**Social Media Analysis**

Objective Questions

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

**Below are queries used to find if there are any NULL values in the given tables**.

* **For Comments Table:**

**SELECT comment\_text, user\_id, photo\_id, created\_at**

**FROM comments**

**where comment\_text is Null**

**and user\_id is null**

**and photo\_id is null**

**and created\_at is null**

**GROUP BY 1,2,3,4**

* **For Follows Table:**

**select follower\_id, followee\_id, created\_at**

**from follows**

**where follower\_id is NULL**

**and followee\_id is NULL**

**and created\_at is NULL**

**group by 1,2,3**

* **For Likes Table:**

**select user\_id, photo\_id, created\_at**

**from likes**

**where user\_id is NULL**

**or photo\_id is NULL**

**or created\_at is NULL**

**group by 1,2,3**

* **For Photo\_Tags table:**

**select photo\_id, tag\_id**

**from photo\_tags**

**where photo\_id is NULL**

**and tag\_id is NULL**

**Group by 1,2**

* **For Photos table:**

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

**from photos**

**where id is NULL**

**and image\_url is NULL**

**and user\_id is NULL**

**and created\_dat is NULL**

**Group by 1,2,3,4**

* **For Tags table;**

**select id, tag\_name, created\_at**

**from tags**

**where id is NULL**

**and tag\_name is NULL**

**and created\_at is NULL**

Group by 1,2,3

* **For Users table;**

**select id, username, created\_at**

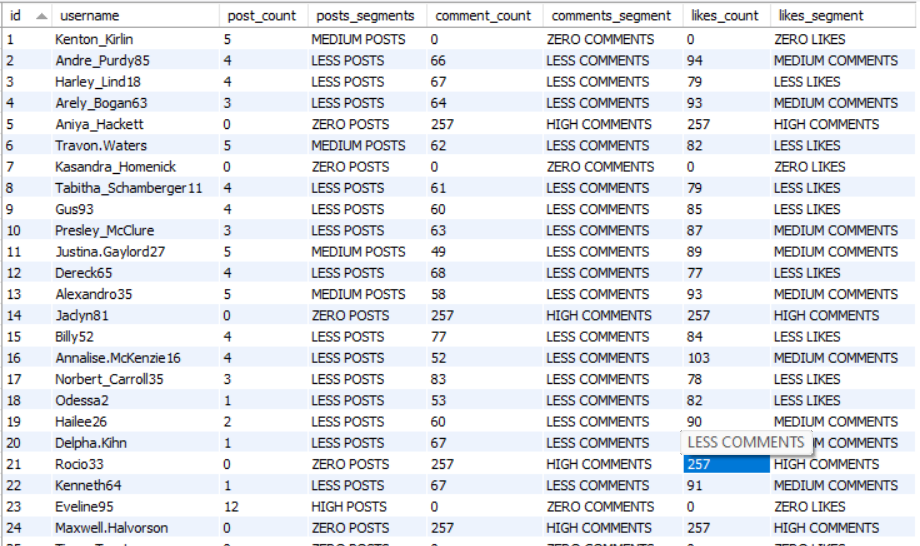
**from users**

**where id is NULL**

**and username is NULL**

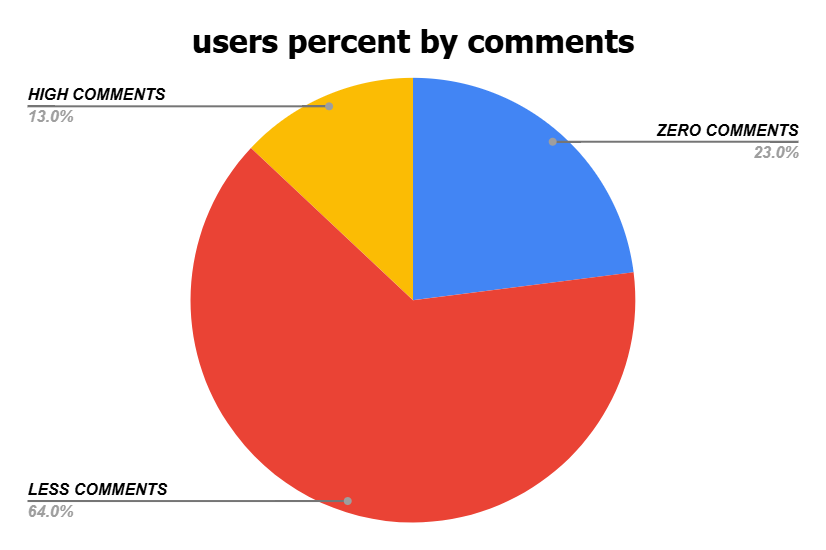
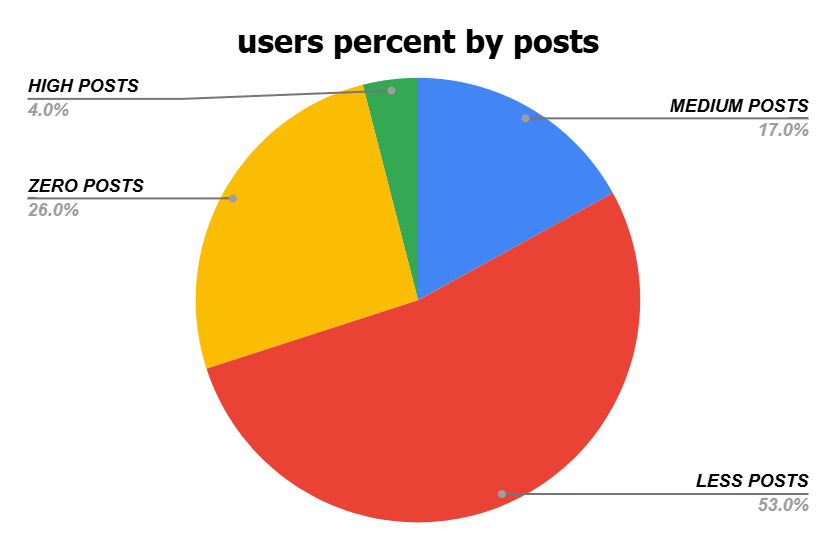
**and created\_at is NULL**

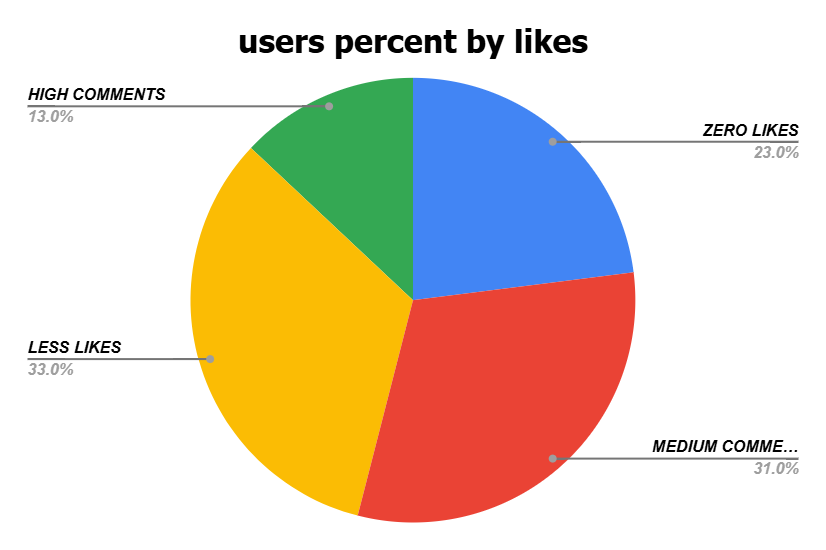
1. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?



