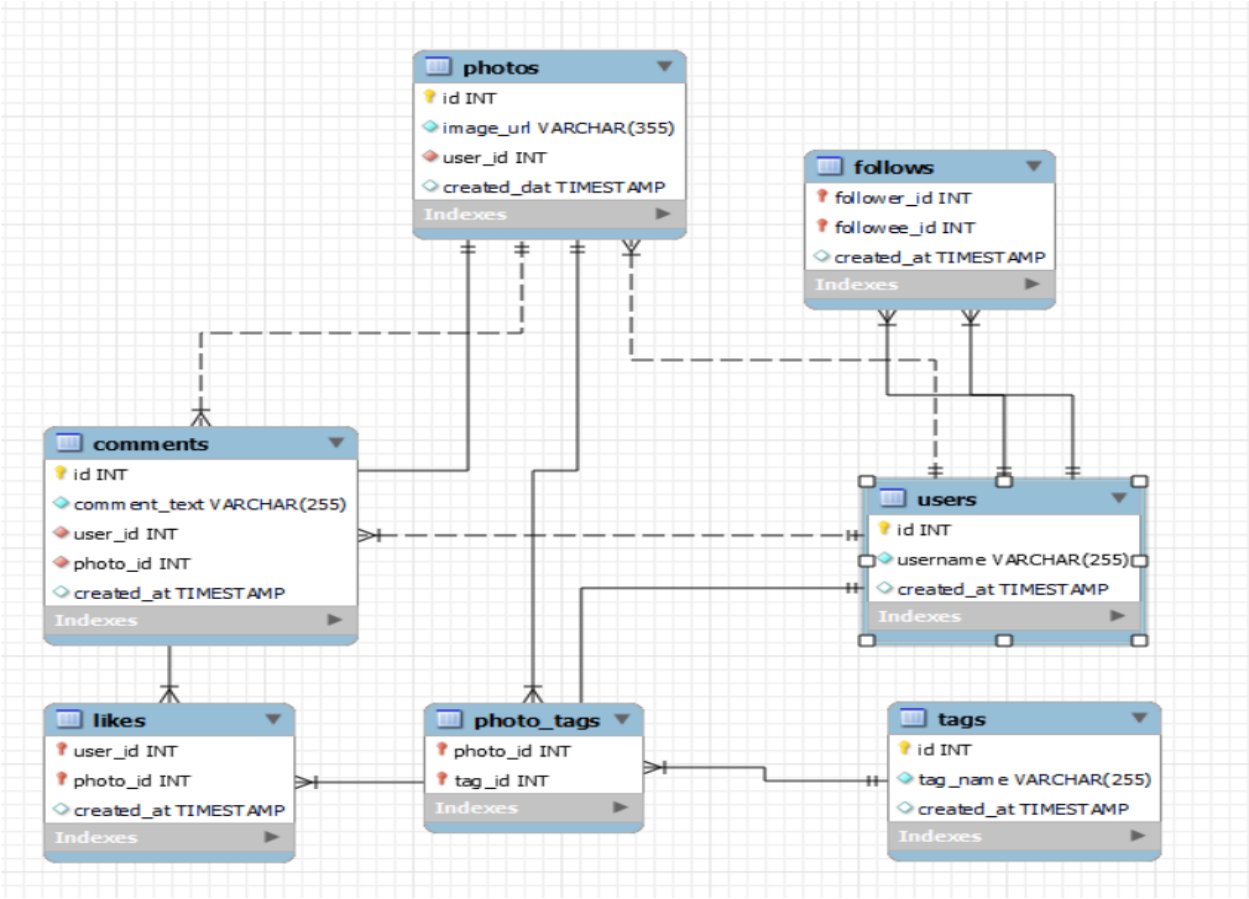
Social Media Analysis (META)

Project By - Aminesh Kumar Singh

SCHEMA



Data Description

- comments_id : unique identifier for each comment
- comment_text : text content of a given comment
- user_id : unique identifier for each user
- photo_id : unique identifier for each photo
- created_at : date of interaction in the form like, photos, tags
- follower_id : user_id of the follower for a certain user
- followee_id : user_id of followee for a certain user
- tag_id : unique identifier for each tag
- image_url : link to the image posted on the platform
- username : username chosen by the user

VIEW

On the Basis of Data Provided and Tasks to be performed, I Have decided to Create a View(Virtual Table) named - “user_summary” by Joining All the Tables and Summarizing/Aggregating them to derive KPIs that can be utilized at different levels to answer the questions. It Helps in removing redundancy and provides reusability, optimizing performance.

