

---

## Social Media

### Objective and Subjective Questions Analysis

---

## Objective Questions:

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

Ans:

No, there are no duplicate or missing null values in every table.

If there any missing value occurs then will handle through 2 ways

- Delete the missing or null values.
- Use coalesce() to fill those NULL values with 0.

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

Ans:

Need to aggregate the number of posts, likes, and comments per user from the respective tables (photos, likes, and comments).

**Query**

```
SELECT u.id AS user_id, u.username,
       COALESCE(COUNT(DISTINCT p.id), 0) AS post_count,
       COALESCE(COUNT(DISTINCT l.photo_id), 0) AS like_count,
       COALESCE(COUNT(DISTINCT c.id), 0) AS comment_count
FROM users u
LEFT JOIN photos p ON p.user_id = u.id
LEFT JOIN likes l ON l.user_id = u.id
LEFT JOIN comments c ON c.user_id = u.id
GROUP BY u.id;
```

**Query Result:**

user_id	username	post_count	like_count	comment_count
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

#### Suggestions:

- Identify users with many posts but low engagement (low engagement\_ratio) and consider offering content suggestions or engagement strategies, such as prompts for interaction or community features.
- Users with high engagement (likes and comments) relative to their post count can be encouraged to create more content through rewards or special recognition.

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

Ans:

Query:

```
SELECT
    AVG(tag_count) AS avg_tags_per_post
FROM (
    SELECT photo_id, COUNT(tag_id) AS tag_count
    FROM photo_tags
    GROUP BY photo_id
) AS tag_counts;
```

#### Query Result:

Avg\_tags\_per\_cost: 2.6368

#### Explanation:

- Subquery: For each photo\_id, count the number of associated tag\_id from photo\_tags.
- Main Query: Calculate the average of these tag counts to get the average number of tags per post.

#### Suggestions:

- I could extend the query to show the average tags per post **by user**. This will give insights into which users are utilizing tags more effectively to increase post visibility or engagement.

- This query helps to understand user behaviour around tagging and could identify power users or influencers who use tags strategically.

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

Ans:

Query:

```
SELECT u.user_id,
       (COALESCE(total_likes, 0) + COALESCE(total_comments, 0)) / COALESCE(total_posts, 1) AS engagement_rate,
       RANK() OVER (ORDER BY (COALESCE(total_likes, 0) + COALESCE(total_comments, 0)) / COALESCE(total_posts, 1) DESC) AS `rank`
FROM (
  SELECT p.user_id, COALESCE(SUM(likes_count), 0) AS total_likes, COALESCE(SUM(comments_count), 0) AS total_comments
  FROM photos p
  LEFT JOIN (
    SELECT photo_id, COUNT(*) AS likes_count FROM likes
    GROUP BY photo_id
  ) l ON p.id = l.photo_id
  LEFT JOIN (
    SELECT photo_id, COUNT(*) AS comments_count FROM comments
    GROUP BY photo_id
  ) c ON p.id = c.photo_id
  GROUP BY p.user_id
) AS u
JOIN (SELECT user_id, COUNT(*) AS total_posts
      FROM photos GROUP BY user_id
) AS p_count ON u.user_id = p_count.user_id
ORDER BY engagement_rate DESC;
```

Query Result:

user_id	engagement_rate	rank
55	75	1
73	73	2
48	71	3
22	70	4
94	68	5
87	68	5
69	68	5
18	67	8
43	66.8	9

Explanation:

- Total likes and comments: The first query aggregates the likes and comments each user has received on their posts.
- Total posts per user: The second query calculates the total number of posts each user has made.

- Engagement rate: In the final query, the engagement rate is calculated as the sum of likes and comments divided by the number of posts, and users are ranked based on this rate.

#### Suggestions:

- The query offers a good balance between likes and comments, making it valuable to measure overall user impact, not just based on likes but on user interactions as well (through comments).
- It provides a summary of user activity by combining the total likes and comments received by each user. This can be used to gauge user engagement on the platform.

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

Ans:

#### Query:

User with most followers:

```
SELECT
    followee_id AS user_id, COUNT(follower_id) AS followers_count
FROM follows
GROUP BY followee_id
ORDER BY followers_count DESC
LIMIT 1;
```

This query identifies the user with the most followers by counting how many users are following each followee\_id.

#### Query Result:

user_id	followers_count
1	77

Users with the most followings:

```
SELECT
    follower_id AS user_id, COUNT(followee_id) AS followings_count
FROM follows
GROUP BY follower_id
ORDER BY followings_count DESC
LIMIT 1;
```

This query identifies the user with the most followings by counting how many users each follower\_id is following.