Breakdown of segments by percentage:

* Posts Segment: Less Posts 53%, Medium Posts 17%, High Posts 4.0%, Zero Posts 26%
* Comments Segments: Less Comments 64%, Zero Comments 23%, High Comments 13%
* Likes Segment: Less Likes 33%, Medium Likes 31%, Zero Likes 23%, High Likes 13%





Below is the query to find the above data

**select u.id, u.username, count(distinct p.id) as post\_count,**

**case**

**when count(distinct p.id) = 0 then 'ZERO POSTS'**

**when count(distinct p.id)<= max(count(distinct p.id)) over() / 3 then 'LESS POSTS'**

**when count(distinct p.id)<=2\* max(count(distinct p.id)) over() / 3 then 'MEDIUM POSTS'**

**else 'HIGH POSTS'**

**end as posts\_segments,**

**count(distinct c.id) comment\_count,**

**case**

**when count(distinct c.id) = 0 then 'ZERO COMMENTS'**

**when count(distinct c.id)<= max(count(distinct c.id)) over() / 3 then 'LESS COMMENTS'**

**when count(distinct c.id)<=2\* max(count(distinct c.id)) over() / 3 then 'MEDIUM COMMENTS'**

**else 'HIGH COMMENTS'**

**end as comments\_segment,**

**count(distinct l.photo\_id) as likes\_count,**

**case**

**when count(distinct l.photo\_id) = 0 then 'ZERO LIKES'**

**when count(distinct l.photo\_id)<= max(count(distinct l.photo\_id)) over() / 3 then 'LESS LIKES'**

**when count(distinct l.photo\_id)<=2\* max(count(distinct l.photo\_id)) over() / 3 then 'MEDIUM COMMENTS'**

**else 'HIGH COMMENTS'**

**end as likes\_segment**

**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 1, 2;**

Concepts used for the above query:

1. COUNT() with DISTINCT - To count the distinct (unique) number of posts, comments and likes
2. CASE Expression for Categorization - To classify users based on their count of posts, comments and likes
3. Window Functions (i.e., OVER() and MAX() - To calculate the maximum count of posts, comments, or likes across all users
4. LEFT JOIN - To join the tables
5. Calculate the average number of tags per post (photo\_tags and photos tables).



Below is the query used to get the above value:

**with cte as (select p.id as photo\_id,**

**count(pt.tag\_id) as no\_of\_tags**

**from photos p**

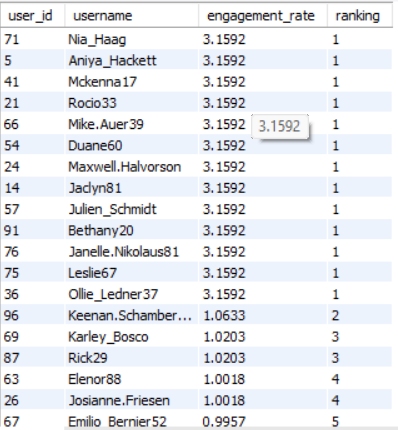
**left join photo\_tags pt on p.id = pt.photo\_id**

**group by p.id)**

**select round(avg(no\_of\_tags),2) as avg\_no\_of\_tags\_per\_post**

**from cte**

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



**WITH cte AS (**

**SELECT u.id AS user\_id, u.username,**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) AS engagement,**

**(COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) \* 100 / SUM((COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id))) OVER() AS engagement\_rate,**

**DENSE\_RANK() OVER(ORDER BY (COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) DESC) AS highest\_engagements**

**FROM users u**

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

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

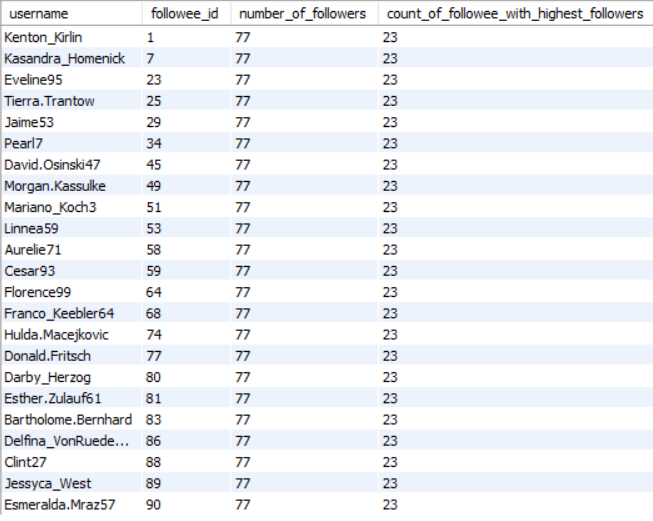
**LEFT JOIN comments c ON u.id = c.user\_id**

**GROUP BY 1**

**)**

**SELECT user\_id, username, engagement\_rate, highest\_engagements AS ranking FROM cte**

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



Above data gives us the details about the users with the highest number of followers.

Below is the query used to get the above output:

**WITH follower\_counts AS (**

**SELECT**

**followee\_id,**

**COUNT(follower\_id) AS number\_of\_followers**

**FROM follows**

**GROUP BY followee\_id**

**),**

**max\_followers AS (**

**SELECT MAX(number\_of\_followers) AS max\_followers**

**FROM follower\_counts**

**)**

**SELECT**

**u.username,**

**fc.followee\_id,**

**fc.number\_of\_followers,**

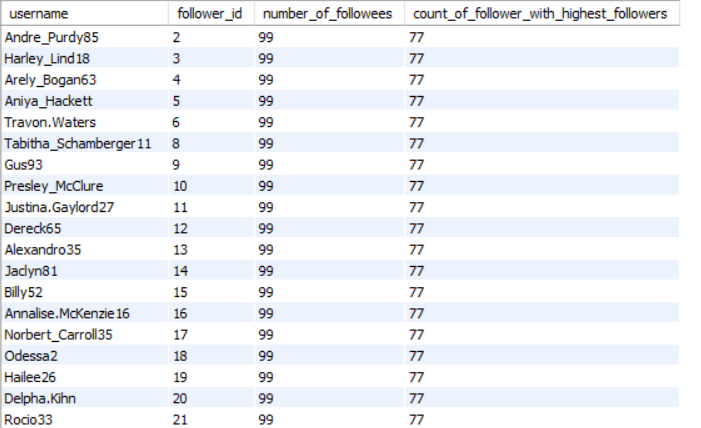
**COUNT(\*) OVER () AS count\_of\_followee\_with\_highest\_followers**

**FROM follower\_counts fc**

**JOIN users u ON fc.followee\_id = u.id**

**JOIN max\_followers mf ON fc.number\_of\_followers = mf.max\_followers**

**ORDER BY fc.followee\_id;**



Above data gives us the details about users with the highest number of followees.

Below is the query used to get the above data:

**WITH follower\_counts AS (**

**SELECT**

**follower\_id,**

**COUNT(followee\_id) AS number\_of\_followees**

**FROM follows**

**GROUP BY follower\_id**

**),**

**max\_followers AS (**

**SELECT MAX(number\_of\_followees) AS max\_followees**

**FROM follower\_counts**

**)**

**SELECT**

**u.username,**

**fc.follower\_id,**

**fc.number\_of\_followees,**

**(SELECT COUNT(\*)**

**FROM follower\_counts fc2**

**WHERE fc2.number\_of\_followees = fc.number\_of\_followees) AS count\_of\_follower\_with\_highest\_followers**

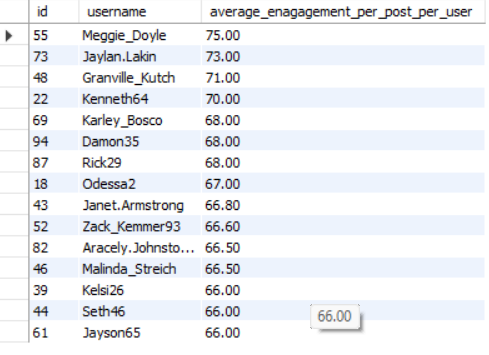
**FROM follower\_counts fc**

**JOIN users u ON fc.follower\_id = u.id**

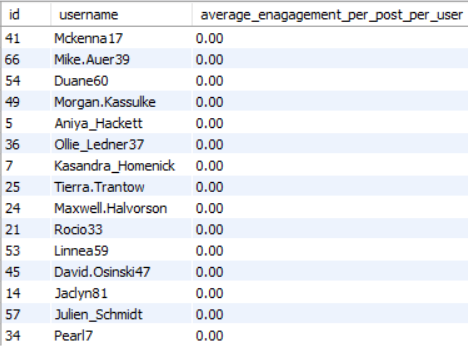
**JOIN max\_followers mf ON fc.number\_of\_followees = mf.max\_followees**

