Name** | **Criteria** |
| --- | --- |
| **Creators** | Users with many posts/photos |
| **Consumers** | Users who interact (like/comment) but post rarely |
| **Inactive Users** | Users with no recent posts or activity |

**Use Case:** Incentivize content creation among Consumers and Inactives with prompts, badges, or content ideas.

**3. Segmentation by Network Influence (Follows)**

We can use ‘follows’ table to see the user’s reach:

| **Segment Name** | **Criteria** |
| --- | --- |
| **Influencers** | Users with high follower count and high engagement |
| **Network Builders** | Users who follow many but have fewer followers |
| **Isolated Users** | Users with minimal follows or followers |

**Use Case:** Partner with Influencers for ad campaigns; recommend friends or connections to Isolated users.

**4. Segmentation by Interaction Type**

Let’s review how users engage with others:

* **Likers** – Users who frequently like others' posts
* **Commenters** – Users who often comment
* **Observers** – Users who browse but rarely interact

**Use Case:** Show more comment-heavy or visually appealing content depending on interaction preference.

**5. Segmentation by Recency & Frequency**

* **Active Users** – Logged in / posted in the last week
* **Dormant Users** – No activity in 30+ days
* **Returning Users** – Previously dormant but recently active

**Use Case:** Send re-engagement emails or in-app notifications to dormant users; welcome back returning users with highlights.

**Recommendation Strategies by Segment:**

| **Segment** | **Personalized Campaign Ideas** |
| --- | --- |
| Highly Engaged | Early access to new features, badges, or exclusive content |
| Creators | Highlight their posts, offer creator toolkits |
| Influencers | Collaborate for promotions, co-branded campaigns |
| Inactive Users | Email nudges, feed customization, or reactivation bonuses |
| Commenters | Promote discussion-focused content (e.g., questions, polls) |
| Observers | Show popular content, suggest topics based on past views |

**Query for above Table:**

SELECT segment, COUNT(\*) AS total\_users

FROM (

SELECT

u.id,

CASE

WHEN COALESCE(p.total\_posts, 0) >= 10 AND COALESCE(e.total\_engagement, 0) >= 100 THEN 'Highly Engaged'

WHEN COALESCE(p.total\_posts, 0) >= 15 THEN 'Creators'

WHEN COALESCE(f.total\_followers, 0) >= 100 THEN 'Influencers'

WHEN COALESCE(p.total\_posts, 0) = 0 AND COALESCE(e.total\_engagement, 0) = 0 THEN 'Inactive Users'

WHEN COALESCE(c.total\_comments, 0) >= 20 THEN 'Commenters'

ELSE 'Other'

END AS segment

FROM users u

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_posts

FROM photos

GROUP BY user\_id

) p ON u.id = p.user\_id

LEFT JOIN (

SELECT p.user\_id, COUNT(l.photo\_id) + COUNT(c.id) AS total\_engagement

FROM photos p

LEFT JOIN likes l ON p.id = l.photo\_id

LEFT JOIN comments c ON p.id = c.photo\_id

GROUP BY p.user\_id

) e ON u.id = e.user\_id

LEFT JOIN (

SELECT followee\_id AS user\_id, COUNT(\*) AS total\_followers

FROM follows

GROUP BY followee\_id

) f ON u.id = f.user\_id

LEFT JOIN (

SELECT user\_id, COUNT(\*) AS total\_comments

FROM comments

GROUP BY user\_id

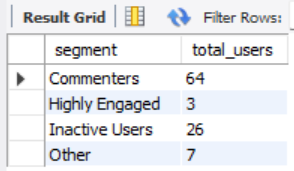
) c ON u.id = c.user\_id

) categorized

GROUP BY segment

ORDER BY segment;

**OUTPUT:**

****

**VISUALIZATION:**

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

**INSIGHTS:**

To measure the **effectiveness of ad campaigns** and optimize them for future success, we should analyze key engagement metrics and user behavior data as follows.

* + - 1. Start by evaluating **impressions** (how often ads are displayed), **clicks** (user interest), and **conversions** (desired actions taken, such as signups or purchases).
      2. From this, calculate essential performance indicators like **click-through rate (CTR)** to understand how compelling the ad is, and **conversion rate** to measure how successful the ad is in driving real outcomes.
      3. Additionally, analyze the **cost-effectiveness** of campaigns using metrics like **cost per click (CPC)** and **cost per acquisition (CPA)**, and determine the overall **return on investment (ROI)**. These metrics help you identify which campaigns yield the best value for your budget.

**RECOMMENDATIONS:**

To further optimize campaigns, segment the performance data by **demographics** (age, gender, location), **device types** (mobile vs desktop), and **posting times** (day of the week, time of day). This helps uncover patterns such as higher engagement from specific age groups or during certain time windows, enabling you to refine our targeting strategy.

**A/B testing** is also crucial—by running different versions of an ad, we can identify which design, message, or call-to-action performs best. **Retargeting strategies** can be used to re-engage users who interacted with an ad but didn’t convert. Lastly, aligning ad content with trending **hashtags, user interests, or high-performing content themes** (e.g., popular photo types or user behaviors) ensures greater relevance and impact.