**Query Result:**

user_id	followings_count
2	99

**Suggestions**

- The first query reveals the most followed user on the platform, which can indicate the most influential or popular account. This is useful for understanding which users are most prominent or influential in the community.
- The second query highlights the user who follows the most accounts. This user is likely highly engaged, interacting with many other users on the platform. It shows user behaviour in terms of connection-building or interest in other users' activities.

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

Ans:

**Query:**

```
SELECT
    users.id AS user_id, users.username,
    ROUND((SUM(likes_count + comments_count) / COUNT(photos.id)) * 100,2) AS avg_engagement_rate
FROM users
JOIN photos ON photos.user_id = users.id
LEFT JOIN (SELECT photo_id, COUNT(user_id) AS likes_count FROM likes GROUP BY photo_id) AS likes ON likes.photo_id = photos.id
LEFT JOIN (SELECT photo_id, COUNT(user_id) AS comments_count FROM comments GROUP BY photo_id) AS comments ON comments.photo_id = photos.id
GROUP BY users.id
ORDER BY avg_engagement_rate DESC;
```

**Query Result:**

user_id	username	avg_engagement_rate
55	Meggie_Doyle	7500
73	Jaylan.Lakin	7300
48	Granville_Kutch	7100
22	Kenneth64	7000
94	Damon35	6800
87	Rick29	6800
69	Karley_Bosco	6800
18	Odessa2	6700

**Explanation:**

- likes\_count: Number of likes per post.
- comments\_count: Number of comments per post.

- **avg\_engagement\_rate:** Total engagements per post for each user, expressed as a percentage.

#### Suggestions:

- Identifies users whose content generates high engagement (likes and comments), highlighting potential influencers or active users.
- The query shows the average engagement per post, which helps evaluate how well users' posts are performing in terms of interaction.

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

Ans:

#### Query:

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

#### Query Result:

user_id	username
1	Kenton_Kirlin
7	Kasandra_Homenick
23	Eveline95
25	Tierra.Trantow
29	Jaime53
34	Pearl7
45	David.Osinski47
49	Morgan.Kassulke

#### Explanation:

- LEFT JOIN ensures all users from the users table are returned, along with any corresponding entries from the likes table.
- WHERE l.user\_id IS NULL filters the result to show only users who have no entries in the likes table, meaning they haven't liked any posts.

#### Suggestions:

- This query helps identify inactive or less-engaged users who are not interacting with content by liking photos. This can indicate passive users who view content but do not engage.
- These users can be targeted with engagement strategies, such as special notifications, recommendations, or prompts to encourage interaction.

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

Ans:

Leveraging user-generated content like posts, hashtags, and photo tags can significantly enhance personalized and engaging ad campaigns. Here's how the Instagram clone data (users, posts, hashtags, and photo tags) can be utilized:

### 1. Personalized Ads Based on User Interests:

- Content: Analyze the posts users have liked or commented on.
- Action: Create targeted ads for products or services similar to the content they frequently engage with.

### 2. Trending Hashtags for Ad Campaigns:

- Content: Track which hashtags are most frequently used or engaged with by users.
- Action: Use these trending hashtags to create ad campaigns that feel relevant and timely.

### 3. Ad Campaigns Based on Follower Behaviour:

- Content: Monitor users' interactions with followers (likes, comments, follows).
- Action: Offer ads reflecting community or mutual interests within groups of followers.

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

Ans:

Yes, there are correlations between user activity levels and specific content types, such as photos, videos, and reels. Research and platform insights suggest that different content types engage users in varying ways. Here's how these correlations can influence content strategies:

#### 1. Photos:

- Users tend to engage with photos quickly, especially when the content is visually striking or aligned with their interests (e.g., visually appealing landscapes, lifestyle imagery).
- Use high-quality, relevant images to captivate viewers' attention and drive engagement.

#### 2. Videos:

- Videos, particularly short-form videos (e.g., reels), tend to have higher engagement levels. These formats can captivate users longer and drive interactions such as likes, shares, and comments.

#### 3. Reels:

- Platforms have seen a significant uptick in engagement with short-form content like reels, especially among younger audiences. They encourage frequent interaction and repeat visits.
- Include reels and other short-form content in your strategy to capitalize on their viral potential and user engagement.