**ORDER BY fc.follower\_id;**

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



Above table gives us the data of average engagement rate per post per user of top 15 users



Above table gives us the data of average engagement rate per post per user of bottom 15 users

Below is the query to get the above data:

**with cte as (select u.id, u.username, p.id as post\_id,**

**count(distinct l.user\_id) + count(distinct c.id) as total\_engagement**

**from users u**

**left join photos p on u.id = p.user\_id**

**left join likes l on p.id = l.photo\_id**

**left join comments c on p.id = c.photo\_id**

**group by 1,2,3**

**)**

**select id, username,**

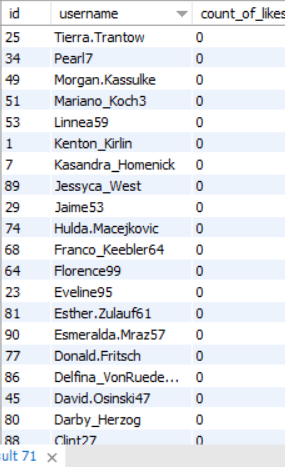
**round(avg(total\_engagement),2) as average\_enagagement\_per\_post\_per\_user**

**from cte**

**group by id,username**

**order by average\_enagagement\_per\_post\_per\_user asc limit 15;**

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



Above table gives us the data about users who have never liked any posts i.e., 23 users.

Below is the query used to find the above data:

**select u.id, u.username, count(distinct l.photo\_id) as count\_of\_likes**

**from users u left join likes l**

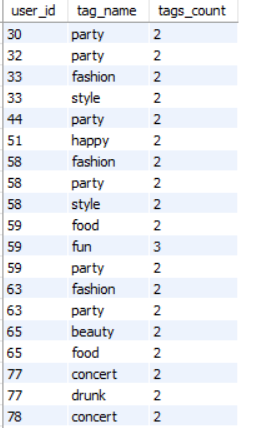
**on u.id = l.user\_id**

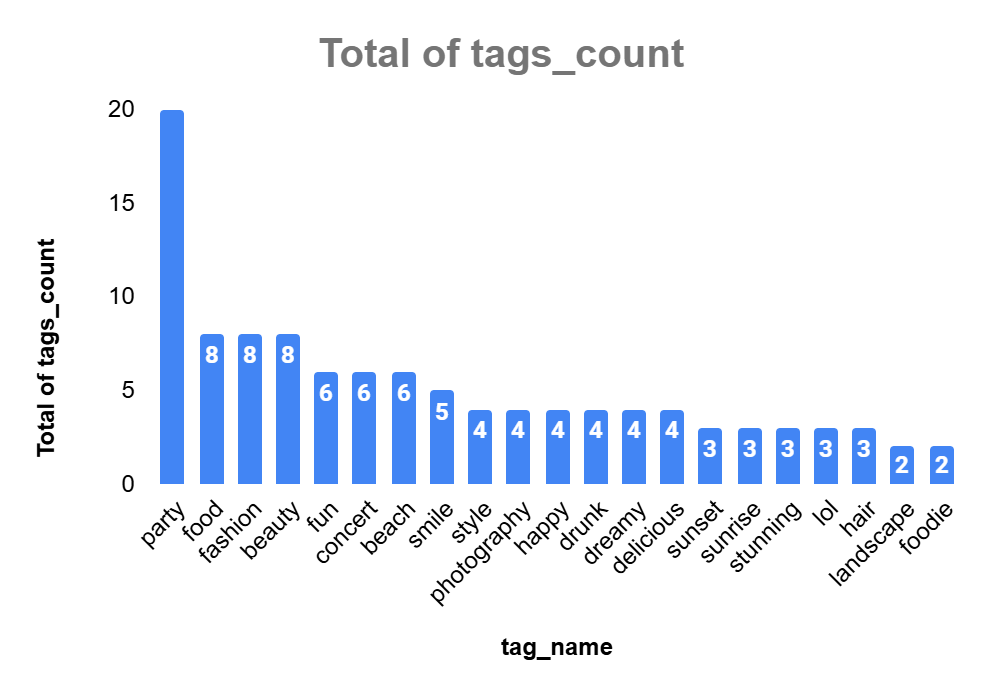
**group by 1,2**

**having count(distinct l.photo\_id) = 0**

**order by id, username asc;**

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





Above data shows us the details about total tag count used by the users.

From the above chart we can observe that:

* **party** is the most used tag by users
* **foodie** is the least used tag by users
* **beauty, food, fashion, smile, beach, concert** and **fun** are also some of the other tags which were used more by the users

Below are some suggestions to create more personalized and engaging ad campaigns.

* Our data reveals that "**Party**" is the most popular tag among users, we can use vibrant imagery and dynamic videos showcasing epic parties, dance floors, and celebrations. This campaign can target users who frequently engage with party-related content, promoting event tickets, party accessories, or venues.
* Both "**food**" and "**foodie**" appear in the data, with "food" being more popular.A campaign featuring mouth-watering dishes, recipes, and dining experiences can resonate well. Collaborate with popular food influencers to create content that showcases delicious meals, cooking tutorials, and restaurant reviews.
* Tags like "**fashion**" and "**beauty**" are equally popular, indicating a keen interest in personal style and aesthetics. An ad campaign focused on the latest fashion trends, beauty tips, and style guides can captivate this audience

Below is the query written to get the above output and insights:

**WITH tag\_counts AS (**

**SELECT**

**u.id AS user\_id,**

**t.tag\_name,**

**COUNT(t.tag\_name) AS tags\_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, t.tag\_name**

**),**

**max\_tag\_counts AS (**

**SELECT**

**tag\_name,**

**MAX(tags\_count) AS max\_count**

**FROM tag\_counts**

**GROUP BY tag\_name**

**)**

**SELECT**

**tc.user\_id,**

**tc.tag\_name,**

**tc.tags\_count**

**FROM tag\_counts tc**

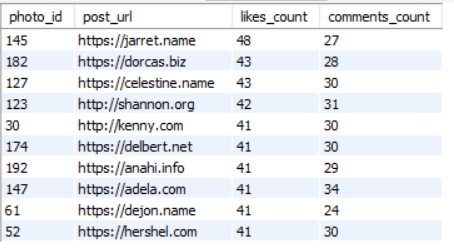
**JOIN max\_tag\_counts mtc**

**ON tc.tag\_name = mtc.tag\_name**

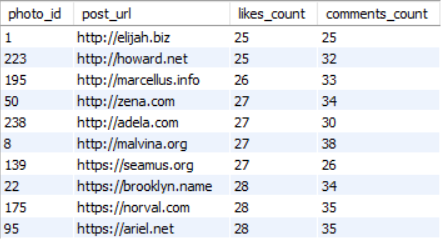
**AND tc.tags\_count = mtc.max\_count**

**ORDER BY tc.user\_id, tc.tag\_name;**

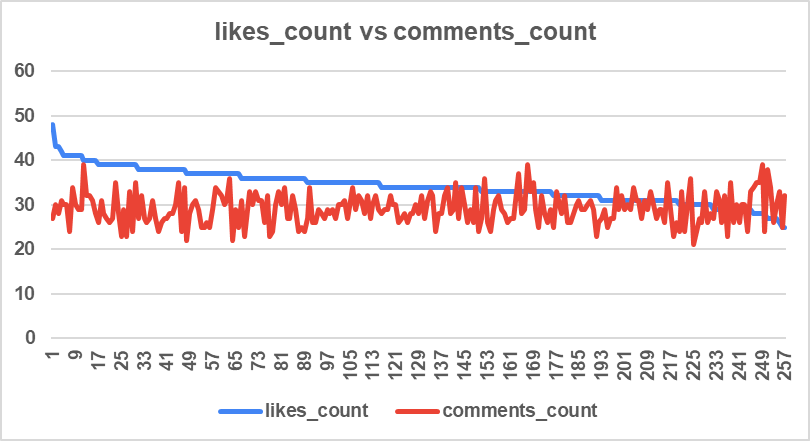
1. Are there any correlations between user activity levels and specific content types (e.g., photos, videos, reels)? How can this information guide content creation and curation strategies?