Related Query -

```
CREATE VIEW user_summary AS
  SELECT
    a.id AS UserID,
    a.username AS Username,
    COALESCE(COUNT(b.id), 0) AS Total_Photos,
    COALESCE(SUM(Likes), 0) AS Total_Likes,
    COALESCE(SUM(Comments), 0) AS Total_Comments,
    COALESCE(SUM(Tags), 0) AS Total_Tags,
    COALESCE(Posts_Liked, 0) AS Posts_Liked,
    COALESCE(Comments_Made, 0) AS Comments_Made,
    COALESCE(Total_followers, 0) AS Total_followers,
    COALESCE(Total_followings, 0) AS Total_followings
  FROM
    users a
      LEFT JOIN
    photos b ON a.id = b.user_id
      LEFT JOIN
    (SELECT
      photo_id, COUNT(*) AS Likes
    FROM
      likes
    GROUP BY photo_id) c ON b.id = c.photo_id
      LEFT JOIN
    (SELECT
      photo_id, COUNT(*) AS Comments
    FROM
      comments
    GROUP BY photo_id) d ON b.id = d.photo_id
      LEFT JOIN
    (SELECT
      photo_id, COUNT(*) AS Tags
    FROM
      photo_tags
    GROUP BY photo_id) e ON b.id = e.photo_id
      LEFT JOIN
    (SELECT
      user_id, COUNT(id) AS Comments_Made
```

```
    FROM
        comments
    GROUP BY user_id) f ON a.id = f.user_id
        LEFT JOIN
    (SELECT
        user_id, COUNT(photo_id) AS Posts_Liked
    FROM
        likes
    GROUP BY user_id) g ON a.id = g.user_id
        LEFT JOIN
    (SELECT
        followee_id AS id, COUNT(follower_id) AS Total_followers
    FROM
        follows
    GROUP BY followee_id) h ON a.id = h.id
        LEFT JOIN
    (SELECT
        follower_id AS id, COUNT(followee_id) AS Total_followings
    FROM
        follows
    GROUP BY follower_id) i ON a.id = i.id
GROUP BY a.id , a.username
```

The Output of the View is -

It gives the summary of user activity and user engagement, which can be consumed for answers several questions.

UserID	Username	Total_Photos	Total_Likes	Total_Comments	Total_Tags	Posts_Liked	Comments_Made	Total_followers	Total_followings
1	Kenton_Kirlin	5	168	142	18	0	0	77	0
2	Andre_Purdy85	4	127	119	13	94	66	76	99
3	Harley_Lind18	4	132	117	7	79	67	76	99
4	Arely_Bogan63	3	106	77	2	93	64	76	99
5	Aniya_Hackett	0	0	0	0	257	257	76	99
6	Travon.Waters	5	173	139	8	82	62	76	99
7	Kasandra_Homenick	0	0	0	0	0	0	77	0
8	Tabitha_Schamberger11	4	137	119	13	79	61	76	99
9	Gus93	4	130	126	11	85	60	76	99
10	Presley_McClure	3	105	90	10	87	63	76	99
11	Justina.Gaylord27	5	166	147	12	89	49	76	99
12	Dereck65	4	140	117	2	77	68	76	99
13	Alexandro35	5	181	148	7	93	58	76	99

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Objective Questions

*** Please Refer to SQL File for ALL the Related Queries used to answer these Questions*

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

Ans - There are total 7 tables in the database and each table have its primary key & definition of all the columns in all the tables has NOT NULL constraint. Hence there is no possibility of any duplicates or null values in the tables.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

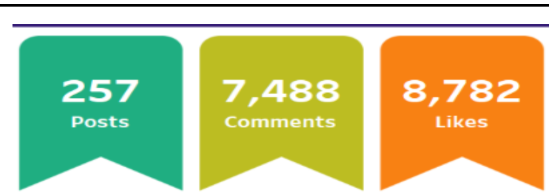
Ans - The overall user Activity is calculated using Three tables
photos, likes & comments.