## 10. Calculate the total number of likes, comments, and photo tags for each user.?

Ans:

Query:

```
SELECT u.id AS user_id, u.username,  
       COUNT(DISTINCT l.photo_id) AS total_likes,  
       COUNT(DISTINCT c.id) AS total_comments,  
       COUNT(DISTINCT pt.tag_id) AS total_tags  
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 u.id = c.user_id  
LEFT JOIN photo_tags pt ON p.id = pt.photo_id  
GROUP BY u.id;
```

Query Result:



user_id	username
1	Kenton_Kirlin
7	Kasandra_Homenick
23	Eveline95
25	Tierra.Trantow
29	Jaime53
34	Pearl7
45	David.Osinski47
49	Morgan.Kassulke

### Explanation:

- The query retrieves data for each user (u.id and u.username) and counts the distinct number of likes, comments, and tags associated with photos uploaded by the user.
- It uses LEFT JOIN to connect the users table with the photos, likes, comments, and photo\_tags tables, ensuring that even if there are no likes, comments, or tags, the user is still included in the results

### Suggestions:

- The query gives a summary of how active each user is based on the number of photos they have that were liked (total\_likes), how many comments they made (total\_comments), and the total tags associated with their photos (total\_tags).
- Users with high numbers of likes or comments indicate more engagement with their content.
- Users with many tags on their photos may have more visibility or engagement through different categories or topics.

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

Ans:

### Query Result

```
SELECT u.id AS user_id, u.username, COALESCE(SUM(l.total_likes), 0) + COALESCE(SUM(c.total_comments), 0) AS total_engagement
FROM users u
LEFT JOIN ( SELECT user_id, COUNT(photo_id) AS total_likes FROM likes
WHERE created_at >= DATE_SUB(CURDATE(), INTERVAL 1 MONTH) GROUP BY user_id
) l ON u.id = l.user_id
LEFT JOIN ( SELECT user_id, COUNT(id) AS total_comments FROM comments
WHERE created_at >= DATE_SUB(CURDATE(), INTERVAL 1 MONTH) GROUP BY user_id
) c ON u.id = c.user_id
GROUP BY u.id
ORDER BY total_engagement DESC;
```

### Query Result:

user_id	username	total_engagement
71	Nia_Haag	514
5	Aniya_Hackett	514
41	Mckenna17	514
21	Rocio33	514
66	Mike.Auer39	514
54	Duane60	514
24	Maxwell.Halvorson	514
14	Jaclyn81	514
57	Julien_Schmidt	514

#### Explanation:

- The query calculates the total engagement for each user, which is the sum of likes, comments, and shares over the past month. It uses `COALESCE` to handle cases where there are no likes, comments, or shares, ensuring a default value of 0.
- The `LEFT JOIN` is used to join the users table with subqueries that count likes, comments, and shares, ensuring that all users are included, even those without activity in the past month.

#### Suggestions:

- Tailor content to users' preferences to keep them actively engaged. Personalized recommendations, notifications, or content can significantly boost interactions such as likes and comments.
- Ensure that interacting with content (liking, commenting, or sharing) is easy and intuitive. A frictionless user interface can increase engagement rates and user satisfaction.

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

Ans:

Query:

```

WITH hashtag_avg_likes AS (
    SELECT pt.tag_id, t.tag_name, AVG(l.likes_count) AS avg_likes
    FROM photo_tags pt
    JOIN tags t ON pt.tag_id = t.id
    JOIN photos p ON pt.photo_id = p.id
    JOIN likes l ON p.id = l.photo_id
    GROUP BY pt.tag_id, t.tag_name
)
SELECT tag_name, avg_likes
FROM hashtag_avg_likes
ORDER BY avg_likes DESC
LIMIT 10;

```

### Query Result:

tag_name	avg_likes
photography	157.5725
landscape	154.373
sunset	152.4877
sunrise	146.8415
beach	142.3867
style	138.372
dreamy	136.1594
stunning	133.7281
fashion	129.8219
delicious	127.9542

### Explanation:

- This calculates the average number of likes (AVG(l.likes\_count)) for each hashtag (t.tag\_name) by joining the photo\_tags, photos, tags, and likes tables.
- The results are grouped by the tag\_name.
- The main query retrieves the hashtags and their average likes from the CTE, ordered by avg\_likes in descending order to get the most popular hashtags.
- LIMIT 10 ensures that only the top 10 hashtags with the highest average likes are returned.

### Suggestions:

- This query helps identify the most popular hashtags based on average likes, which can give insight into what type of content is engaging users the most.
- Analyze these hashtags further to understand why they attract higher engagement—whether due to their topical relevance, timing, or user preferences.

