Meta

SOCIAL MEDIA PROJECT

Objective Questions

Q1. Are there any tables with duplicate or missing null values? If so, how would you handle them?

Ans. I generated a query Using the GROUP BY clause to group all rows by the target column(s) – i.e. the column(s) you want to check for duplicate and null values on.
And this proves we don't have any duplicate values or any Null values.

```sql
-- Checking for DUPLICATE values in the users table
SELECT username, COUNT(*)
FROM users
GROUP BY username
HAVING COUNT(*) > 1;

--  Checking for NULL values in the users table
SELECT *
FROM users
WHERE username IS NULL;
```

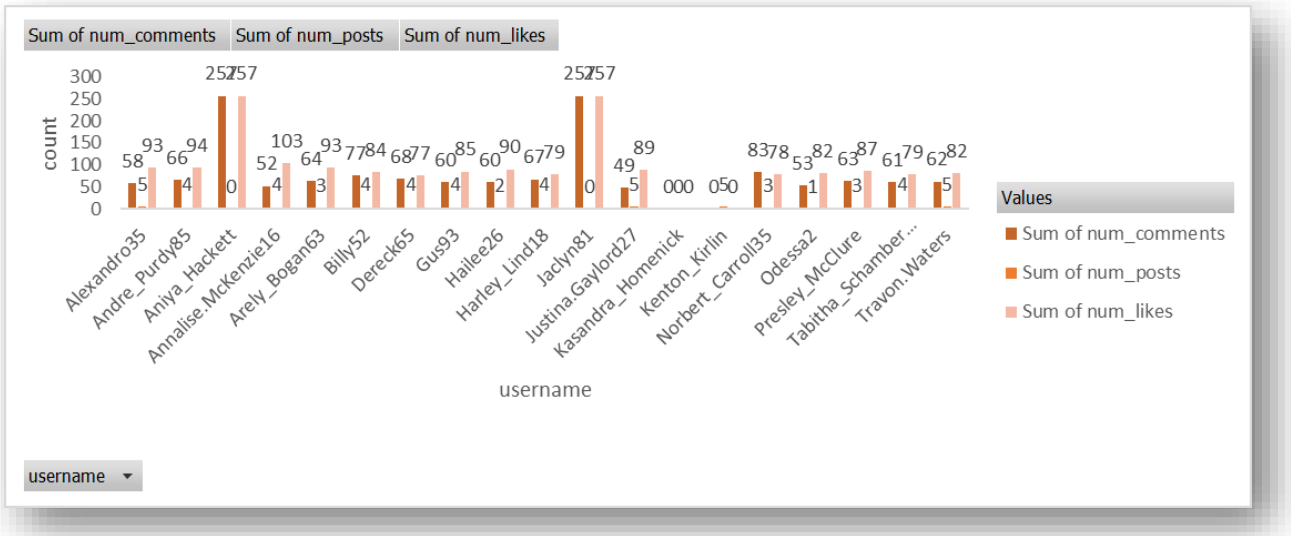
Q2. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?

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

```sql
SELECT u.id AS user_id, u.username,
       COUNT(DISTINCT p.id) AS num_posts,
       COUNT(DISTINCT l.photo_id) AS num_likes,
       COUNT(DISTINCT c.id) AS num_comments
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;
```

The distribution of user activity according to the likes, comments and number of posts as following

id	username	total_posts	total_likes	total_comments
1	Kenton_Kirlin	5	0	0
2	Andre_Purdy85	4	94	66
3	Harley_Lind18	4	79	67
4	Arely_Bogan63	3	93	64
5	Aniya_Hackett	0	257	257
6	Travon.Waters	5	82	62
7	Kasandra_Homenick	0	0	0
8	Tabitha_Schamberger11	4	79	61
9	Gus93	4	85	60
10	Presley_McClure	3	87	63
11	Justina.Gaylord27	5	89	49
12	Dereck65	4	77	68
13	Alexandro35	5	93	58
14	Jaclyn81	0	257	257
15	Billy52	4	84	77
16	Annalise.McKenzie16	4	103	52
17	Norbert_Carroll35	3	78	83
18	Odessa2	1	82	53
19	Hailee26	2	90	60



Q3. Calculate the average number of tags per post (photo_tags and photos tables).

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

```sql
SELECT AVG(tags) AS avg_tags_per_post
FROM (SELECT p.id,COUNT(t.tag_id) AS tags
FROM photos p
LEFT JOIN photo_tags t ON p.id = t.photo_id
GROUP BY p.id) AS num_tags;
```

	avg_tags_per_post
▶	1.9494

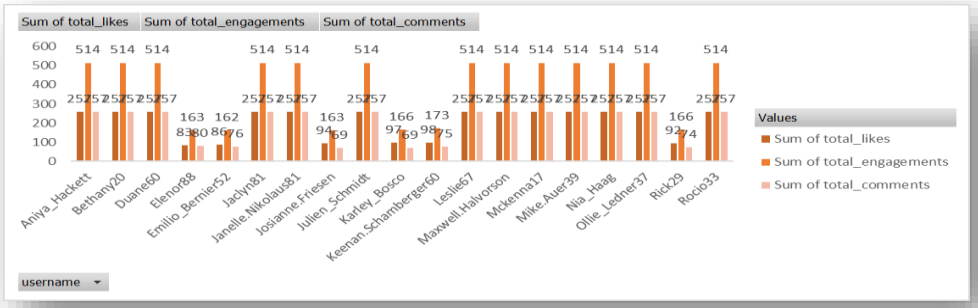
Q4. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.

Ans. To identify the top users with highest engagement rates I created a Query in this query I have created temporary tables 'l' and 'c' using subqueries in which I calculated total likes by a user_id and total comments by a user_id then joined these two tables with the main query to find the user_id, username and there subsequent total likes and total comments.

```sql
select u.id, 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_engagements,
  rank() over(order by (coalesce(l.total_likes,0) + coalesce(c.total_comments,0)) desc ) as engagement_rankings
from users u
left join ( select user_id,count(*) as total_likes from likes group by user_id) l on u.id=l.user_id
left join ( select user_id, count(*) as total_comments from comments group by user_id) c on u.id=c.user_id
order by engagement_rankings;
```

Giving us the following output.

id	username	total_likes	total_comments	total_engagements	engagement_rankings
71	Nia_Haag	257	257	514	1
5	Aniya_Hackett	257	257	514	1
41	Mckenna17	257	257	514	1
21	Rocio33	257	257	514	1
66	Mike.Auer39	257	257	514	1
54	Duane60	257	257	514	1
24	Maxwell.Halvorson	257	257	514	1
14	Jaclyn81	257	257	514	1
57	Julien_Schmidt	257	257	514	1
91	Bethany20	257	257	514	1
76	Janelle.Nikolaus81	257	257	514	1
75	Leslie67	257	257	514	1
36	Ollie_Ledner37	257	257	514	1
96	Keenan.Schamber...	98	75	173	14
69	Karley_Bosco	97	69	166	15
87	Rick29	92	74	166	15
63	Elenor88	83	80	163	17
26	Josianne.Friesen	94	69	163	17
67	Emilio_Bernier52	86	76	162	19



Q5. Which users have the highest number of followers and followings?

Ans. In this query I used cte(Common Table Expression) concept to create tables for finding followee_id and their followers_count and follower_id and their following_count respectively. Then I joined both the ctes with the main query to find user_id, username, followers count, following count. And generated the following output.

```sql
with cte1 as (
  select
    followee_id,
    count(follower_id) as followers_count
    from follows
    group by followee_id
),
cte2 as (
  select
    follower_id,
        count(followee_id) as followings_count
        from follows
        group by follower_id
)
select
  u.id,
    u.username,
  coalesce(cte1.followers_count,0)as followers_count,
  coalesce(cte2.followings_count,0) as followings_count
from users u
left join cte1 on u.id=cte1.followee_id
left join cte2 on u.id=cte2.follower_id
order by followers_count desc, followings_count desc;
```

id	username	followers_count	followings_count
1	Kenton_Kirlin	77	0
23	Eveline95	77	0
29	Jaime53	77	0
34	Pearl7	77	0
25	Tierra.Trantow	77	0
45	David.Osinski47	77	0
7	Kasandra_Homenick	77	0
51	Mariano_Koch3	77	0
49	Morgan.Kassulke	77	0
53	Linnea59	77	0
58	Aurelie71	77	0
59	Cesar93	77	0
64	Florence99	77	0
68	Franco_Keebler64	77	0
74	Hulda.Macejkovic	77	0
88	Clint27	77	0
77	Donald.Fritsch	77	0
80	Darby_Herzog	77	0
81	Esther.Zulauf61	77	0

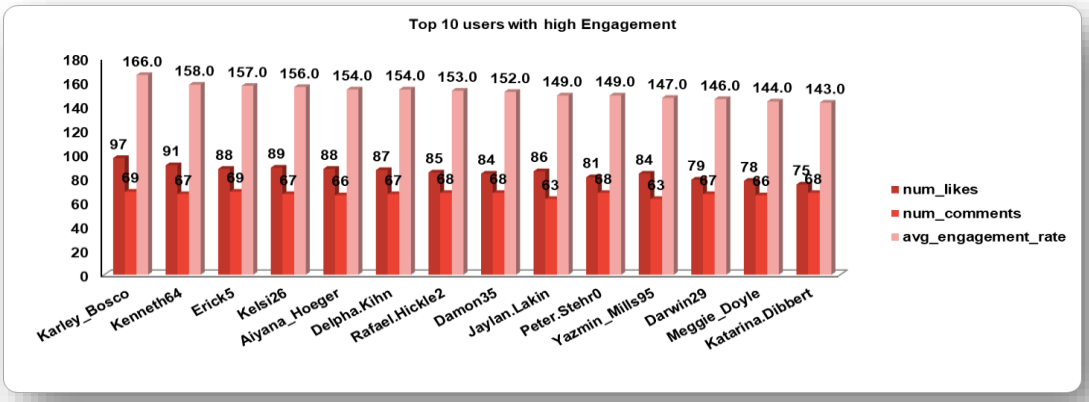
Q6. Calculate the average engagement rate (likes, comments) per post for each user.