Since we only have the data of photos, the above table gives us the data about the top 10 posts with the highest likes.



Above table gives us the data about the bottom 10 posts with lowest likes count.



Above chart represents the correlation between likes and comments on the posts. As we can observe, posts with the highest likes have a relatively high number of comments.

Understanding that likes and comments go hand-in-hand, we can refine our content strategies in three ways:

* **Create Engaging Content:** Focus on crafting posts that are visually appealing, emotionally engaging, or thought-provoking. Interactive posts, like polls, questions, or challenges, can encourage users to not only like but also comment, sparking conversations.
* **Curate Comment-Worthy Topics:** Encourage more interaction by addressing topics that invite discussion. Controversial topics, trending news, or heartfelt stories often prompt users to share their opinions or experiences in the comments section. This can foster a sense of community and keep users coming back to engage with your posts.
* **Monitor and Adapt:** Track which types of posts receive the most likes and comments. Analyze the themes, formats, and timing of these successful posts to replicate and refine your strategy.

Below is the query written to generate the above data:

**select p.id as photo\_id, p.image\_url as post\_url,**

**count(distinct l.user\_id) as likes\_count,**

**count(distinct c.id) as comments\_count**

**from photos p**

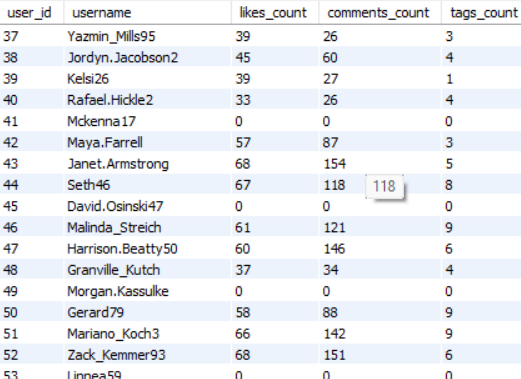
**left join likes l on p.id = l.photo\_id**

**left join comments c on p.id = c.photo\_id**

**group by 1**

**order by 3 desc;**

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



The above table represents the data about the total number of likes, comments, and photo tags for each user.

Below is the query written to get the above data:

**WITH user\_activity AS (**

**SELECT u.id AS user\_id,**

**u.username,**

**COUNT(DISTINCT l.user\_id) AS likes\_count,**

**COUNT(DISTINCT pt.tag\_id) AS tags\_count**

**FROM users u**

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

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

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

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

**GROUP BY**

**u.id,**

**u.username**

**)**

**SELECT user\_id, username,**

**SUM(likes\_count) AS likes\_count,**

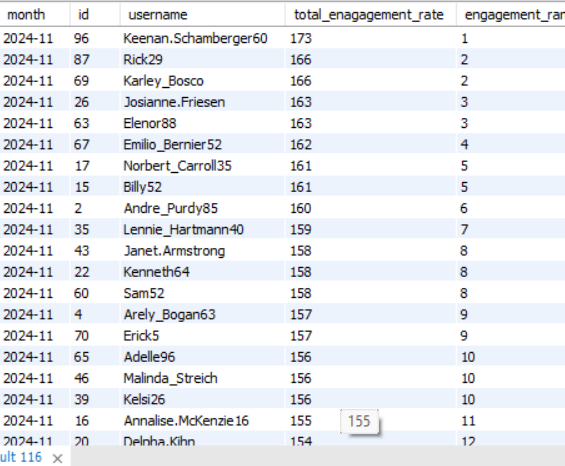
**SUM(comments\_count) AS comments\_count,**

**SUM(tags\_count) AS tags\_count**

**FROM user\_activity**

**GROUP BY user\_id, username;**

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



Above table gives us the rank by engagement rate, total engagement rate is calculated by SUM of distinct likes and distinct comments.

Below is the query written to get the above output:

**select date\_format(p.created\_dat, '%Y-%m') as month, u.id, u.username,**

**(count(distinct l.photo\_id) + count(distinct c.id)) as total\_enagagement\_rate,**

**dense\_rank() over(partition by date\_format(p.created\_dat, '%Y-%m') order by**

**(count(distinct l.photo\_id) + count(distinct c.id)) desc) as engagement\_rank**

**from users u**

**left join photos p on u.id = p.user\_id**

**left join likes l on u.id = l.user\_id and p.created\_dat = l.created\_at**

**left join comments c on u.id = c.user\_id and p.created\_dat = c.created\_at**

**where date\_format(p.created\_dat, '%Y-%m') is not null**

**group by 1,2,3**

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.



Above are the hashtags that have been used in posts with the highest average number of likes.

Query written to get the output is:

**with highest\_avg\_likes AS (**

**select post\_id**

**from (**

**select p.id as post\_id, avg\_likes, rank() over(order by avg\_likes DESC) AS ranking**

**from photos p**

**left join (**

**select photo\_id, count(user\_id) / (select COUNT(distinct user\_id) from photos) as avg\_likes**

**from likes**

**group by 1**

**) l on p.id = l.photo\_id**

**) dt**

**where ranking = 1**

**)**

**select al.post\_id, group\_concat(t.tag\_name) as hashtags**

**from highest\_avg\_likes al**

**join photo\_tags pt on al.post\_id = pt.photo\_id**

**join tags t on pt.tag\_id = t.id**

**group by 1;**

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



There are no one who started following someone after being followed by that person

Query written to get the output is:

**select f1.follower\_id as followed\_back, f1.followee\_id as followee**

**from follows f1**

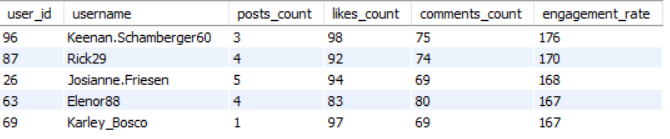
**join follows f2 on f1.follower\_id = f2.followee\_id**

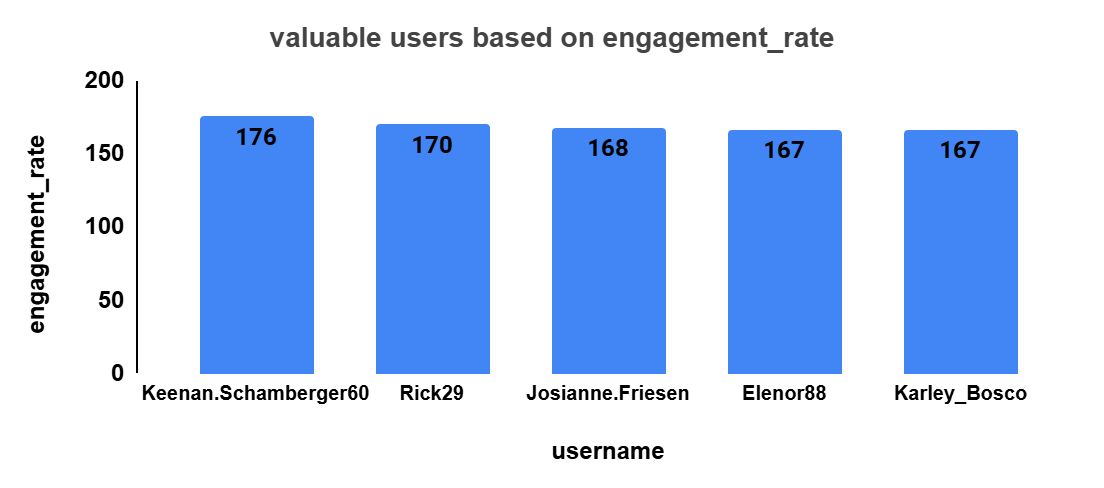
**and f1.followee\_id = f2.followee\_id and**

**f1.created\_at>f2.created\_at**

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?