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

Ans:

Query:

```
SELECT f1.follower_id, f1.followee_id
FROM follows f1
WHERE EXISTS (
    SELECT 1
    FROM follows f2
    WHERE f1.follower_id = f2.followee_id
    AND f1.followee_id = f2.follower_id
);
```

Query Result:

follower_id	followee_id
3	2
4	2
5	2
6	2
8	2
9	2
10	2
11	2

Explanation:

- follows table: Contains two columns (follower\_id, followee\_id), representing who follows whom.
- EXISTS clause: The query checks if there is a reverse relationship for every follower\_id in the follows table (i.e., if the followee\_id has also followed back the follower\_id).

Suggestions:

- This query helps identify mutual followers, showing how many users reciprocate follow actions.
- Target users who follow each other for potential friend suggestions or mutual-interest recommendations.

## 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?

Ans:

Users with high levels of engagement (e.g., likes, comments, and follows) are typically the most loyal or valuable. They regularly interact with content, indicating strong interest and a deeper connection with the platform or brand.

Query:

```
SELECT
    u.id AS user_id, u.username,
    COALESCE(SUM(l.total_likes), 0) AS total_likes,
    COALESCE(SUM(c.total_comments), 0) AS total_comments,
    COALESCE(SUM(f.total_follows), 0) AS total_follows,
    (COALESCE(SUM(l.total_likes), 0) + COALESCE(SUM(c.total_comments), 0) + COALESCE(SUM(f.total_follows), 0)) AS total_engagement
FROM users u
> LEFT JOIN ( SELECT user_id, COUNT(*) AS total_likes
    FROM likes WHERE created_at >= DATE_SUB(CURDATE(), INTERVAL 1 MONTH) GROUP BY user_id
~ ) l ON u.id = l.user_id
> LEFT JOIN ( SELECT user_id, COUNT(*) AS total_comments
    FROM comments WHERE created_at >= DATE_SUB(CURDATE(), INTERVAL 1 MONTH) GROUP BY user_id
~ ) c ON u.id = c.user_id
> LEFT JOIN (
    SELECT follower_id AS user_id, COUNT(*) AS total_follows
    FROM follows WHERE created_at >= DATE_SUB(CURDATE(), INTERVAL 1 MONTH) GROUP BY follower_id
~ ) f ON u.id = f.user_id
GROUP BY u.id
ORDER BY total_engagement DESC;
```

Query Result:

user_id	username	total_likes	total_comments	total_follows	total_engagement
71	Nia_Haag	257	257	99	613
5	Aniya_Hackett	257	257	99	613
41	Mckenna17	257	257	99	613
21	Rocio33	257	257	99	613
66	Mike.Auer39	257	257	99	613
54	Duane60	257	257	99	613
24	Maxwell.Halvorson	257	257	99	613
14	Jaclyn81	257	257	99	613
57	Julien_Schmidt	257	257	99	613

Explanation:

- **User Engagement:** This query calculates the total number of likes, comments, and follows each user has received in the past month.

- **COALESCE & Aggregation:** The COALESCE function ensures that users with no activity in a given category get a 0 count instead of NULL. The total engagement is the sum of these values.
- **Sorting by Engagement:** The users are sorted by total\_engagement in descending order, with the most active users at the top.

#### **Approach to Reward or Incentivize Loyal Users:**

- **Exclusive Content Access:** Provide top users with early access to new features, content, or promotions.
- **Recognition:** Feature the most loyal users in a "Top Users" list or provide badges to signify their contribution.
- **Discounts or Gifts:** Offer special discounts, exclusive offers, or rewards (e.g., merchandise) for consistent engagement.

#### **Criteria for Loyalty:**

- **High Total Engagement:** Users who consistently like, comment, and follow others demonstrate a strong connection with the platform.
- **Activity Over Time:** Consider consistent activity over time, not just one-off high engagement

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

Ans:

#### **Query:**

```
SELECT u.id, u.username
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
WHERE (p.created_at IS NULL OR p.created_at < NOW() - INTERVAL 30 DAY)
AND (l.created_at IS NULL OR l.created_at < NOW() - INTERVAL 30 DAY)
AND (c.created_at IS NULL OR c.created_at < NOW() - INTERVAL 30 DAY);
```

#### **Query Result:**

id	username
7	Kassandra_Homenick
25	Tierra.Trantow
34	Pearl7
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
68	Franco_Keebler64
74	Hulda.Macejkovic
80	Darby_Herzog

#### Explanation:

- **Inactive Users:** The query identifies users who haven't posted, liked, or commented in the last 30 days.
- **Left Join:** Ensures that user with no activity records are included.
- **Filter:** The conditions filters users who have no engagement within the last 30 days.