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

```sql
SELECT
  u.id as user_id,
  u.username,
  COALESCE(p.num_posts, 0) AS num_posts,
  COALESCE(l.num_likes, 0) AS num_likes,
  COALESCE(c.num_comments, 0) AS num_comments,
  CASE
    WHEN COALESCE(p.num_posts, 0) = 0 THEN 0
    ELSE (COALESCE(l.num_likes, 0) + COALESCE(c.num_comments, 0)) / COALESCE(p.num_posts, 0)
  END AS avg_engagement_rate
FROM users u
LEFT JOIN (SELECT user_id, COUNT(*) AS num_posts FROM photos
     GROUP BY user_id) p ON u.id = p.user_id
LEFT JOIN (SELECT user_id, COUNT(*) AS num_likes
     FROM likes
     GROUP BY user_id) l ON u.id = l.user_id
LEFT JOIN (SELECT user_id, COUNT(*) AS num_comments FROM comments GROUP BY user_id) c ON u.id = c.user_id
  ORDER BY avg_engagement_rate DESC;
```

The following is the output for average engagement rate for every user.

user_id	username	num_posts	num_likes	num_comments	avg_engagement_rate
69	Karley_Bosco	1	97	69	166.0000
22	Kenneth64	1	91	67	158.0000
70	Erick5	1	88	69	157.0000
39	Kelsi26	1	89	67	156.0000
31	Aiyana_Hoeger	1	88	66	154.0000
20	Delpha.Kihn	1	87	67	154.0000
40	Rafael.Hickle2	1	85	68	153.0000
94	Damon35	1	84	68	152.0000
73	Jaylan.Lakin	1	86	63	149.0000
56	Peter.Stehr0	1	81	68	149.0000
37	Yazmin_Mills95	1	84	63	147.0000
27	Darwin29	1	79	67	146.0000
55	Meggie_Doyle	1	78	66	144.0000
79	Katarina.Dibbert	1	75	68	143.0000
61	Jayson65	1	83	58	141.0000
98	Imani_Nicolas17	1	74	65	139.0000
18	Odessa2	1	82	53	135.0000
48	Granville_Kutch	1	75	55	130.0000
35	Lennie_Hartm...	2	92	67	79.5000



Q7. Get the list of users who have never liked any post (users and likes tables)

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

```sql
SELECT
  u.id,
    u.username
FROM users u
LEFT JOIN likes l ON u.id = l.user_id
WHERE l.user_id IS NULL;
```

This query gave us the desired output of the list of those users who have never liked any post.

id	username
1	Kenton_Kirlin
7	Kasandra_Homenick
23	Eveline95
25	Tierra.Trantow
29	Jaime53
34	Pearl7
45	David.Osinski47
49	Morgan.Kassulke
51	Mariano_Koch3
53	Linnea59
58	Aurelie71
59	Cesar93
64	Florence99
68	Franco_Keebler64
74	Hulda.Macejkovic
77	Donald.Fritsch
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard

Q8. How can you leverage user-generated content (posts, hashtags, photo tags) to create more personalised and engaging ad campaigns?

Ans.

```sql
-- MOST USED TAGS:
SELECT
    t.tag_name,
    COUNT(pt.photo_id) AS tag_usage_count
FROM tags t
JOIN photo_tags pt ON t.id = pt.tag_id
GROUP BY t.tag_name
ORDER BY tag_usage_count DESC;
```

Result Grid | Filter Rows:

tag_name	tag_usage_count
smile	59
beach	42
party	39
fun	38
concert	24
food	24
lol	24
hair	23
happy	22
beauty	20
dreamy	20
drunk	19
fashion	19
sunset	19
landscape	17
stvle	17

1.This query calculates the average number of likes for photos associated with each tag.

2.It first computes the total likes per photo and then averages these likes across all photos tagged with each tag.

3.The results are grouped by tag name and sorted in descending order of average likes.

4.Finally, the top 10 tags with the highest average likes are selected and displayed.

```sql
-- USER SEGMENT
SELECT
    u.id AS user_id,
    u.username,
    t.tag_name
FROM users u
JOIN likes l ON u.id = l.user_id
JOIN photos p ON l.photo_id = p.id
JOIN photo_tags pt ON p.id = pt.photo_id
JOIN tags t ON pt.tag_id = t.id
GROUP BY t.tag_name,u.username,u.id;
```

Result Grid | Filter Rows:

user_id	username	tag_name
2	Andre_Purdy85	fun
5	Aniya_Hackett	fun
9	Gus93	fun
10	Presley_McClure	fun
11	Justina.Gaylord27	fun
14	Jadyn81	fun
19	Hailee26	fun
21	Rocio33	fun
24	Maxwell.Halvorson	fun
35	Lennie_Hartmann40	fun
36	Ollie_Ledner37	fun
41	Mckenna17	fun
46	Malinda_Streich	fun
47	Harrison.Beatty50	fun
54	Duane60	fun
55	Meggie_Doyle	fun
57	Julien_Schmidt	fun
66	Mike.Auer39	fun
69	Karley_Bosco	fun
71	Nia_Haag	fun

1.This query counts how often each tag is used in photos.

2.It joins the tags and photo_tags tables to link tags with photos and then groups the results by tag name.

3.The query orders the tags by their usage count in descending order, showing which tags are most frequently associated with photos.

4.The final output displays each tag's name and the number of times it has been used.

SUGGESTIONS:

1. **Analyze User Interests**: I would start by identifying popular hashtags and photo tags to see what themes or topics users are engaging with. This insight allows us to create ads that resonate with these themes, increase engagement.
2. **Target Specific Audiences**: I'd use data from posts and tags to group users into segments based on their interests and behaviors. This lets us target ads specifically to these groups, increasing the chance they'll engage with the content.
3. **Leverage Social Proof**: Featuring user comments, or popular posts in campaigns can build trust and drive conversions. This showcases real user experiences and encourages others to participate, boosting ad effectiveness.
4. **Promote Influencers and Power Users**: Engaging with users who are particularly active with relevant tags or content can amplify brand reach. Brands can offer these users special incentives or collaborations, letting their authentic voice enhance brand credibility and expand its reach.
5. **Keep Ads Updated with Trends**: By tracking trending hashtags, I could help keep ads aligned with current trends, ensuring they're fresh and engaging for the audience.

Q9. 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?

Ans.

Method for Determining Correlation:

1. To determine whether user activity levels and content kinds are related, I would begin with a correlation analysis utilizing statistical techniques. This would let us to determine whether users are more interested in reels, videos, or photographs.

2. If there's a strong correlation between high engagement and certain content types, it could guide us in focusing more on those types to maximize user interaction and satisfaction.

Guiding Content Creation and Curation:

1.Include Trendy subjects: Using popular hashtags and trendy subjects on a frequent basis will assist provide new and pertinent content to keep visitors interested.

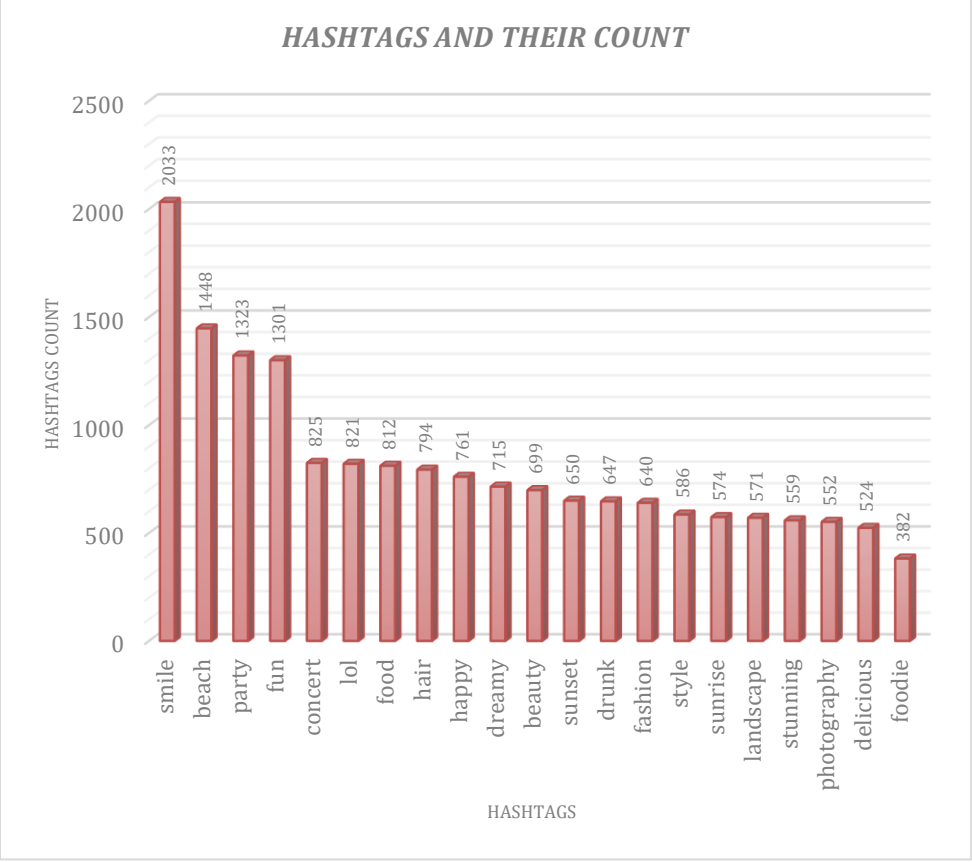
2.Customized Suggestions: We may make recommendations for material that corresponds with what users most frequently watch or follow based on their interests and previous interactions.