The above users will be considered as most loyal and valuable based on their overall engagement rate i.e., posts, likes and comments and overall engagement rate.

I would suggest the below ideas to reward or incentivize the above users;

**Personalized Shoutouts:** Imagine their surprise when they open Instagram to see a personalized shoutout from our brand, celebrating their amazing engagement. We’ll feature their profile, highlighting their unique style and top posts.

**Special Invitations:** We’ll invite them to exclusive virtual events, like behind-the-scenes tours of our latest projects, Q&A sessions with influencers, or sneak peeks of new products. These VIP events will make them feel special and valued.

**Custom Gifts:** Each of our top users will receive a customized gift package. It could include branded merchandise, a handwritten thank-you note, and a special coupon for their favorite items. These thoughtful touches will strengthen their connection to our brand.

**Early Access:** Grant them early access to new features or content. Whether it’s a new app feature, a special product line, or an upcoming event, being the first to experience it will make them feel like insiders.

**Engagement Challenges:**Launch exclusive engagement challenges where they can earn points and badges. For instance, the more they interact with our content, the more rewards they can unlock, like additional discounts, special recognition.

Below is the query written to get the above insights:

**with user\_activity as (select u.id as user\_id,u.username,**

**count(distinct p.id) as posts\_count,**

**count(distinct l.photo\_id) as likes\_count,**

**count(distinct c.id) as comments\_count,**

**count(distinct p.id) + count(distinct l.photo\_id) + count(distinct c.id) as engagement\_rate,**

**dense\_rank() over(order by count(distinct p.id) + count(distinct l.photo\_id) + count(distinct c.id) desc) user\_rank**

**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**

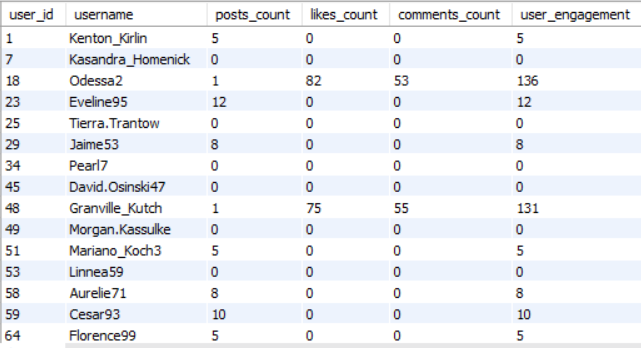
**)**

**select user\_id, username, posts\_count, likes\_count, comments\_count, engagement\_rate**

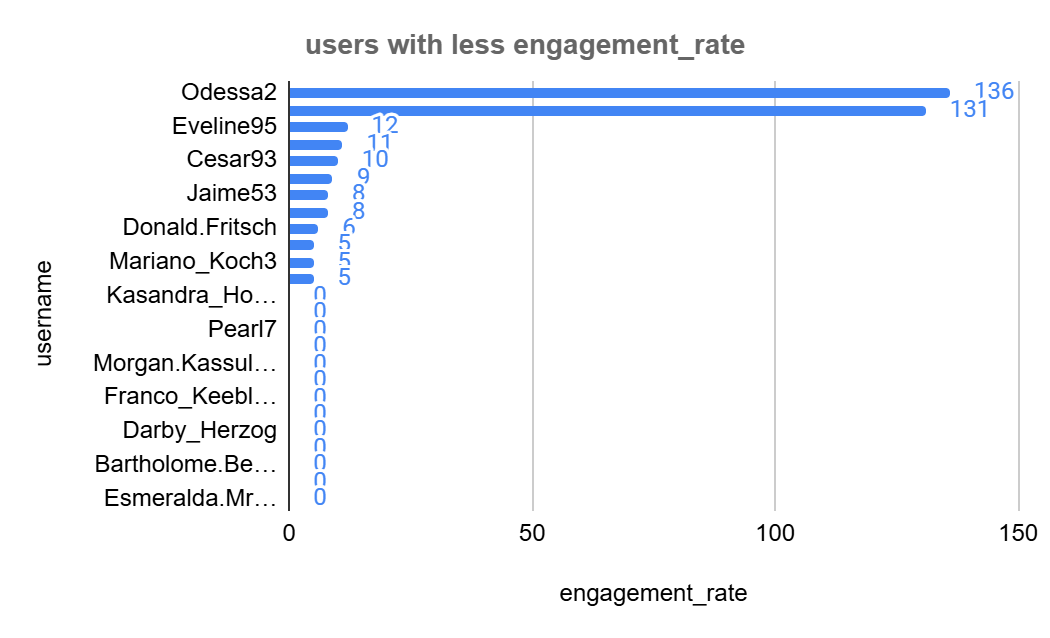
**from user\_activity**

**where user\_rank between 1 and 5 and posts\_count >0;**

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



Above table represents users who are inactive based on their count of posts, likes, comments and overall engagement.



Below are the few strategies I would recommend to reawaken the inactive users:

**Personalized Invitations**: A personalized message saying, “We miss your amazing posts, [Username]! Here’s a little something to welcome you back.” Accompanied by a special badge or a limited-time feature, this makes users feel valued and missed.

**Flashback Features:** Highlight their past posts with a “Remember This?” feature. Showcasing their most-liked posts or favorite memories encourages nostalgia and reminds them of the joy they once found in sharing their moments.

**Engagement Challenges:** Launch a fun challenge exclusively for inactive users. For example, a “7-Day Comeback Challenge” where each day has a new theme or prompt. Completing each day’s challenge earns them points, rewards, or a chance to be featured on Instagram’s official page

**Learning and Growth:** Offer resources or mini-courses on improving their Instagram game. Topics could include photography tips, editing hacks, or storytelling techniques. Empowering them with new skills can reignite their interest and confidence in posting.

Below is the query written to get the above insights;

**WITH cte AS (**

**SELECT u.id AS user\_id, u.username,**

**COUNT(DISTINCT p.id) AS posts\_count,**

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

**COUNT(DISTINCT c.id) AS comments\_count,**

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

**DENSE\_RANK() OVER(ORDER BY COUNT(DISTINCT p.id) + COUNT(DISTINCT l.photo\_id) + COUNT(DISTINCT c.id)) AS drank**

**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 1, 2**

**)**

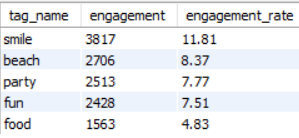
**SELECT user\_id, username, posts\_count, likes\_count, comments\_count, user\_engagement**

**FROM cte**

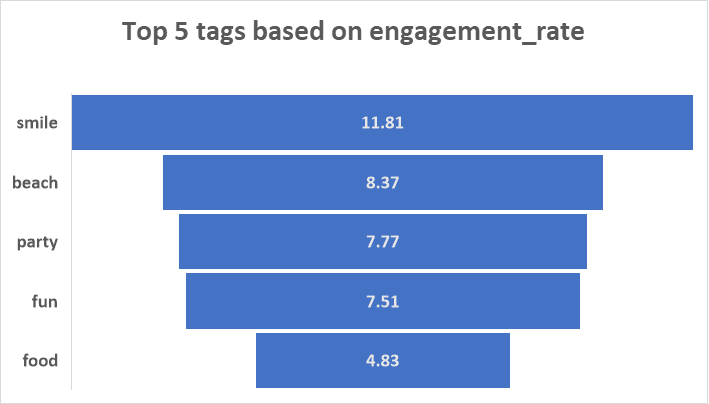
**WHERE drank BETWEEN 1 AND 10**

**ORDER BY 1;**

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



The above data represents Top 5 tags with highest engagement\_rates.