#### Strategies to Re-engage Inactive Users:

- **Survey the users:** Identify why they became inactive by sending surveys. Understanding their reasons can help you address concerns effectively.
- **Offer personalized incentives:** Provide exclusive deals or time-sensitive discounts to encourage them to return.
- **Segment and send tailored emails:** Use segmentation to send personalized messages, reminding users about features or benefits they might be missing.
- **Create a sense of community:** Foster a community where users can feel part of a group, encouraging them to participate.
- **Use multiple communication channels:** Reach out through social media, SMS, or even phone calls to make it more personal and direct.

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

Ans:

Query:

```

Select tag_id, tag_name, count(tag_name) as count_tage, count(l.photo_id) as likes_photo, count(c.photo_id) as comments_photo from tags t
left join photo_tags p on t.id = p.tag_id
left join likes l on l.photo_id = p.photo_id
left join comments c on c.photo_id = p.photo_id

group by tag_id, tag_name
order by count(tag_name) desc
limit 5

```

### Query Result:

tag_id	tag_name	count_tage	likes_photo	comments_photo
21	smile	59199	59199	59199
20	beach	41935	41935	41935
17	party	38939	38939	38939
13	fun	37145	37145	37145
5	food	24619	24619	24619
11	lol	23643	23643	23643
18	concert	23065	23065	23065

### Explanation:

- Joins: We join the tags table with photo\_tags, photos, likes, and comments to calculate engagement metrics (likes + comments).
- Aggregation: The query calculates the total number of likes and comments for each hashtag.
- Order and Limit: Results are ordered by total engagement in descending order, and the top 10 hashtags with the highest engagement are displayed.

### Analyzing hashtags:

In the above query result smile, beach, party, fun and food has the highest engagement and provides valuable insights for shaping content strategy and ad campaigns:

#### 1. Content Strategy:

- Focus content creation around these themes. For instance, promoting a joyful, fun atmosphere with visuals that represent parties, food, or beach scenes will attract higher engagement.
- Leverage user-generated content where people naturally tag these themes, encouraging more interactions.

#### 2. Ad Campaigns:

- Tailor ad creatives to incorporate these popular hashtags, making the campaigns resonate with user preferences. Ads featuring "beach" and



"party" visuals can boost the reach among audiences interested in lifestyle and leisure.

- Use these hashtags in paid promotions to target specific user groups who engage with similar content, increasing click-through rates and conversions.

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

Ans:

Query:

```
SELECT
    u.id AS user_id,
    u.username,
    HOUR(l.created_at) AS engagement_hour,
    COUNT(l.user_id) AS total_likes,
    COUNT(c.user_id) AS total_comments
FROM users u
LEFT JOIN likes l ON u.id = l.user_id
LEFT JOIN comments c ON u.id = c.user_id
GROUP BY u.id, HOUR(l.created_at)
ORDER BY total_likes DESC, total_comments DESC;
```

Query Result:

user_id	username	engagement_hour	total_likes	total_comments
36	Ollie_Ledner37	17	66049	66049
71	Nia_Haag	17	66049	66049
57	Julien_Schmidt	17	66049	66049
5	Aniya_Hackett	17	66049	66049
24	Maxwell.Halvorson	17	66049	66049
41	Mckenna17	17	66049	66049
21	Rocio33	17	66049	66049
14	Jaclyn81	17	66049	66049

Explanation:

- **LEFT JOIN:** We join the tags table with users, likes, and comments to calculate engagement metrics (likes + comments).

- Aggregation: Query calculate the number of total\_likes, total\_comments and engagement hours.
- Order : Results are ordered by total like and total comments in descending order.

Analyzing user engagement based on demographics and posting times can reveal significant trends that help tailor marketing strategies. Some possible patterns could include:

#### **Demographic Insights:**

- Age groups, locations, or gender might engage with certain content more often.
- Example: Younger users may engage more with fun, trendy, and lifestyle content, while older users may prefer informative or professional content.

#### **Posting Times:**

- Engagement might vary based on the time of day or week.
- Example: Posts made during peak hours could attract higher engagement, while off-peak times might work for specific user types

#### **Marketing Campaign Implications:**

- **Targeting:** Marketing campaigns can be adjusted to post at times when the target audience is most active. Demographics can be segmented to create more personalized ads (e.g., fun content for younger users, informative posts for professionals).
- **Content Personalization:** By analyzing which demographic engages with which type of content, brands can create more relevant and effective ads that resonate with their specific audience.