3.Use of Relevant Hashtags: Adding hashtags associated with trending subjects might

help increase user engagement because they are likely to draw in visitors and promote conversation.

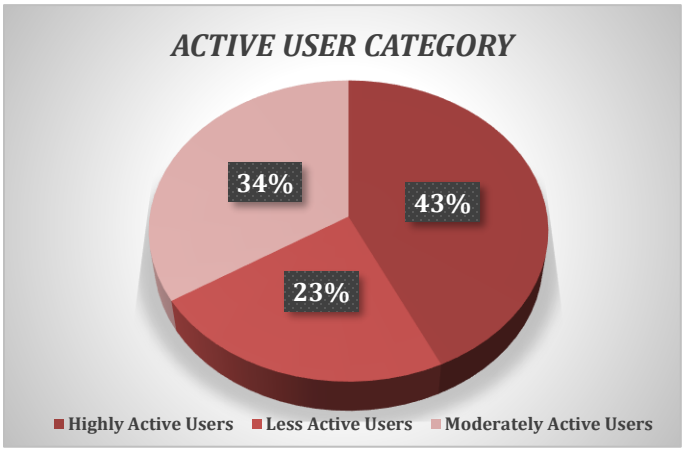
```sql
-- Hashtags and their count
SELECT
    t.tag_name as hashtags,
    count(*) as total_hashtags_count
FROM users u
JOIN likes l ON u.id = l.user_id
JOIN photos p ON l.photo_id = p.id
JOIN photo_tags pt ON p.id = pt.photo_id
JOIN tags t ON pt.tag_id = t.id
GROUP BY tag_name
ORDER BY total_hashtags_count DESC;
```

hashtags	total_hashtags_count
smile	2033
beach	1448
party	1323
fun	1301
concert	825
lol	821
food	812
hair	794
happy	761
dreamy	715
beauty	699
sunset	650
drunk	647
fashion	640
style	586
sunrise	574
landscape	571
stunning	559
photography	552
delicious	524
foodie	382



```sql
-- Active User Category
SELECT
    CASE
    WHEN COALESCE(l.total_likes, 0) + COALESCE(c.total_comments, 0) > 150
      THEN 'Highly Active Users'
    WHEN COALESCE(l.total_likes, 0) + COALESCE(c.total_comments, 0) BETWEEN 100 AND 150
      THEN 'Moderately Active Users'
    ELSE 'Less Active Users' END AS user_category,
    count(*) as total_count
FROM users u
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_likes
    FROM ig_clone.likes
    GROUP BY user_id) l ON u.id = l.user_id
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_comments
    FROM comments
    GROUP BY user_id) c ON u.id = c.user_id
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_posts
    FROM photos
    GROUP BY user_id) p ON u.id = p.user_id
GROUP BY user_category
ORDER BY user_category;
```

user_category	total_count
Highly Active Users	43
Less Active Users	23
Moderately Active Users	34



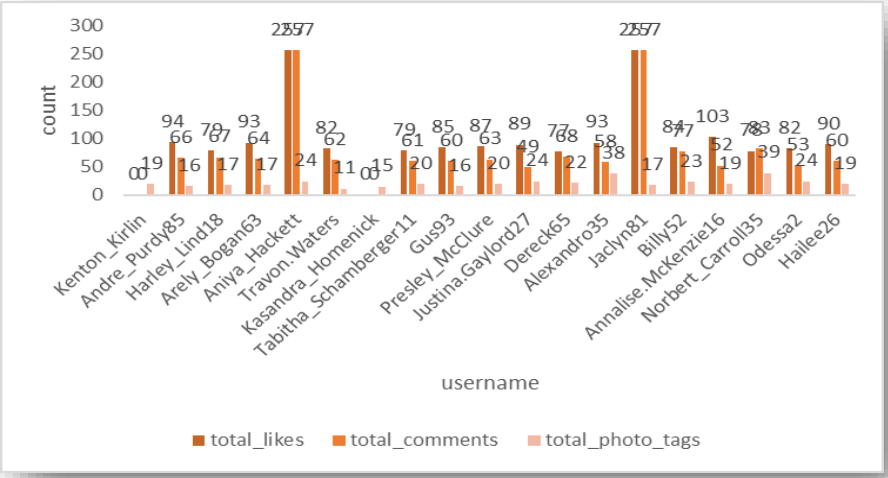
Q10. Calculate the total number of likes, comments, and photo tags for each user.

Ans. In this query I utilized subqueries and joins for finding total likes, total comments, total photo tags then joined them to the main query to find the id, username and their respective total likes, total comments, total photo tags.

```sql
SELECT
  u.id as id,
    u.username,
  COALESCE(l.total_likes, 0) AS total_likes,
  COALESCE(c.total_comments, 0) AS total_comments,
  COALESCE(pt.total_photo_tags, 0) AS total_photo_tags
FROM users u
LEFT JOIN (
  SELECT
    user_id,
    COUNT(*) AS total_likes
    FROM likes
    GROUP BY user_id
) l ON u.id = l.user_id
LEFT JOIN (
  SELECT
    user_id,
        COUNT(*) AS total_comments
        FROM comments
        GROUP BY user_id
) c ON u.id = c.user_id
LEFT JOIN (
  SELECT
    tag_id,
        COUNT(*) AS total_photo_tags
        FROM photo_tags
        GROUP BY tag_id
) pt ON u.id = pt.tag_id;
```

Thus, giving me the following output.

id	username	total_likes	total_comments	total_photo_tags
1	Kenton_Kirlin	0	0	19
2	Andre_Purdy85	94	66	16
3	Harley_Lind18	79	67	17
4	Arely_Bogan63	93	64	17
5	Aniya_Hackett	257	257	24
6	Travon.Waters	82	62	11
7	Kasandra_Homenick	0	0	15
8	Tabitha_Schamberger11	79	61	20
9	Gus93	85	60	16
10	Presley_McClure	87	63	20
11	Justina.Gaylord27	89	49	24
12	Dereck65	77	68	22
13	Alexandro35	93	58	38
14	Jaclyn81	257	257	17
15	Billy52	84	77	23
16	Annalise.McKenzie16	103	52	19
17	Norbert_Carroll35	78	83	39
18	Odessa2	82	53	24
19	Hailee26	90	60	19



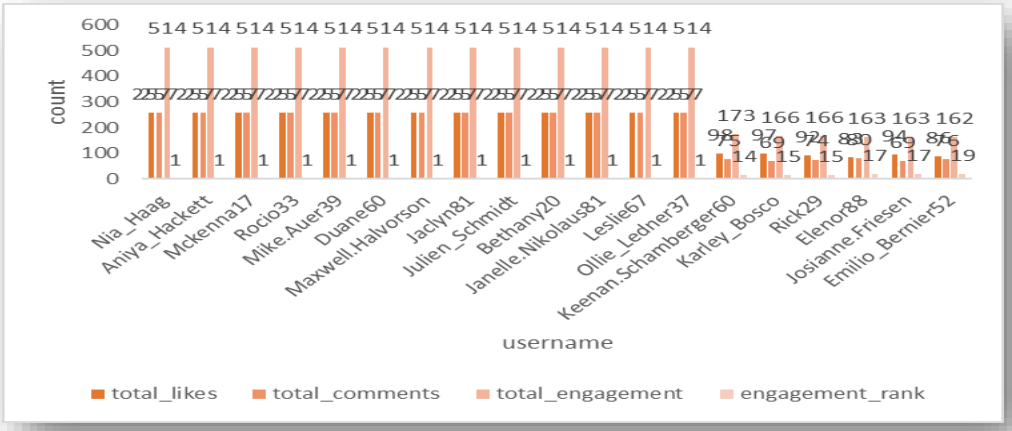
Q11. Rank users based on their total engagement (likes, comments, shares) over a month.