Based on the above insights I would recommend the below strategies to guide content strategy and ad campaigns

**Spread the Joy with #Smile**

**Content Strategy:** Create uplifting and positive posts that capture genuine smiles and happy moments. Use #Smile to boost visibility and engagement

**Ad Campaign:** Promote products or services that bring joy, like wellness programs, holiday packages, or happiness quotes. Highlight how they make life brighter.

**Embrace the Adventure with #Beach**

**Content Strategy:** Post stunning beach photos, travel guides, and relaxing seascapes. Use #Beach to attract travel enthusiasts and nature lovers.

**Ad Campaign:** Advertise beach resorts, travel gear, or summer clothing. Emphasize the serene and adventurous aspects of beach vacations.

**Celebrate with #Party**

**Content Strategy:** Share vibrant party scenes, event highlights, and celebration tips. Use #Party to engage users looking for fun and excitement.

**Ad Campaign:** Promote event planning services, party supplies, or nightlife venues. Showcase how your brand can make parties unforgettable.

**Capture Fun Moments with #Fun**

**Content Strategy:** Post entertaining and light-hearted content, from funny videos to playful activities. Use #Fun to attract a wide audience seeking amusement.

**Ad Campaign:** Advertise games, amusement parks, or humor-related products. Highlight the fun and enjoyment they bring.

**Satisfy Cravings with #Food**

**Content Strategy:** Share delicious recipes, food photography, and culinary tips. Use #Food to engage food enthusiasts and chefs.

**Ad Campaign:** Promote restaurants, cooking classes, or food products. Emphasize the taste and quality of your offerings.

**Below is the query written to generate the above insights:**

**with likes\_count as (**

**select t.tag\_name,count(l.user\_id) as likes\_count**

**from tags t**

**left join photo\_tags pt on t.id = pt.tag\_id**

**left join likes l on pt.photo\_id = l.photo\_id**

**group by t.tag\_name**

**),**

**posts\_count as (**

**select t.tag\_name,count(p.id) as posts\_count**

**from tags t**

**left join photo\_tags pt on t.id = pt.tag\_id**

**left join photos p on pt.photo\_id = p.id**

**group by t.tag\_name**

**),**

**comments\_count as (**

**select t.tag\_name, count(c.id) as comments\_count**

**from tags t**

**left join photo\_tags pt on t.id = pt.tag\_id**

**left join comments c on pt.photo\_id = c.photo\_id**

**group by t.tag\_name**

**)**

**select cl.tag\_name,**

**cl.likes\_count + cp.posts\_count + cc.comments\_count as engagement,**

**round((cl.likes\_count + cp.posts\_count + cc.comments\_count) \* 100 /**

**sum(cl.likes\_count + cp.posts\_count + cc.comments\_count) over(), 2) as engagement\_rate**

**from likes\_count cl**

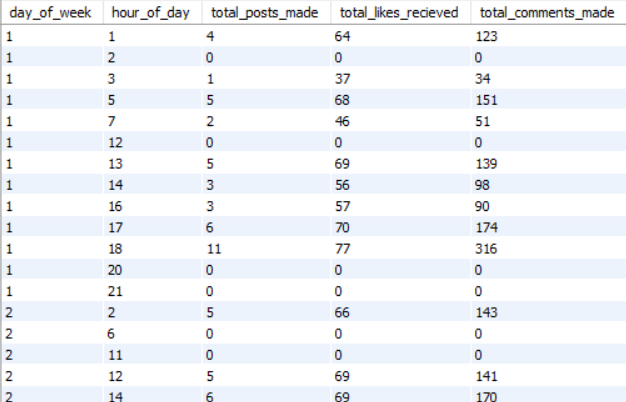
**join posts\_count cp ON cl.tag\_name = cp.tag\_name**

**join comments\_count cc ON cl.tag\_name = cc.tag\_name**

**order by 2 desc**

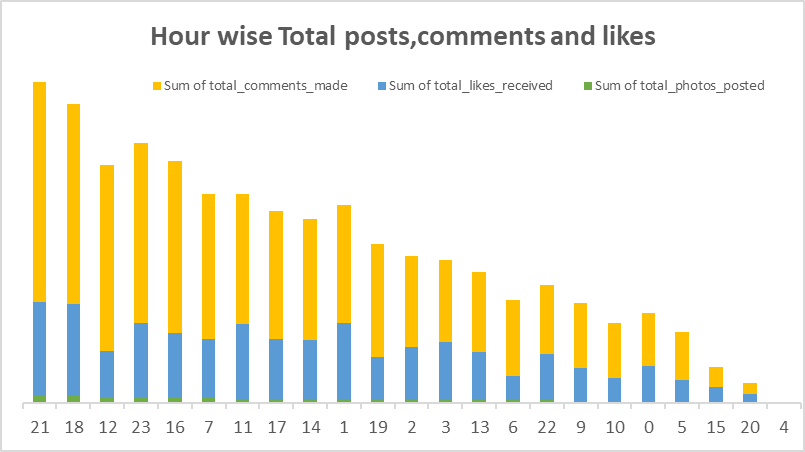
**limit 5;**

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?



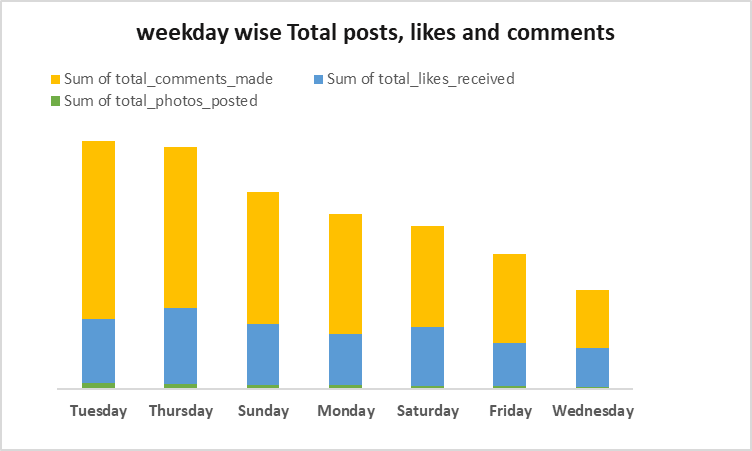
The above data represents user engagement patterns (photos posted, likes received, and comments made) by day of the week and hour of the day.

High Engagement Times:



* Peak engagement times for photos, likes, and comments are in the late morning (11 AM to 12 PM) and evening (4 PM to 7 PM, 9 PM to 11 PM).

Day of week Trends:



* Bursts of activity on Tuesdays and Thursdays, with steady engagement on Sundays and Mondays. Engagement dips on Fridays and Wednesdays
* Strong engagement from 1AM to 3AM
* **Marketing Strategies:**
* **Afternoon and Evening Blitz:**

Imagine your audience scrolling through their feeds on a leisurely Tuesday, Thursday, or Sunday afternoon and evening. That’s when engagement is at its peak. Schedule your key campaigns during these golden hours to capture their attention and drive maximum interaction.

* **Late-Night Delight:**

On early weekdays and late Wednesdays, people are winding down, but their phones are still buzzing. Utilize these late-night windows for promotions and interactive content. Think of exciting late-night deals, interactive polls, or fun quizzes to keep them engaged and coming back for more.

* **Weekend Wonders:**

Weekends, particularly Saturdays and Sundays, are prime time for brand awareness. People are relaxed, exploring new things, and more open to discovering your brand. Launch targeted ads and awareness campaigns to reach them when they are most receptive.