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

**Ans:**

**Query:**

```

> with cte as (
  SELECT p.user_id, COALESCE(SUM(likes_count), 0) AS total_likes, COALESCE(SUM(comments_count), 0) AS total_comments
  FROM photos p
  LEFT JOIN ( SELECT photo_id, COUNT(*) AS likes_count FROM likes GROUP BY photo_id
  ) l ON p.id = l.photo_id
  LEFT JOIN ( SELECT photo_id, COUNT(*) AS comments_count FROM comments GROUP BY photo_id
  ) c ON p.id = c.photo_id GROUP BY p.user_id order by total_likes desc, total_comments desc
),
cte2 as (
  SELECT follower_id AS user_id, COUNT(followee_id) AS followings_count
  FROM follows
  GROUP BY follower_id ORDER BY followings_count DESC
)
Select cte.user_id, u.username, total_likes, total_comments, followings_count from cte inner join cte2 on cte.user_id = cte2.user_id
inner join users u on u.id = cte.user_id
order by total_likes desc, total_comments desc, followings_count desc

```

### Query result:

user_id	username	total_likes	total_comments	followings_count
52	Zack_Kemmer93	182	151	99
13	Alexandro35	181	148	99
43	Janet.Armstrong	180	154	99
65	Adelle96	179	142	99
78	Colten.Harris76	177	143	99
33	Yvette.Gottlieb91	175	141	99

### Explanation:

CTE (Common Table Expressions):

- The first CTE calculates total likes and comments for each user's photos.
- The second CTE calculates how many users each influencer is following (followings\_count).

Main Query: Joins the CTE results with user data to rank influencers based on total likes, total comments, and followings count, identifying the most engaging users.

The ideal candidates for influencer marketing campaigns are users with high levels of engagement (likes, comments) and a substantial number of followings. These metrics indicate active users who generate and interact with content frequently, making them effective influencers.

For example, **Zack\_Kemmer93** has 182 likes, 151 comments, and 99 followings, which makes him an excellent candidate. Similarly, **Alexandro35** and **Janet.Armstrong** also have high engagement levels, placing them among the top candidates.

### **Approach for Collaboration:**

1. **Initial Contact:** Reach out with personalized messages to explain why they were selected based on their high engagement rates and relevance to your brand.
2. **Offer Value:** Provide attractive incentives like exclusive deals, early access to products, or financial compensation.
3. **Define Expectations:** Clearly outline the campaign goals, expected content types, and posting schedules.
4. **Foster Long-Term Relationships:** Build ongoing partnerships to keep influencers engaged and create authentic promotions.

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

**Ans:**

you can segment users into various groups based on their behaviour and engagement patterns. Here's how you might define and segment your user base:

1. **Engaged Users:**
  - High number of likes, comments, and posts.
  - These users frequently interact with content and are prime candidates for personalized recommendations or promotional content
2. **Influencers:**
  - Users with a high number of followers and engagement (likes, comments on their posts).
  - They are ideal for influencer campaigns or exclusive product recommendations.
3. **Content Consumers:**
  - Users who like and comment on many posts but do not create much content themselves.
  - Target them with content recommendations or curated content newsletters.
4. **Inactive Users:**
  - Users with low or no interaction (few or no likes, comments, or posts).
  - These users can be re-engaged through email campaigns or special offers.
5. **Topic-Specific Enthusiasts:**

- Users who frequently interact with content that is tagged with specific categories (e.g., specific hashtags or tags).
- These users can receive targeted product recommendations based on their interest areas.

### Query:

```

> WITH engagement AS (
  SELECT u.id AS user_id, u.username, COALESCE(SUM(l.likes_count), 0) AS total_likes,
    COALESCE(SUM(c.comments_count), 0) AS total_comments, COUNT(p.id) AS total_posts, COUNT(DISTINCT f.follower_id) AS followers_count,
    COUNT(DISTINCT f.followee_id) AS followings_count
  FROM users u LEFT JOIN photos p ON u.id = p.user_id
  LEFT JOIN (SELECT photo_id, COUNT(*) AS likes_count FROM likes GROUP BY photo_id) l ON p.id = l.photo_id
  LEFT JOIN (SELECT photo_id, COUNT(*) AS comments_count FROM comments GROUP BY photo_id) c ON p.id = c.photo_id
  LEFT JOIN follows f ON u.id = f.followee_id
  GROUP BY u.id
)
SELECT user_id, username,
  CASE
    WHEN total_likes > 50 AND total_comments > 50 AND total_posts > 5 THEN 'Engaged User'
    WHEN followers_count > 100 THEN 'Influencer'
    WHEN total_likes > 50 AND total_posts < 5 THEN 'Content Consumer'
    WHEN total_likes = 0 AND total_comments = 0 AND total_posts = 0 THEN 'Inactive User'
    ELSE 'Topic-Specific Enthusiast'
  END AS user_segment
FROM engagement
ORDER BY user_segment;