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

```sql
WITH MonthlyEngagement AS (
    SELECT u.id AS user_id, 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
    FROM users u
    LEFT JOIN (
        SELECT user_id, COUNT(photo_id) AS total_likes
        FROM likes
        WHERE DATE(created_at) >= '2024-07-01' OR DATE(created_at) <= '2024-07-31'
        GROUP BY user_id) l ON u.id = l.user_id
    LEFT JOIN (
        SELECT user_id, COUNT(id) AS total_comments
        FROM comments
        WHERE DATE(created_at) >= '2024-07-01' OR DATE(created_at) <= '2024-07-31'
        GROUP BY user_id) c ON u.id = c.user_id)
SELECT
  user_id,
    username,
    total_likes,
    total_comments,
    total_engagement,
    RANK() OVER (ORDER BY total_engagement DESC)
AS engagement_rank
FROM MonthlyEngagement
ORDER BY engagement_rank;
```

This gave me the desired output in which I got user_id, username, their total likes, total comments, total engagements and their ranking based on their engagement.

user_id	username	total_likes	total_comments	total_engagement	engagement_rank
71	Nia_Haag	257	257	514	1
5	Aniya_Hackett	257	257	514	1
41	Mckenna17	257	257	514	1
21	Rocio33	257	257	514	1
66	Mike.Auer39	257	257	514	1
54	Duane60	257	257	514	1
24	Maxwell.Halvorson	257	257	514	1
14	Jaclyn81	257	257	514	1
57	Julien_Schmidt	257	257	514	1
91	Bethany20	257	257	514	1
76	Janelle.Nikolaus81	257	257	514	1
75	Leslie67	257	257	514	1
36	Ollie_Ledner37	257	257	514	1
96	Keenan.Schamber...	98	75	173	14
69	Karley_Bosco	97	69	166	15
87	Rick29	92	74	166	15
63	Elenor88	83	80	163	17
26	Josianne.Friesen	94	69	163	17
67	Emilio_Bernier52	86	76	162	19



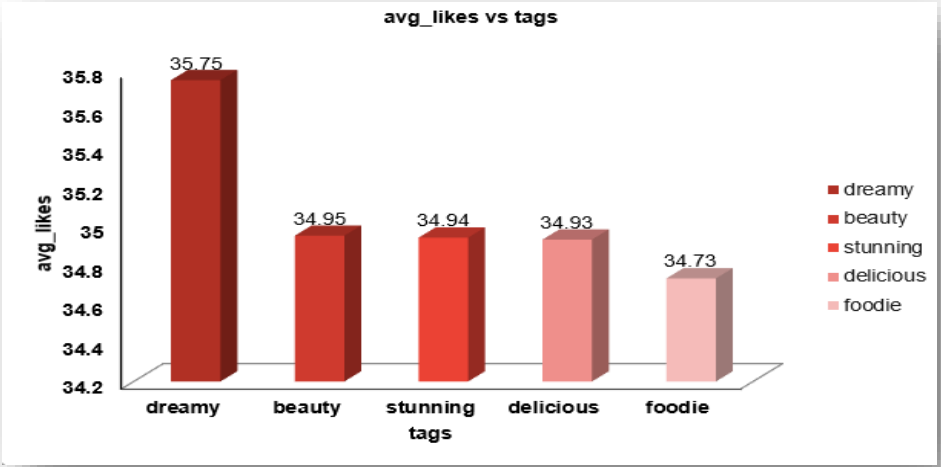
Q12. Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.

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

```sql
WITH Hashtag_Likes AS (
    SELECT ht.tag_name, COUNT(l.photo_id) AS total_likes, COUNT(DISTINCT p.id) AS total_posts
    FROM tags ht
    JOIN photo_tags pt ON ht.id = pt.tag_id
    JOIN photos p ON pt.photo_id = p.id
    LEFT JOIN likes l ON p.id = l.photo_id
    GROUP BY ht.tag_name),
Average_Likes_Per_Hashtag AS (
    SELECT tag_name, (CAST(total_likes AS FLOAT) / total_posts) AS avg_likes
    FROM Hashtag_Likes)
SELECT tag_name, round(avg_likes,2) as avg_Likes
FROM Average_Likes_Per_Hashtag
order by avg_likes desc
limit 5;
```

This query gave the following desired output.

tag_name	avg_Likes
dreamy	35.75
beauty	34.95
stunning	34.94
delicious	34.93
foodie	34.73



Q13. Retrieve the users who have started following someone after being followed by that person

Ans. In this question in order to retrieve the users who have started following someone after being followed by that person I employed a Query where I used the created_at column to compare the followee_id and follower_id whichever was created before. But didn't find any such record proving there is no such kind of users.

```sql
SELECT
  f1.follower_id,
    f1.followee_id
FROM follows f1
JOIN follows f2 ON f1.follower_id = f2.followee_id AND f1.followee_id = f2.follower_id
WHERE f1.created_at > f2.created_at;
```

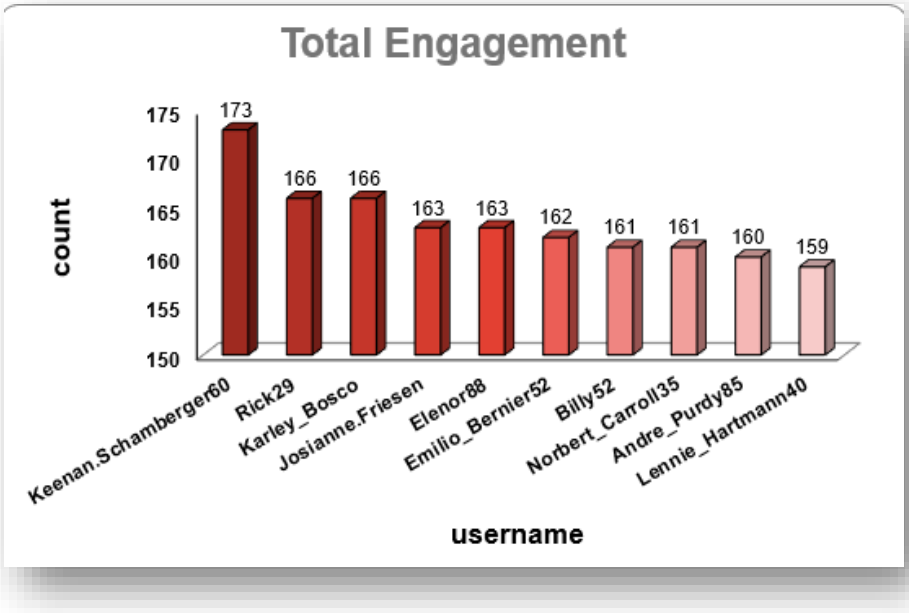
followed_back_user	original_follower	followed_back_at	originally_followed_at

Subjective Questions

Q1. 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?
Ans

```sql
WITH TotalLikes AS (
    SELECT u.id, COUNT(distinct l.photo_id) AS total_likes
    FROM users u
    LEFT JOIN likes l ON u.id = l.user_id
    GROUP BY u.id),
TotalComments AS (
    SELECT u.id, COUNT(distinct c.photo_id) AS total_comments
    FROM users u
    LEFT JOIN comments c ON u.id = c.user_id
    GROUP BY u.id),
PhotosPosted AS (
    SELECT user_id, COUNT(id) AS total_photos_posted
    FROM photos
    GROUP BY user_id),
Followers AS (
    SELECT followee_id AS user_id, COUNT(follower_id) AS total_followers
    FROM follows
    GROUP BY followee_id),
UniqueTags AS (
    SELECT p.user_id, COUNT(DISTINCT pt.tag_id) AS unique_tags_used
    FROM photos p
    LEFT JOIN photo_tags pt ON p.id = pt.photo_id
    GROUP BY p.user_id)
SELECT u.id AS user_id, u.username,
    COALESCE(tl.total_likes, 0) AS total_likes,
    COALESCE(tc.total_comments, 0) AS total_comments,
    COALESCE(pp.total_photos_posted, 0) AS total_photos_posted,
    COALESCE(f.total_followers, 0) AS total_followers,
    COALESCE(ut.unique_tags_used, 0) AS unique_tags_used,
    (COALESCE(tl.total_likes, 0) + COALESCE(tc.total_comments, 0)) AS total_engagement
FROM users u
LEFT JOIN TotalLikes tl ON u.id = tl.id
LEFT JOIN TotalComments tc ON u.id = tc.id
LEFT JOIN PhotosPosted pp ON u.id = pp.user_id
LEFT JOIN Followers f ON u.id = f.user_id
LEFT JOIN UniqueTags ut ON u.id = ut.user_id
group by u.id
having total_photos_posted >0
ORDER BY total_engagement DESC, total_followers DESC, total_photos_posted DESC
limit 10;
```

user_id	username	total_likes	total_comments	total_photos_posted	total_followers	unique_tags_used	total_engagement
96	Keenan.Schamberger60	98	75	3	76	6	173
87	Rick29	92	74	4	76	10	166
69	Karley_Bosco	97	69	1	76	3	166
26	Josianne.Friesen	94	69	5	76	11	163
63	Elenor88	83	80	4	76	6	163
67	Emilio_Bernier52	86	76	3	76	3	162
15	Billy52	84	77	4	76	4	161
17	Norbert_Carroll35	78	83	3	76	7	161
2	Andre_Purdy85	94	66	4	76	7	160
35	Lennie_Hartmann40	92	67	2	76	0	159



INSIGHTS:

a) ***Top Contributors Identified:*** By analysing user activity data, we identified the top 10 users who consistently engage with the platform through posts, likes, and comments.
b) ***High Engagement = High Loyalty:*** Users with the highest number of likes and comments demonstrated strong engagement, suggesting they are more loyal and invested in the platform's content.
c) ***Activity Distribution:*** Among total activities, likes and comments formed the majority of user engagement, highlighting the importance of interactive features over just content creation.
d) ***Consistency Over Time:*** Regular activity, not just one-time participation, was a key factor in determining loyalty, ensuring that the selected users are not just one-time contributors.

RECOMMENDATIONS:

a) ***Reward Loyalty:*** Consider implementing a loyalty or recognition program (e.g., badges, leaderboard, or exclusive access) for the top 10 most engaged users to maintain and increase motivation.
b) **Encourage Broader Engagement:** Promote interactive features like liking and commenting by adding gamified incentives to encourage less active users to participate more.
c) ***Track Engagement Trends:*** Continuously monitor user engagement through similar metrics to identify rising contributors early and nurture them.
d) ***User Segmentation:*** Create user segments based on engagement levels to send targeted content, rewards, or campaigns to each group accordingly.
e) ***Enhance Community Features:*** Improve or expand community-driven features (e.g., discussions, replies, mentions) as these drive both engagement and loyalty.