Below are the queries written to generate the above insights;

**select**

**weekday(p.created\_dat) as day\_of\_week,**

**extract(hour from p.created\_dat) as hour\_of\_day,**

**count(distinct p.id) as total\_posts\_made,**

**count(distinct l.user\_id) as total\_likes\_recieved,**

**count(distinct c.id) as total\_comments\_made**

**from photos p**

**left join likes l on p.id = l.photo\_id**

**left join comments c on p.id = c.photo\_id**

**where extract(hour from p.created\_dat) is not null**

**group by day\_of\_week, hour\_of\_day**

**order by day\_of\_week, hour\_of\_day;**



**select**

**dayofweek(u.created\_at) as day\_of\_week,**

**extract(hour from u.created\_at) as hour\_of\_day,**

**count(distinct p.id) as total\_posts\_made,**

**count(distinct l.user\_id) as total\_likes\_recieved,**

**count(distinct c.id) as total\_comments\_made**

**from users u**

**left join photos p on u.id = p.user\_id**

**left join likes l on p.id = l.photo\_id**

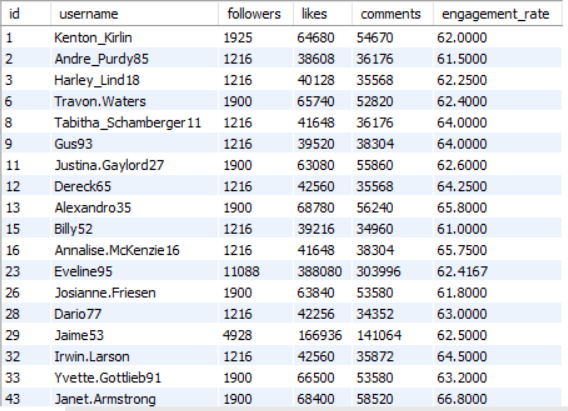
**left join comments c on p.id = c.photo\_id**

**where extract(hour from u.created\_at) is not null**

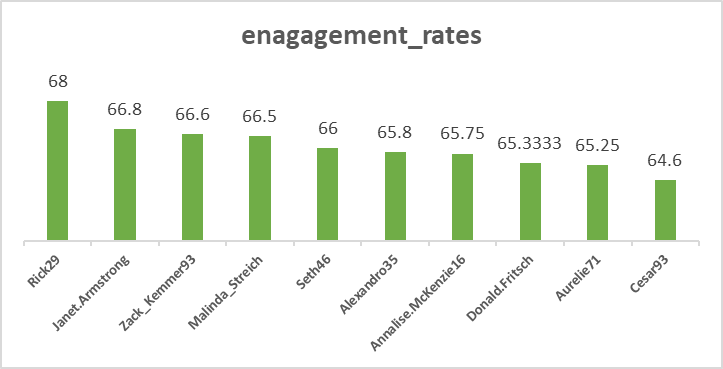
**group by 1,2**

**order by 1,2;**

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?



The above data represents the users with overall followers, likes, comments and engagement rates.



* Above insights are calculated by summing up the total number of likes and comments on their posts and dividing it by their total number of followers
* I filtered out users with less than 1000 followers and an engagement rate below 5% to ensure only users with a significant following and high engagement rates are considered for influencer marketing campaigns.

**Below are few recommendations to approach and collaborate with these influencers:**

**1. Personal Connection**

Start by making a personal connection. Imagine receiving a message that stands out:

“Hey [Influencer's Name], we’ve been huge fans of your content here at Instagram! Your amazing engagement rates and unique style truly inspire us. We’d love to collaborate and create something extraordinary together!”

**2. Highlight the Value**

Show them what’s in it for them. Paint a picture of the mutual benefits:

“By partnering with Instagram, you’ll have access to exclusive features, promotional support, and opportunities to reach an even larger audience. Together, we can amplify your influence and create content that your followers will love.”

**3. Showcase Previous Success**

Share stories of past successful collaborations to build trust:

“We’ve previously worked with influencers like [Example Influencer], and the results were phenomenal! Their engagement soared, and their followers loved the exclusive content we created together. We believe we can achieve similar success with you.”

**4. Offer Creative Freedom**

Respect their creative vision and give them space to innovate:

“We value your unique voice and creativity. Our collaboration will be a blend of your distinctive style and our innovative features. Let’s brainstorm ideas and

create something that truly resonates with your audience.”

**5. Provide Clear Next Steps**

Make it easy for them to say yes by outlining clear steps:

“If you’re interested, let’s set up a quick call to discuss potential ideas and see how we can make this partnership incredible. We can’t wait to explore this opportunity with you!”

**In Summary:**

* **Personal Connection:** Stand out with a heartfelt message.
* **Highlight the Value:** Emphasize mutual benefits.
* **Showcase Success:** Share inspiring stories of past collaborations.
* **Offer Creative Freedom:** Respect and blend their style with your features.
* **Provide Clear Next Steps:** Make the process straightforward and exciting.

**By approaching influencers with genuine enthusiasm, clear value propositions, and respect for their creativity, you can build strong, impactful partnerships that resonate with both their followers and your brand.**

**Below is the query written to generate the above insights:**

**with user\_engagement as (**

**select**

**u.id,**

**u.username,**

**count(f.follower\_id) as followers,**

**sum(l.likes) as likes,**

**sum(c.comments) as comments,**

**(sum(l.likes) + sum(c.comments)) / count(f.follower\_id) as engagement\_rate**

**from**

**users u**

**left join**

**follows f on u.id = f.followee\_id**

**left join**

**(select photo\_id, count(\*) as likes from likes group by photo\_id) l on l.photo\_id in (select id from photos where user\_id = u.id)**

**left join**

**(select photo\_id, count(\*) as comments from comments group by photo\_id) c on c.photo\_id in (select id from photos where user\_id = u.id)**

**group by**

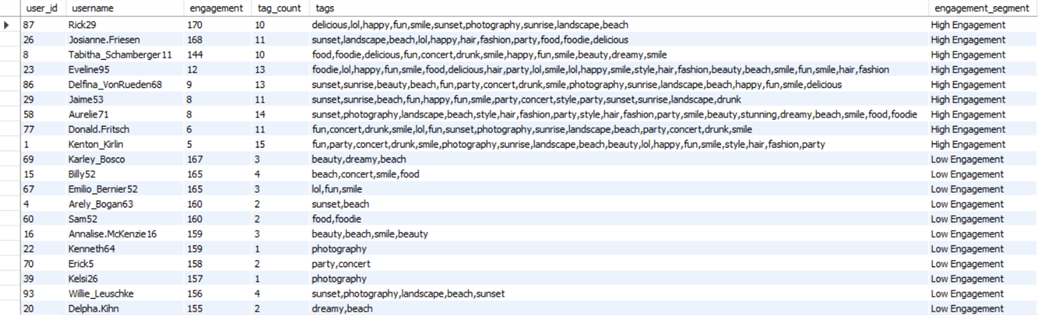
**u.id, u.username**

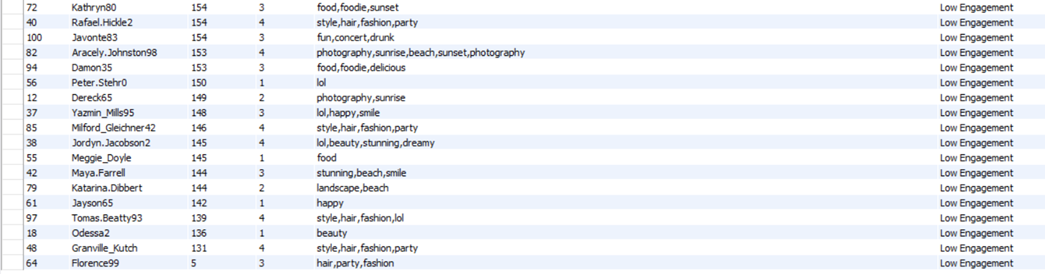
**)**

**select \* from user\_engagement**

**where followers > 1000 and engagement\_rate > 0.05;**

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





Segmentation of users’ base for targeted marketing campaign or personalized recommendation:

**High Engagement Users:**

Meet our High Engagement Users – the life of our social media party. These users interact frequently and use a wide variety of tags. They are perfect candidates for personalized recommendations and exclusive rewards.

**Re-engaging the Inactive Users:**

Now, let's talk about our Low Engagement Users – those who need a little nudge to become more active. These users have minimal interactions and use fewer tags. To re-engage them, we can create targeted re-engagement strategies. Think of special incentives, like personalized discounts or interactive content, designed to spark their interest and encourage more activity.