```

### Query result:

user_id	username	user_segment
1	Kenton_Kirlin	Engaged User
2	Andre_Purdy85	Engaged User
3	Harley_Lind18	Engaged User
4	Arely_Bogan63	Engaged User
86	Delfina_VonRueden68	Engaged User
6	Travon.Waters	Engaged User
87	Rick29	Engaged User
8	Tabitha_Schamberger11	Engaged User

### Explanation:

#### Engagement CTE:

- This calculates the total number of likes, comments, posts, and followers for each user.
- It aggregates the data from the likes, comments, photos, and follows tables.

#### Segmentation:

- Users are segmented based on their total engagement.

## 7. 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 and optimize future campaigns, you need to evaluate key metrics such as:

1. **Impression Count:** How many times an ad is shown.
2. **Click-Through Rate (CTR):** The ratio of clicks to impressions.
3. **Conversion Rate:** The percentage of users who clicked the ad and performed a desired action (conversion).
4. **Cost Per Click (CPC):** The average cost paid per click.
5. **Return on Investment (ROI):** The revenue generated compared to the ad spend.

With data on impressions, clicks, conversions, and costs, we can perform the following analysis:

Let's assume we have a table `ad\_campaigns` with the following structure:

**Query:**

```
CREATE TABLE ad_campaigns (  
  id INT AUTO_INCREMENT PRIMARY KEY,  
  campaign_name VARCHAR(255),  
  impressions INT,  
  clicks INT,  
  conversions INT,  
  cost DECIMAL(10, 2),  
  revenue DECIMAL(10, 2),  
  start_date DATE,  
  end_date DATE  
);
```

**Explanation:**

- **CTR (Click-Through Rate):** Calculated as  $(\text{clicks} / \text{impressions}) * 100$ . It measures the effectiveness of ads in generating clicks. A higher CTR indicates that the ad is appealing to users.
- **Conversion Rate:** Calculated as  $(\text{conversions} / \text{clicks}) * 100$ . It shows how well clicks are turning into actual conversions (such as purchases or sign-ups).
- **CPC (Cost Per Click):** Calculated as  $\text{cost} / \text{clicks}$ . This tells you how much you are spending for each click. Lower CPC means more efficient spending.
- **ROI (Return on Investment):** Calculated as  $((\text{revenue} - \text{cost}) / \text{cost}) * 100$ . This is the key metric to determine whether the campaign is profitable.

**Optimizing Future Campaigns:**

- **Improve CTR:** Experiment with different ad creatives, headlines, and targeting options to increase the click-through rate.
- **Boost Conversion Rate:** Focus on landing page optimization and better call-to-actions (CTAs) to ensure that more clicks convert into actions.
- **Lower CPC:** Try adjusting bidding strategies, targeting more relevant audiences, or using A/B testing for keywords and ad copy.
- **Maximize ROI:** Allocate more budget to high-ROI campaigns and reduce spend on underperforming ones.

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

Ans:

To identify potential brand ambassadors or advocates using user activity data on Instagram, we can analyze various metrics related to their engagement and influence within the platform. The key attributes that can help identify such individuals include:

### Key Metrics for Brand Ambassadors:

- **Follower Count:** Users with a large follower base, indicating they have reach and influence over others.
- **Engagement Rate:** High activity levels in terms of likes, comments, and posts, demonstrating active participation.
- **Content Creation:** Users who post a large number of photos and actively use tags, indicating that they frequently share content.
- **Interaction:** Users who receive a high number of likes and comments on their posts, showing that others engage with their content.
- **Consistency:** Users who post consistently and engage over time, indicating long-term commitment and presence on the platform.