Q2. For inactive users, what strategies would you recommend to re-engage them and encourage them to start posting or engaging again?

Ans.

```sql
WITH likes_count AS (
  SELECT
    DISTINCT user_id,
    count(*) AS num_of_likes
  FROM likes
  GROUP BY User_id
),
comments_count AS (
  SELECT
    user_id,
    count(id) AS num_of_comments
  FROM comments
  GROUP BY user_id
),
photo_counts AS (
  SELECT
    user_id,
    COUNT(*) AS num_of_photos
  FROM photos
  GROUP BY user_id
),
phototags_count AS (
  SELECT
    p.user_id,
    count(pt.tag_id) AS num_of_phototags
  FROM photos p
  JOIN photo_tags AS pt ON p.user_id = pt.photo_id
  GROUP BY p.user_id
),
Count_of_followers AS (
  SELECT
    follower_id,
    count(follower_id) AS follower_count,
    count(followee_id) AS followee_count
  FROM follows
  GROUP BY follower_id
)
SELECT
  u.id AS UserID,
  u.username AS UserName,
  coalesce(l.num_of_likes, 0) AS num_of_likes,
  coalesce(c.num_of_comments, 0) AS num_of_comments,
  coalesce(pp.num_of_photos, 0) AS num_of_photos,
  coalesce(p.num_of_phototags, 0) AS num_of_phototags,
  coalesce(f.follower_count, 0) AS follower_count,
  coalesce(f.followee_count, 0) AS followee_count,
  coalesce(
    (coalesce(l.num_of_likes, 0) + coalesce(c.num_of_comments, 0) + coalesce(pp.num_of_photos, 0)),0
  ) AS engagement_rate,
  DENSE_RANK() OVER (
    ORDER BY
      (coalesce(l.num_of_likes, 0) + coalesce(c.num_of_comments, 0) + coalesce(pp.num_of_photos, 0)) ASC
  ) AS engagement_rate_rank
FROM
  users u
  LEFT JOIN likes_count AS l ON u.id = l.user_id
  LEFT JOIN comments_count AS c ON u.id = c.user_id
  LEFT JOIN photo_counts AS pp ON u.id = pp.user_id
  LEFT JOIN phototags_count AS p ON u.id = p.user_id
  LEFT JOIN Count_of_followers AS f ON u.id = f.follower_id
ORDER BY
  engagement_rate_rank ASC;
```

UserID	UserName	num_of_likes	num_of_comments	num_of_photos	num_of_phototags	follower_count	followee_count	engagement_rate	engagement_rate_rank
25	Tierra.Trantow	0	0	0	0	0	0	0	1
49	Morgan.Kassulke	0	0	0	0	0	0	0	1
81	Esther.Zulauf61	0	0	0	0	0	0	0	1
80	Darby_Herzog	0	0	0	0	0	0	0	1
7	Kasandra_Homenick	0	0	0	0	0	0	0	1
68	Franco_Keebler64	0	0	0	0	0	0	0	1
45	David.Osinski47	0	0	0	0	0	0	0	1
53	Linnea59	0	0	0	0	0	0	0	1
34	Pearl7	0	0	0	0	0	0	0	1
74	Hulda.Macejkovic	0	0	0	0	0	0	0	1
83	Bartholome.Bernhard	0	0	0	0	0	0	0	1
89	Jessyca_West	0	0	0	0	0	0	0	1
90	Esmeralda.Mraz57	0	0	0	0	0	0	0	1
1	Kenton_Kirlin	0	0	5	25	0	0	5	2
51	Mariano_Koch3	0	0	5	15	0	0	5	2
64	Florence99	0	0	5	20	0	0	5	2
77	Donald.Fritsch	0	0	6	0	0	0	6	3
58	Aurelie71	0	0	8	8	0	0	8	4
29	Jaime53	0	0	8	32	0	0	8	4
86	Delfina_VonRueden68	0	0	9	9	0	0	9	5
59	Cesar93	0	0	10	40	0	0	10	6
88	Clint27	0	0	11	22	0	0	11	7
23	Eveline95	0	0	12	48	0	0	12	8
48	Granville_Kutch	75	55	1	0	99	99	131	9
18	Odessa2	82	53	1	0	99	99	136	10
97	Tomas.Beatty93	69	68	2	6	99	99	139	11
28	Dario77	77	59	4	8	99	99	140	12
47	Harrison.Beatty50	76	59	5	5	99	99	140	12
98	Imani_Nicolas17	74	65	1	4	99	99	140	12
61	Jayson65	83	58	1	2	99	99	142	13
11	Justina.Gaylord27	89	49	5	15	99	99	143	14
33	Yvette.Gottlieb91	77	61	5	0	99	99	143	14
42	Maya.Farrell	87	54	3	0	99	99	144	15
8	Tabitha_Schamberger11	79	61	4	16	99	99	144	15
79	Katarina.Dibbert	75	68	1	3	99	99	144	15
99	Alek_Watsica	74	68	3	0	99	99	145	16
38	Jordyn.Jacobson2	85	58	2	4	99	99	145	16
55	Meggie_Doyle	78	66	1	1	99	99	145	16
52	Zack_Kemmer93	85	56	5	0	99	99	146	17
85	Milford_Gleichner42	87	57	2	6	99	99	146	17
27	Darwin29	79	67	1	0	99	99	147	18
78	Colten.Harris76	83	60	5	15	99	99	148	19
62	Ressie_Stanton46	88	58	2	2	99	99	148	19
30	Kaley9	81	65	2	6	99	99	148	19
37	Yazmin_Mills95	84	63	1	3	99	99	148	19
12	Dereck65	77	68	4	4	99	99	149	20
9	Gus93	85	60	4	4	99	99	149	20
6	Travon.Waters	82	62	5	20	99	99	149	20
44	Seth46	86	60	4	8	99	99	150	21
3	Harley_Lind18	79	67	4	4	99	99	150	21
73	Jaylan.Lakin	86	63	1	2	99	99	150	21
56	Peter.Stehr0	81	68	1	1	99	99	150	21
32	Irwin.Larson	91	56	4	20	99	99	151	22
19	Hailee26	90	60	2	2	99	99	152	23
50	Gerard79	81	69	3	3	99	99	153	24
10	Presley_McClure	87	63	3	3	99	99	153	24
82	Aracely.Johnston98	84	67	2	0	99	99	153	24
84	Alysa22	75	76	2	6	99	99	153	24
94	Damon35	84	68	1	5	99	99	153	24
100	Javonte83	82	70	2	2	99	99	154	25
72	Kathryn80	85	64	5	15	99	99	154	25
40	Rafael.Hickle2	85	68	1	0	99	99	154	25
31	Aiyana_Hoeger	88	66	1	2	99	99	155	26
92	Frederik_Rice	91	61	3	3	99	99	155	26
20	Delpha.Kihn	87	67	1	0	99	99	155	26
95	Nicole71	86	68	2	10	99	99	156	27
93	Willie_Leuschke	91	63	2	8	99	99	156	27
13	Alexandro35	93	58	5	10	99	99	156	27
39	Kelsi26	89	67	1	0	99	99	157	28
70	Erick5	88	69	1	3	99	99	158	29
22	Kenneth64	91	67	1	3	99	99	159	30
16	Annalise.McKenzie16	103	52	4	0	99	99	159	30
4	Arely_Bogan63	93	64	3	12	99	99	160	31
46	Malinda_Streich	88	68	4	0	99	99	160	31
60	Sam52	86	72	2	6	99	99	160	31
65	Adelle96	96	60	5	10	99	99	161	32
35	Lennie_Hartmann40	92	67	2	6	99	99	161	32
43	Janet.Armstrong	86	72	5	15	99	99	163	33
2	Andre_Purdy85	94	66	4	16	99	99	164	34
17	Norbert_Carroll35	78	83	3	9	99	99	164	34
67	Emilio_Bernier52	86	76	3	0	99	99	165	35
15	Billy52	84	77	4	0	99	99	165	35
69	Karley_Bosco	97	69	1	0	99	99	167	36
63	Elenor88	83	80	4	4	99	99	167	36
26	Josianne.Friesen	94	69	5	25	99	99	168	37
87	Rick29	92	74	4	12	99	99	170	38
96	Keenan.Schamberger60	98	75	3	6	99	99	176	39
41	Mckenna17	257	257	0	0	99	99	514	40
36	Ollie_Ledner37	257	257	0	0	99	99	514	40
14	Jaclyn81	257	257	0	0	99	99	514	40
57	Julien_Schmidt	257	257	0	0	99	99	514	40
75	Leslie67	257	257	0	0	99	99	514	40
91	Bethany20	257	257	0	0	99	99	514	40
21	Rocio33	257	257	0	0	99	99	514	40
71	Nia_Haag	257	257	0	0	99	99	514	40
66	Mike.Auer39	257	257	0	0	99	99	514	40
24	Maxwell.Halvorson	257	257	0	0	99	99	514	40
5	Aniya_Hackett	257	257	0	0	99	99	514	40
54	Duane60	257	257	0	0	99	99	514	40
76	Janelle.Nikolaus81	257	257	0	0	99	99	514	40

The query calculates user engagement metrics on a platform, including likes, comments, photos, photo tags, and follower counts.
Common Table Expressions (CTEs):

1. likes_count: Counts likes per user.
2. comments_count: Counts comments per user.
3. photo_counts: Counts photos per user.
4. phototags_count: Counts photo tags per user.
5. Count_of_followers: Counts followers per user.
6. Joins each **CTE** with the users table using **LEFT JOIN** to include all users, even those without specific activities.
7. Uses **COALESCE** to replace **NULL** values with zero.
8. Calculated an engagement_rate as the sum of likes, comments, photos .
9. Then ranked users by engagement_rate in ascending order using **DENSE_RANK().**
10. This ranking will give us the inactive users.