**Tag-Based Segmentation:**

By leveraging users' tag preferences, we can create even more personalized campaigns. Whether it's food lovers, photography enthusiasts, or fashion aficionados, we can align our content and product recommendations with their interests. This tag-based segmentation ensures that each user sees content that resonates with them, boosting engagement and satisfaction.

**Targeting Strategy:**

* **High Engagement Users:** Offer exclusive access, rewards, and influencer opportunities.
* **Low Engagement Users:** Use personalized campaigns and incentives to boost activity.
* **Tag-Based Segmentation:** Align content with user interests for better engagement

**Marketing Focus:**

Use engagement and tag data to segment users effectively, ensuring relevant content and offers for each group. This strategy maximizes overall engagement and builds a more connected, loyal user base.

**Below is the query written to generate the above insights:**

**WITH user\_engagement AS (**

**SELECT**

**u.id AS user\_id,**

**u.username,**

**COALESCE(p.engagement, 0) + COALESCE(l.engagement, 0) + COALESCE(c.engagement, 0) AS engagement,**

**COALESCE(t.tag\_count, 0) AS tag\_count**

**FROM**

**users u**

**LEFT JOIN (**

**SELECT user\_id, COUNT(DISTINCT id) AS engagement**

**FROM photos**

**GROUP BY user\_id**

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

**LEFT JOIN (**

**SELECT user\_id, COUNT(DISTINCT photo\_id) AS engagement**

**FROM likes**

**GROUP BY user\_id**

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

**LEFT JOIN (**

**SELECT user\_id, COUNT(DISTINCT id) AS engagement**

**FROM comments**

**GROUP BY user\_id**

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

**LEFT JOIN (**

**SELECT u.id AS user\_id,**

**COUNT(DISTINCT t.tag\_name) AS tag\_count**

**FROM**

**users u**

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

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

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

**GROUP BY u.id**

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

**),**

**global\_max AS (**

**SELECT**

**MAX(engagement) AS max\_engagement,**

**MAX(tag\_count) AS max\_tag\_count**

**FROM user\_engagement**

**),**

**user\_tags AS (**

**SELECT**

**u.id AS user\_id,**

**group\_concat(t.tag\_name) AS tags**

**FROM**

**users u**

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

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

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

**GROUP BY u.id**

**),**

**user\_segments AS (**

**SELECT**

**e.user\_id,**

**e.username,**

**e.engagement,**

**e.tag\_count,**

**t.tags,**

**CASE**

**WHEN e.engagement < gm.max\_engagement / 3 AND e.tag\_count < gm.max\_tag\_count / 3 THEN 'Low Engagement'**

**WHEN e.engagement < 2 \* gm.max\_engagement / 3 AND e.tag\_count < 2 \* gm.max\_tag\_count / 3 THEN 'Moderate Engagement'**

**ELSE 'High Engagement'**

**END AS engagement\_segment**

**FROM user\_engagement e**

**LEFT JOIN user\_tags t ON e.user\_id = t.user\_id**

**CROSS JOIN global\_max gm -- Cross join to ensure you can use the global maximums**

**GROUP BY e.user\_id, e.username, e.engagement, e.tag\_count, t.tags, gm.max\_engagement, gm.max\_tag\_count**

**)**

**SELECT \***

**FROM user\_segments**

**WHERE tag\_count > 0 AND tags IS NOT NULL and engagement\_segment in ('High Engagement', 'Low Engagement')**

**ORDER BY engagement\_segment, engagement DESC;**

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

* **Understanding Impressions:**

Think of impressions as the first glance—how often your ad is shown. High impressions mean great visibility, but what’s next?

* **Click-Through Rate (CTR):**

Clicks are like catching the eye of your audience. It’s the moment someone decides to engage with your ad. A high click-through rate (CTR) shows your ad’s message and visuals are resonating well. To calculate CTR:

**CTR (%) = (Total Clicks / Total Impressions) \* 100**

* **Monitoring Conversions (Conversion Rates):**

Conversions are the ultimate goal—be it a sale, signup, or any other desired action. A high conversion rate means your ad effectively leads users to take action. To calculate conversion rate:

**Conversion Rate (%) = (Total Conversions / Total Clicks) \* 100**

* **Continuous Improvement(CPA - Cost per Acquisition)**

Regularly review and adjust your campaigns based on CPA data. Aim to lower your CPA while maintaining or increasing conversion rates. This balance will help you achieve cost-effective and successful marketing efforts.

**CPA = Total Campaign Cost / Number of Conversions**

* **Return on Ad Spend(ROA)**

Constantly strive to improve your ROA by refining your targeting strategies, enhancing ad creatives, and optimizing your ad spend. The goal is to maximize revenue while minimizing costs, ensuring your advertising efforts are both effective and efficient.

**ROA = Revenue from Ads / Total Ad Spend**

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

**I would recommend below strategies to identify potential brand ambassadors or advocates who could help promote Instagram's initiatives or events:**

* **High Engagement Users:**

Start by analyzing user activity data to find those with the highest engagement rates. These are the users who consistently like, comment, and share content. Their active participation and diverse tag usage make them ideal candidates for brand ambassadorship.

* **Influential Interactors:**

Look for users who have a significant influence on their followers. These individuals may not have the largest follower count, but their posts generate a high level of interaction. Their ability to engage and inspire others makes them valuable advocates for promoting initiatives and events.

* **Tag-Based Enthusiasts:**

Examine the tags that users frequently engage with. Identify those who are passionate about specific topics like food, photography, or fashion. These users can be segmented based on their tag preferences, allowing you to match them with relevant campaigns that align with their interests.

* **Consistent Contributors:**

Focus on users who consistently create and share high-quality content. Regular posting and a strong visual aesthetic are key indicators of potential brand ambassadors. Their dedication to content creation ensures they can effectively represent Instagram’s initiatives.

* **Engagement Growth Patterns:**

Identify users who have shown significant growth in their engagement metrics over time. These rising stars are often on an upward trajectory and can bring fresh energy and enthusiasm to your campaigns.

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

**Problem Statement:** You are hired as a data analyst at Meta and asked to collaborate with the Marketing team. Marketing teams want to leverage Instagram's user data to develop targeted marketing strategies that will increase user engagement, retention, and acquisition. Provide insights and recommendations to address the following objectives

1. **Define Clear Objectives**

**Engagement:** How users interact with posts (likes, comments and share)

**Retention:** How often user comes back to the platform

**Acquisition:** Attract new users

1. **Data Collection and Cleaning**

Gather relevant Instagram user data, including:

* Activity Data: Posts, likes, comments, and tags
* Engagement Metrics: Impressions, clicks, conversions
* Retention Metrics: Active vs. inactive users

**3. Exploratory Data Analysis (EDA)**

Conduct exploratory data analysis to uncover patterns and insights:

* **User Segmentation:** Group users based on engagement levels, interests, and demographics.
* **Engagement Analysis:** Identify peak engagement times, popular tags, and content types that drive interactions.
* **Engagement per Post**: Analyze likes and comments per post to determine content effectiveness.

**4. Analyze Hashtag Usage**

* Track frequently used hashtags to understand user preferences.
* Identify which hashtags drive the most likes and comments.
* Motivate users to adopt popular hashtags through engaging campaigns.

**5. Personalization and Targeting**

Leverage insights from the analysis and models to develop personalized marketing strategies:

* **High Engagement Users:** Offer exclusive rewards, personalized content, and influencer collaborations.
* **Low Engagement Users:** Implement re-engagement campaigns with personalized incentives and targeted content.
* **Tag-Based Campaigns:** Create campaigns aligned with users' tag preferences (e.g., food, travel) to boost relevance and engagement.

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?

This can be done using DML(Data Manipulation Language) function:

Query to do this is:

**UPDATE User\_Interactions**

**SET Engagement\_Type = 'Heart'**

**WHERE Engagement\_Type = 'Like';**