Query:

```

SELECT
    u.id AS user_id,u.username,
    (SELECT COUNT(*) FROM follows WHERE followee_id = u.id) AS total_followers,
    (SELECT COUNT(*) FROM photos WHERE user_id = u.id) AS total_photos_posted,
    COALESCE((
        SELECT AVG(like_count) FROM (
            SELECT COUNT(l.user_id) AS like_count FROM likes l
            JOIN photos p ON l.photo_id = p.id
            WHERE p.user_id = u.id GROUP BY p.id
        ) AS like_stats), 0) AS avg_likes_per_photo,
    COALESCE((
        SELECT AVG(comment_count) FROM (SELECT COUNT(c.id) AS comment_count FROM comments c
            JOIN photos p ON c.photo_id = p.id
            WHERE p.user_id = u.id
            GROUP BY p.id
        ) AS comment_stats
    ), 0) AS avg_comments_per_photo,
    COALESCE((
        SELECT MAX(p.created_at) FROM photos p WHERE p.user_id = u.id
    ), u.created_at) AS last_activity_date

FROM users u
HAVING total_followers > 50 AND total_photos_posted > 5 AND avg_likes_per_photo > 30
ORDER BY avg_likes_per_photo DESC;

```

### Query Result:

user_id	username	total_followers	total_photos_posted	avg_likes_per_photo	avg_comments_per_photo	last_activity_date
77	Donald.Fritsch	77	6	36.3333	29	09-10-2024 17:05
23	Eveline95	77	12	35	27.4167	09-10-2024 17:05
58	Aurelie71	77	8	35	30.25	09-10-2024 17:05
29	Jaime53	77	8	33.875	28.625	09-10-2024 17:05
59	Cesar93	77	10	33.8	30.8	09-10-2024 17:05
88	Clint27	77	11	32.8182	27.1818	09-10-2024 17:05
86	Delfina_VonRueden68	77	9	31.6667	30.3333	09-10-2024 17:05

### Explanation:

- **Follower Count:** The subquery `SELECT COUNT(*) FROM follows WHERE followee_id = u.id` calculates how many users follow a particular user, which helps measure their influence.
- **Content Creation:** The number of photos a user has posted is calculated by counting rows in the `photos` table where `user_id = u.id`.
- **Average Likes and Comments:** These subqueries calculate the average likes and comments per photo, giving an idea of how engaging the user's content is.
- **Activity Level:** The most recent post's date (`MAX(p.created_at)`) indicates the user's recent activity.

This query will return a list of potential brand ambassadors ranked by their engagement metrics, including follower count, number of posts, average likes and comments per photo, and recent activity. These users can be targeted for promoting Instagram's initiatives or events based on their high engagement and influence.

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



Ans:

I would approach the problem in a structured manner. Here's a step-by-step outline of how I would tackle this:

**Identify Key Metrics:** Determine which metrics are crucial for evaluating user engagement, retention, and acquisition, such as:

- Number of photos posted
- Number of likes and comments per user
- Growth in followers over time
- Engagement rates (likes and comments per photo)

**Define Key Questions and Metrics:**

1. Engagement Metrics:
  - Average likes and comments per user and per photo.
  - User activity level (frequency of posting, commenting, and liking).
2. Acquisition Metrics:
  - New users gained per month.
  - Conversion rate from visitors to registered users.

**Data Analysis:**

1. Engagement Analysis:
  - Analyze likes and comments to identify trends. For example, which types of photos get the most engagement? What times/days see higher activity?
2. Retention Analysis:
  - Track the number of returning users and analyze their activity levels over time.
3. Acquisition Analysis:
  - Examine new user registrations and track growth over time. Analyze if new users engage similarly to older users.
4. Acquisition:
  - **Targeted Content Strategy:**
    - Identify high-performing content (based on likes/comments) and recommend strategies for encouraging similar content creation among users.
  - **User Engagement Initiatives:**
    - Propose initiatives such as challenges, contests, or campaigns to incentivize users to post more actively.

- **Retention Programs:**
  - Suggest loyalty programs or features that encourage users to log in daily or weekly, such as personalized content feeds based on their interaction history.
- **Acquisition Channels:**
  - Recommend marketing strategies focusing on platforms where potential users are most active, utilizing data from current user demographics and behaviour.

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

Ans:

To update the "Engagement\_Type" column in the "User\_Interactions" table and change all instances of "Like" to "Heart," you can use the SQL `UPDATE` statement. Here's how you can do it:

Query:

```
UPDATE User_Interactions
SET Engagement_Type = 'Heart'
WHERE Engagement_Type = 'Like';
```

**Explanation:**

- **UPDATE User\_Interactions:** This specifies the table you want to update, which in this case is "User\_Interactions."
- **SET Engagement\_Type = 'Heart':** This indicates that you want to change the value of the "Engagement\_Type" column to "Heart."
- **WHERE Engagement\_Type = 'Like':** This condition ensures that only the rows where "Engagement\_Type" is currently set to "Like" will be affected by the update.