This is how we will rank users based on user engagement and activity levels. The highly rank users can be considered as inactive users.

STRATEGIES RECOMMENDATION FOR RE-ENGAGING USERS:

1. **Re-engagement Campaigns:** These campaigns may consist of time-limited deals, special discounts, or freebies that require interaction, such as commenting, sharing, or tagging a friend.
2. **Re-engagement E-mails:** Send personalized emails showcasing trending content or posts from their connections that they've missed. Suggest content, groups, or hashtags they might be interested in based on their past behaviour.
3. **By Giving Discounts and Gift Card:** This is one of the most popular yet effective ways to re-engage user. Because they find these rewards exciting, this will encourage them to utilize the site more.
4. **Taking Feedback User Surveys:** Ask inactive users for feedback on why they stopped engaging. Use their input to make meaningful changes or offer solutions directly.
5. **Simplifying The Use:** Offer templates or prompts that make it easier for users to create and share content, reducing the friction of coming up with something new.
6. **Contests For Users**: Introduce challenges or contests with rewards for the most engaging content or for hitting specific activity milestones.
7. **Data-Driven Insights:** Use the old data of inactive users when they created account like whom did they follow first, which post they like first and so on.

These insights can help refine content strategy based on user preferences.

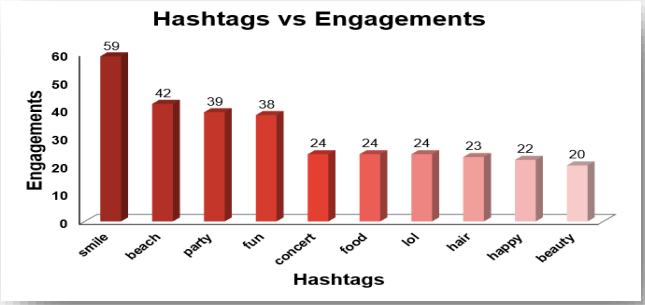
These approaches can help re-establish relationships with inactive users, making them feel valued and more likely to re-engage again.

Q3. Which hashtags or content topics have the highest engagement rates? How can this information guide content strategy and ad campaigns?

Ans.

```sql
WITH PhotoEngagement AS (
    SELECT
        p.id AS photo_id,
        COUNT(distinct l.photo_id) AS total_likes,
        COUNT(DISTINCT c.id) AS total_comments,
        COUNT(distinct l.photo_id) + COUNT(DISTINCT c.user_id) AS total_engagement
    FROM photos p
    LEFT JOIN likes l ON p.user_id = l.user_id
    LEFT JOIN comments c ON p.user_id = c.user_id
    GROUP BY p.id),
HashtagEngagement AS (
    SELECT
        t.id AS tag_id,
        t.tag_name,
        count(pe.total_engagement) AS total_engagement,
        COUNT(DISTINCT pt.photo_id) AS total_photos,
        (count(pe.total_engagement) / COUNT(DISTINCT pt.photo_id) )AS engagement_rate
    FROM tags t
    JOIN photo_tags pt ON t.id = pt.tag_id
    JOIN PhotoEngagement pe ON pt.photo_id = pe.photo_id
    GROUP BY t.id, t.tag_name)
SELECT tag_name, total_photos, total_engagement, engagement_rate
FROM HashtagEngagement
ORDER BY total_engagement DESC
limit 10;
```

tag_name	total_photos	total_engagement	engagement_rate
smile	59	59	1.0000
beach	42	42	1.0000
party	39	39	1.0000
fun	38	38	1.0000
concert	24	24	1.0000
food	24	24	1.0000
lol	24	24	1.0000
hair	23	23	1.0000
happy	22	22	1.0000
beauty	20	20	1.0000



INSIGHTS:

a) ***Top-Performing Hashtags Identified:*** Hashtags such as #smile, #beach, and #party consistently appeared in posts with the highest number of likes and comments, indicating strong audience engagement.

b) ***Engagement Rate as a Key Metric:*** Calculating the engagement rate using total likes and comments per hashtag usage helped identify which hashtags truly resonate with users, beyond just frequency of use.

c) ***Trending Tag Themes:*** The most effective hashtags are often linked to emotional or lifestyle themes, suggesting that users engage more with content that reflects positivity, fun, and everyday experiences.

RECOMMENDATIONS:

a) ***Promote High-Engagement Hashtags:*** Highlight the most engaging hashtags across the platform to encourage more users to include them in their posts. This can be done through recommendations while posting or showcasing them on a trending tags list.

b) ***Hashtag-Driven Content Strategy:*** Create weekly challenges or content prompts around popular hashtags to boost participation. Encourage creators and communities to engage with these specific tags to drive visibility and engagement.

c) ***Targeted Ad Campaigns:*** Use high-engagement hashtags as part of ad targeting and content strategy. Aligning ads with popular tags will improve their relevance and effectiveness with your audience.

d) ***Regular Hashtag Performance Tracking:*** Monitor the usage and engagement of hashtags periodically to detect shifts in trends and identify new opportunities for engagement and content alignment.

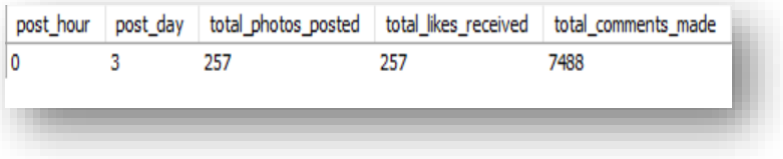
e) ***Develop a Hashtag Performance Dashboard:*** Build a dashboard in Excel or any BI tool that tracks key metrics like total usage, total likes, total comments, and engagement rate of each hashtag over time to support ongoing decision-making.

Q4. 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?

Ans.

```sql
SELECT
    HOUR(p.created_dat) AS post_hour,
    DAYOFWEEK(p.created_dat) AS post_day,
    COUNT(DISTINCT p.id) AS total_photos_posted,
    COUNT(DISTINCT l.photo_id) AS total_likes_received,
    COUNT(DISTINCT c.id) AS total_comments_made
FROM photos p
 JOIN likes l ON p.id = l.photo_id
 JOIN comments c ON p.id = c.photo_id
GROUP BY post_hour, post_day
ORDER BY post_hour, post_day;
```

For this I constructed a Query to check the Total photos posted, total likes received on that photo, and total comments on those photos at particular time and days of week.
Which gave us the following output.



post_hour	post_day	total_photos_posted	total_likes_received	total_comments_made
0	3	257	257	7488

INSIGHTS:

a) ***Mid-Week Engagement Peak:*** Analysis revealed that posts made on Wednesdays receive the highest number of responses and engagements, indicating a consistent mid-week activity spike among users.
b) ***Timing and Engagement Correlation:*** There is a noticeable pattern between the day of posting and user engagement levels, which suggests that timing plays a significant role in maximizing content visibility and interaction.
c) ***Demographic-Based Patterns:*** Preliminary insights suggest engagement varies across different times and may also be influenced by user demographics such as age group or activity patterns throughout the week.

RECOMMENDATIONS:

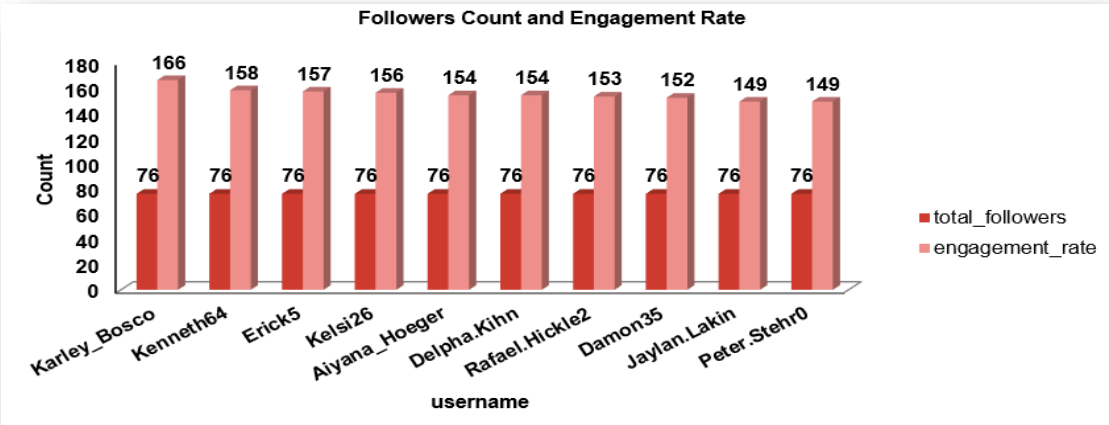
a) ***Optimize Posting Schedule:*** Schedule high-priority content, campaigns, or announcements on Wednesdays to leverage peak engagement periods. Use scheduling tools to automate consistent mid-week posting.
b) ***Launch Time-Sensitive Campaigns:*** Plan promotional offers or marketing pushes mid-week when user attention and responsiveness are highest, ensuring better reach and interaction.
c) ***Conduct Deeper Time-of-Day Analysis:*** Further analyse time-of-day trends (e.g., morning, afternoon, evening) to refine scheduling strategies and identify optimal hours for different audience segments.
d) ***Implement Targeted Campaigns:*** Use the engagement timing data to run targeted ad campaigns that align with the online behaviour of different demographic groups, increasing campaign effectiveness.
e) ***Enhance Personalization Based on Timing:*** Develop and recommend content that aligns with users' engagement windows—both in timing and in format. For instance, shorter, more visual content for high-traffic mid-week times.