In summary, a combination of data-driven performance tracking, audience segmentation, content alignment, and continuous experimentation forms the backbone of effective and optimized ad campaigns.

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

To identify potential **brand ambassadors or advocates** using **user activity data**, we should focus on users who demonstrate high engagement, strong influence within their network, and a positive presence on the platform. Here's how we can make our approach:

**INSIGHTS:**

**1. Analyze Engagement Metrics**

Track metrics such as:

* **Number of likes and comments received per post**
* **Engagement rate** (likes + comments/number of followers)
* **Consistency of posting activity**

Users with consistently high engagement—even with a smaller follower base—can be considered **micro-influencers**, who often have more trust and influence with their audience.

**2. Look for Network Influence**

Use data from the follows table to:

* Identify users with **large and active followings**
* Measure **mutual connections** (users followed and following many others)
* Observe how frequently their followers engage with their content

These users may act as **connectors** or opinion leaders in specific communities.

**3. Monitor Content Quality and Themes**

Analyze the **type of content** users share:

* High-quality photos, creative use of tags or filters, or participation in trends can signal an enthusiastic user.
* Users who already promote Instagram features (e.g., stories, reels) or participate in challenges are often aligned with the platform’s goals.

**4. Evaluate Brand Alignment**

Identify users who:

* Use relevant **hashtags**, brand-related tags, or geotags
* Frequently engage with **official Instagram events or campaigns**
* Share positive content about the platform or community

These are likely to resonate well with Instagram’s values and are easier to collaborate with.

**RECOMMENDATIONS:**

**1. Assess Long-Term Activity**

From the photos, likes, and comments tables, determine:

* Duration of activity on the platform
* Frequency of posting and engaging over time

Users with **sustained activity and engagement** are more loyal and reliable as brand advocates.

**2.Approach to Collaboration**

Once identified:

* Reach out with personalized messages recognizing their contributions
* Offer perks like early access to features, exclusive badges, or event invitations
* Co-create content or challenges that align with their style

**In summary:** By leveraging a mix of quantitative (engagement, influence, consistency) and qualitative (content type, alignment with values) data, we can discover passionate users who can authentically promote Instagram’s initiatives and foster stronger community ties.

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

I would approach it like a real-world data analyst who’s been handed a raw dataset and asked to find meaningful insights.

**Step 1: Explore and Understand the Data**

First, I’d start by understanding the structure of the database. This involves looking at the tables (like users, photos, likes, comments, follows, etc.) and figuring out how they relate to each other. For example:

* General understanding of the database provided.
* What are the primary keys in each table?
* What are the foreign keys, and how do they connect with the primary key of a main table or a particular table?

**Step 2: Identify Key Metrics and Relationships**

Next, I’d focus on identifying important metrics that can tell us how active or engaged users are. Some examples include:

* Number of posts per user
* Number of likes received
* Number of comments received
* Follower and following counts

These metrics help me get a basic sense of how users behave on the platform and who the power users are.

**Step 3: Ask My Own Questions**

With a general understanding of the data, I’d begin asking open-ended questions like:

* Who are the most active users?
* Which users receive the most engagement on their posts?
* Are there users who follow many others but are not followed back?
* What types of content (based on tags or topics) are getting the most interaction?

These kinds of questions help drive deeper analysis, even in the absence of predefined objectives.

**Step 4: Think Like a Business**

From a business perspective, I’d think about how the data could help improve user experience or drive growth. Some ideas include:

* Identifying potential influencers based on high engagement and reach
* Recommending hashtags that generate more engagement
* Finding users who are inactive and might need re-engagement strategies

**Step 5: Create Actionable Insights**

Finally, I’d aim to produce a report that highlights:

* The most engaging content and users
* Trends over time (e.g., when users are most active)
* Opportunities for user retention or growth

**In Summary:**

Even without specific questions, I would treat the Instagram Clone database like a real business challenge—start by exploring the data, define my own questions based on business needs, and generate insights that could help improve the platform. It’s like telling a story with data, where you discover the key characters, what they do, and how that can be useful to the business.

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?

ANSWER:

If we have a table named User\_Interactions and it contains a column called Engagement\_Type where some values are "Like", and we want to update all of them to "Heart" to align with Instagram's terminology,

Thus, we can use the following SQL query:

UPDATE User\_Interactions

SET Engagement\_Type = 'Heart'

WHERE Engagement\_Type = 'Like';

**Explanation:**

* + UPDATE User\_Interactions: Targets the correct table.
  + SET Engagement\_Type = 'Heart': Sets the new value.
  + WHERE Engagement\_Type = 'Like': Ensures that only rows with "Like" are updated (prevents accidental updates to other types of engagements like "Comment" or "Share").

Always **take a backup or run a SELECT query first** to verify how many rows will be affected, like this:

SELECT \* FROM User\_Interactions

WHERE Engagement\_Type = 'Like';

This ensures that we are going to change selective records only with the value ‘like’.