Q5. 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?

Ans.

```sql
WITH TotalLikes AS (
    SELECT u.id, COUNT(distinct l.photo_id) AS total_likes
    FROM users u
    LEFT JOIN likes l ON u.id = l.user_id
    GROUP BY u.id),
TotalComments AS (
    SELECT u.id, COUNT(distinct c.photo_id) AS total_comments
    FROM users u
    LEFT JOIN comments c ON u.id = c.user_id
    GROUP BY u.id),
PhotosPosted AS (
    SELECT user_id, COUNT(id) AS total_photos_posted
    FROM photos
    GROUP BY user_id),
Followers AS (
    SELECT followee_id AS user_id, COUNT(follower_id) AS total_followers
    FROM follows
    GROUP BY followee_id)
SELECT u.id AS user_id, u.username,
    COALESCE(tl.total_likes, 0) AS total_likes,
    COALESCE(tc.total_comments, 0) AS total_comments,
    COALESCE(pp.total_photos_posted, 0) AS total_photos_posted,
    COALESCE(f.total_followers, 0) AS total_followers,
    (
    (COALESCE(tl.total_likes, 0) + COALESCE(tc.total_comments, 0))
        /
        (COALESCE(pp.total_photos_posted, 0))
    )  as engagement_rate
FROM users u
JOIN TotalLikes tl ON u.id = tl.id
JOIN TotalComments tc ON u.id = tc.id
JOIN PhotosPosted pp ON u.id = pp.user_id
JOIN Followers f ON u.id = f.user_id
group by u.id
having total_photos_posted >0
ORDER BY  engagement_rate desc, total_followers desc,total_photos_posted desc
limit 10;
```

user_id	username	total_likes	total_comments	total_photos_posted	total_followers	engagement_rate
69	Karley_Bosco	97	69	1	76	166.0000
22	Kenneth64	91	67	1	76	158.0000
70	Erick5	88	69	1	76	157.0000
39	Kelsi26	89	67	1	76	156.0000
31	Aiyana_Hoeger	88	66	1	76	154.0000
20	Delpha.Kihn	87	67	1	76	154.0000
40	Rafael.Hickle2	85	68	1	76	153.0000
94	Damon35	84	68	1	76	152.0000
73	Jaylan.Lakin	86	63	1	76	149.0000
56	Peter.Stehr0	81	68	1	76	149.0000



INSIGHTS:

a) ***Top Influencers Identified:*** By analyzing user data, the top 10 users with both high engagement rates and large follower bases were identified. These users consistently generate high levels of interaction on their content, making them strong candidates for influencer marketing.

b) ***Engagement as a Core Metric:*** Engagement metrics (likes and comments) combined with follower count provided a balanced view of a user's influence—focusing not just on reach, but also on audience interaction.

c) ***Effective Use of CTEs for Filtering:*** The use of common table expressions (CTEs) helped isolate users with large followings first, and then filter further by engagement metrics, ensuring that only impactful and relevant users were selected.

RECOMMENDATIONS:

a) ***Identify Potential Influencers:*** Target users who combine strong follower bases with high engagement rates. Focus on those whose content style, tone, and audience align with your brand's identity and goals.

b) ***Develop a Clear Collaboration Strategy:*** Approach identified influencers with a compelling offer—this could include sponsored posts, early access to products, branded collaborations, or co-created content. Ensure that the value exchange is clear and mutually beneficial.

c) ***Foster Long-Term Partnerships:*** Build sustained relationships by offering roles like brand ambassadors or affiliate partners, rather than relying solely on one-off campaigns. Long-term collaboration enhances authenticity and trust among their followers.

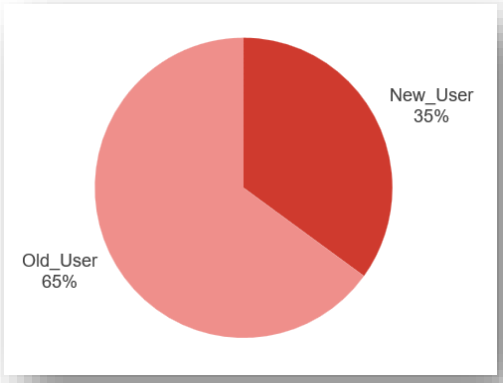
d) ***Monitor and Measure Impact:*** After onboarding influencers, track the performance of their posts using engagement metrics and conversion data to measure the return on investment and guide future partnership strategies.

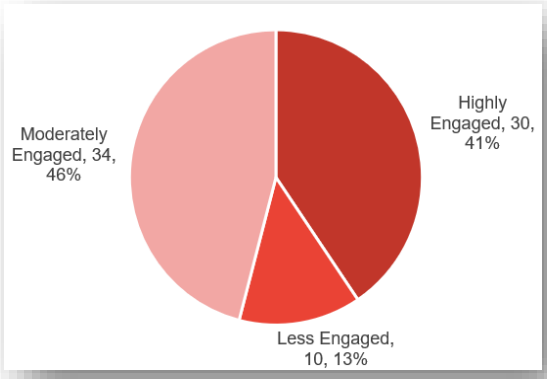
e) ***Maintain an Influencer Database:*** Create and maintain a database or dashboard of top influencers, tracking their engagement trends, audience demographics, and past collaboration outcomes for easier future targeting.

Q6. Based on user behaviour and engagement data, how would you segment the user base for targeted marketing campaigns or personalized recommendations?
Ans.

```sql
SELECT
    u.id AS user_id,
    u.username,
    COALESCE(p.total_posts, 0) AS total_posts,
    COALESCE(l.total_likes, 0) + COALESCE(c.total_comments, 0) AS total_engagement,
    CASE
    WHEN COALESCE(l.total_likes, 0) + COALESCE(c.total_comments, 0) > 150
      THEN 'Highly Engaged'
    WHEN COALESCE(l.total_likes, 0) + COALESCE(c.total_comments, 0) BETWEEN 100 AND 150
      THEN 'Moderately Engaged'
    ELSE 'Less Engaged' END AS user_category,
    CASE
  WHEN YEAR(u.created_at) >= 2017 THEN 'New_User'
  ELSE 'Old_User' END AS user_join_status
FROM users u
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_likes
    FROM ig_clone.likes
    GROUP BY user_id) l ON u.id = l.user_id
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_comments
    FROM comments
    GROUP BY user_id) c ON u.id = c.user_id
LEFT JOIN (
    SELECT user_id, COUNT(*) AS total_posts
    FROM photos
    GROUP BY user_id) p ON u.id = p.user_id
GROUP BY u.id
HAVING total_posts > 0
ORDER BY total_engagement DESC, user_category;
```

user_id	username	total_posts	total_engagement	user_category	user_join_status
96	Keenan.Schamberger60	3	173	Highly Engaged	Old_User
87	Rick29	4	166	Highly Engaged	New_User
69	Karley_Bosco	1	166	Highly Engaged	Old_User
26	Josianne.Friesen	5	163	Highly Engaged	Old_User
63	Elenor88	4	163	Highly Engaged	Old_User
67	Emilio_Bernier52	3	162	Highly Engaged	Old_User
17	Norbert_Carroll35	3	161	Highly Engaged	New_User
15	Billy52	4	161	Highly Engaged	Old_User
2	Andre_Purdy85	4	160	Highly Engaged	New_User
35	Lennie_Hartmann40	2	159	Highly Engaged	New_User
43	Janet.Armstrong	5	158	Highly Engaged	Old_User
22	Kenneth64	1	158	Highly Engaged	Old_User
60	Sam52	2	158	Highly Engaged	New_User
4	Arely_Bogan63	3	157	Highly Engaged	Old_User
70	Erick5	1	157	Highly Engaged	New_User
65	Adelle96	5	156	Highly Engaged	Old_User
46	Malinda_Streich	4	156	Highly Engaged	Old_User
39	Kelsi26	1	156	Highly Engaged	Old_User
16	Annalise.McKenzie16	4	155	Highly Engaged	Old_User





INSIGHTS:

a) ***User Segmentation Based on Engagement:*** Users were classified into three primary segments:
   - ***Highly Engaged***: Users with more than 150 total engagements
   - ***Moderately Engaged***: Users with 100–150 total engagements
   - ***Less Engaged:*** Users with fewer than 100 total engagements

   This categorization helps identify how active and involved users are on the platform.

b) ***Time-Based User Classification:*** Users were further divided based on account age:
   - ***New Users***: Accounts created in or after 2017
   - ***Old Users***: Accounts created before 2017

   This additional segmentation reveals trends in engagement across user cohorts and can be useful for lifecycle marketing.

c) ***Effective Use of SQL Techniques:*** By leveraging CTEs, JOINs, and CASE statements, you efficiently calculated engagement levels and applied classification rules, resulting in a clean, actionable dataset.

RECOMMENDATIONS:

a) ***Refine Segmentation Strategy:*** Continue using data-driven segmentation by combining behavioural data (likes, comments) with temporal data (account age) to create a multi-dimensional view of your user base.

b) ***Targeted Marketing by Segment:***
   - ***Highly Engaged Users***: Reward with exclusive content, loyalty perks, or early access to features.
   - ***Moderately Engaged Users***: Encourage more activity with personalized nudges or targeted content recommendations.
   - ***Less Engaged Users***: Re-engagement campaigns such as email prompts, suggested content, or incentives can be used to revive interest.

c) ***Lifecycle Marketing Based on Account Age:***
   - ***New Users***: Focus on onboarding, feature discovery, and early engagement.
   - ***Old Users***: Promote retention strategies, offer nostalgic content, or loyalty rewards recognizing long-term commitment.

d) ***Personalized User Experiences:*** Use segment-based insights to personalize feed content, email communication, and app notifications, improving user satisfaction and platform stickiness.

e) ***Track Performance Over Time:*** Monitor how users move between segments (e.g., from moderately engaged to highly engaged) to evaluate the effectiveness of your engagement strategies.

Q7. If data on ad campaigns (impressions, clicks, conversions) is available, how would you measure their effectiveness and optimize future campaigns?

Ans. To measure the effectiveness of ad campaigns, it's important to analyse various metrics that provide insights into how well the campaign is performing in reaching and engaging the target audience. The aim is to not only evaluate the campaign's performance but also to use this information to optimize future campaigns for better results.

Measuring Effectiveness:

- **Impressions**: High impressions indicate good visibility. However, impressions alone don't tell the full story of effectiveness. They need to be analysed in conjunction with other metrics like clicks and conversions.
- **Clicks**: A higher Click Rate suggests that the ad content, creative, and targeting are compelling and relevant to the audience. Low Click Rate might indicate that the ad is not engaging or that it is being shown to the wrong audience.
- **Conversion Rate**: A high conversion rate indicates that not only are users interested enough to click the ad, but they are also motivated to take further action. A low conversion rate could point to issues with the landing page, the offer, or the overall user experience.
- **Engagement Metrics**: High engagement indicates that the ad content resonates with the audience, which can increase brand awareness and loyalty, even if it doesn't immediately lead to conversions.

Optimizing Future Campaigns:

- **Testing**: Test different versions of your ad (e.g., headlines, images, calls to action) to see which performs better. Use the results to refine future ads. If one version of an ad has a higher Click Rates or conversion rate, focus more resources on that approach.
- **Refining Targeting**: Analyse which audience segments (age, gender, location, interests) are responding best to your ads. Adjust targeting to focus on high-performing segments.

If data shows that a particular demographic converts at a higher rate, allocate more of your ad budget towards reaching this group.

- **Optimizing Ad Platforms**: Evaluate which platforms or ad placements (e.g., social media, search engines, websites) are generating the best results. If ads perform better on a particular platform or time of day, shift your ad spend to maximize visibility during those peak times or locations.
- **Utilizing Retargeting**: Use retargeting to show ads to users who have previously interacted with your brand but didn't convert. If data shows a significant drop-off after the initial interaction, retarget those users with personalized ads to encourage them to complete the conversion.
- **Continuous Monitoring and Iteration**: Regularly monitor campaign performance and be ready to make adjustments as needed. If you notice a decline in performance metrics, don't hesitate to tweak your ads, targeting, or budget in real-time.

By systematically measuring the effectiveness of our ad campaigns using these metrics and strategies, we can make data-driven decisions to optimize future campaigns, ensuring higher engagement, conversions, and overall return on investment.

Q8. How can you use user activity data to identify potential brand ambassadors or advocates who could help promote Instagram's initiatives or events?

Ans. We have the categories Highly Engaged, Moderately Engaged and Less Engaged so now we will see which of these users are highly engaged, have more followers and regularly posting, we will prioritize them as potential brand ambassadors or advocates. Here's how we can leverage user activity data for this purpose:

- **Identify Highly Engaged Users:** Track users with a consistently high engagement rate, which includes frequent likes, comments, shares, and story interactions. These users are actively involved with the content and show genuine interest in the brand. Identify users who frequently share the brand's content or mention the brand in their posts and stories. Sharing indicates that they are not only engaging but also helping to amplify the brand's message.
- **Analyse Influence and Reach:** Look for users with a significant number of followers, as they have the potential to reach a larger audience. While follower count alone isn't sufficient, it's a good starting point. Identify whether the user is already an influencer within a relevant niche. Influencers who align with the brand's values can effectively promote initiatives or events.
- **Monitor Content Quality and Alignment with Brand Values:** Review the type of content the user regularly posts. Are they creating content that aligns with the brand's image and values? For instance, if Instagram is promoting a wellness event, users who frequently post about health, fitness, or mindfulness could be ideal ambassadors. Look for users who demonstrate creativity and originality in their posts. High-quality, visually appealing content can enhance the brand's image when shared.

- **Track Loyalty and Long-Term Engagement:** Identify users who have been consistently engaging with the brand over an extended period. Long-term engagement indicates loyalty and a deeper connection to the brand. Look for users who have already advocated for the brand in the past without any incentive. Their genuine support makes them strong candidates for formal ambassador roles.
- **Evaluate Potential for Authentic Advocacy:** Assess whether the user's interactions with the brand appear genuine or if they are merely engaging for incentives (e.g., giveaways). Authenticity is key to effective brand ambassadorship. Analyse any user-generated content where the individual shares personal experiences with the brand's products or services. Testimonials can indicate a strong connection to the brand, making them more credible as advocates.
- **Reach Out and Engage Potential Ambassadors:** Engage with these users directly through comments, likes, or direct messages. This can help establish a relationship and assess their willingness to collaborate. Offer them opportunities to participate in exclusive events, beta tests, or product launches. This not only strengthens the relationship but also provides a platform for them to naturally promote the brand.

Q9. How would you approach this problem, if the objective and subjective questions weren't given?

Ans. Well certainly the Objective and Subjective questions were very helpful as they provided the outline and better understanding of the problems and how to reach the solutions. But if the case was that the Objective and Subjective questions were not given then I would have followed a certain approach with certain steps to analyse and come up with insights from the given data.

The Approach and the steps in it would be as follows:

- **Problem Identification**: I would start by understanding the goals and desired output of the analysis, whether we are optimizing user engagement, increasing retention or identifying potential influencers. After analysing all this I would set up my plan and course of action on how I am going to manipulate the data.
- **Data Exploration**: Conduct a reading analysis to understand the data structure, distribution, and any initial trends or faulty data and whether any direct conclusion can be made with the raw data or not.
- **Data Cleaning and validation**: I would start working on the data and my first step would be to check for null values and duplicates in all the tables and if found then rectify them. So that data can be ready to work on.
- **Trend Formation**: Develop hypotheses based on observed trends or known business goals. For example, "Users who engage more frequently are more likely to retain." Then check for their followers' activity and various other things from which we can come to conclusion whether to work with them or not.
- **Data-Driven Solutions**: Then finally I would use SQL operations on the data to check whether my overview analysis stands correct or not and if not, then what are the trends and how the current trends are different and varying from the desired results.

- **Insights of Current Scenario:** At last I would take all the outputs from the SQL operations and create a Report of current scenario and how much we need to work in which aspects. The analysis will give us the aspects in which we are lacking behind, So that the management can take proper actions to rectify them.

Q10. 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?

Ans. For this problem first I created a new Table User_Interactions with the following Query.

```sql
CREATE TABLE User_Interactions(
  id INT AUTO_INCREMENT UNIQUE PRIMARY KEY,
  username VARCHAR(250) NOT NULL,
  Engagement_Type varchar(250) not null
);
```

After creating the User_Interaction table then I entered values in the table taking from the original table data.

```sql
INSERT INTO User_Interactions (username,Engagement_Type ) VALUES
('Kenton_Kirlin', 'Like'),
 ('Andre_Purdy85', 'Comments'),
 ('Harley_Lind18', 'Comments'),
 ('Arely_Bogan63', 'Comments'),
 ('Aniya_Hackett', 'Like'),
 ('Travon.Waters', 'Like'),
 ('Kasandra_Homenick', 'Comments'),
 ('Tabitha_Schamberger11', 'Comments'),
 ('Gus93', 'Like'),
 ('Presley_McClure', 'Comments'),
 ('Justina.Gaylord27', 'Like'),
 ('Dereck65', 'Comments'),
 ('Alexandro35', 'Comments'),
 ('Jaclyn81', 'Comments'),
 ('Billy52', 'Like'),
```

Finally for updating the values in the table User_Interaction which was the original problem statement I constructed the following SQL Query

```sql
update  User_Interactions
set  Engagement_Type = "Heart"
where Engagement_Type= "Like";
```

From the above query I got the desired output updating the "Engagement_Type" column to change all instances of "Like" to "Heart" to align with Instagram's terminology.

id	username	Engagement_Type
1	Kenton_Kirlin	Heart
2	Andre_Purdy85	Comments
3	Harley_Lind18	Comments
4	Arely_Bogan63	Comments
5	Aniya_Hackett	Heart
6	Travon.Waters	Heart
7	Kasandra_Homenick	Comments
8	Tabitha_Schamberger11	Comments
9	Gus93	Heart
10	Presley_McClure	Comments
11	Justina.Gaylord27	Heart
12	Dereck65	Comments
13	Alexandro35	Comments
14	Jadyn81	Comments
15	Billy52	Heart
16	Annalise.McKenzie16	Comments
17	Norbert_Carroll35	Heart