<u>QUERY</u>

```
SELECT
    COUNT(a.id) AS Total_Photos,
    SUM(Likes) AS Total_Likes,
    SUM(Comments) AS Total_Comments
FROM photos a
LEFT JOIN
    (SELECT photo_id, COUNT(*) AS Likes
    FROM likes
    GROUP BY photo_id) b ON a.id = b.photo_id
LEFT JOIN
    (SELECT photo_id, COUNT(*) AS Comments
    FROM comments
    GROUP BY photo_id) c ON a.id = c.photo_id;
```

<u>OUTPUT</u>

Total_Photos	Total_Likes	Total_Comments
257	8782	7488



xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans - Average number of tags is calculated using two tables

photos and photo_tags

QUERY

```
SELECT
    ROUND(AVG(COALESCE(tags, 0)), 2) AS Avg_Tags_Per_Post
FROM
    photos a
        LEFT JOIN
    (SELECT
        photo_id, COUNT(tag_id) AS tags
    FROM
        photo_tags
    GROUP BY photo_id) b ON a.id = b.photo_id;
```

OUTPUT

Avg_Tags_Per_Post
1.95

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

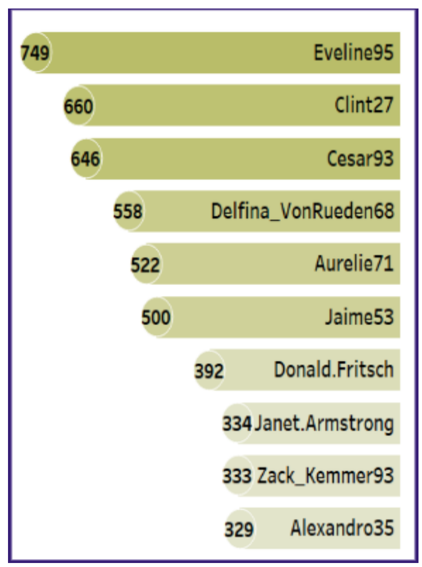
Ans - We are using the user_summary view to find total likes, comments on individual users level, then finding total engagement and using rank function to identify users with highest engagement rates.

<u>QUERY</u>

```
SELECT
    UserID,
    username,
    Total_likes,
    Total_Comments,
    (Total_likes + Total_Comments) AS Total_Engagement,
    RANK() OVER(ORDER BY (Total_likes + Total_Comments) DESC) AS Top_Ranked_Users
FROM
    user_summary;
```

<u>OUTPUT</u>

UserID	username	Total_likes	Total_Comments	Total_Engagement	Top_Ranked_Users
23	Eveline95	420	329	749	1
88	Clint27	361	299	660	2
59	Cesar93	338	308	646	3
86	Delfina_VonRueden68	285	273	558	4
58	Aurelie71	280	242	522	5
29	Jaime53	271	229	500	6
77	Donald.Fritsch	218	174	392	7
43	Janet.Armstrong	180	154	334	8
52	Zack_Kemmer93	182	151	333	9
13	Alexandro35	181	148	329	10
65	Adelle96	179	142	321	11



xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans - Using user_summary view to find followers and following count for each user and then to display users with highest number of followers and followings at top, sorted the output in descending order of sum total of followers and followings.

QUERY

```
SELECT
    UserID, username, Total_Followers, Total_Followings
FROM
    user_summary
ORDER BY (Total_Followers + Total_Followings) DESC;
```

OUTPUT

user_id	username	Total_Followers	Total_Followings
13	Alexandro35	76	99
2	Andre_Purdy85	76	99
3	Harley_Lind18	76	99
4	Arely_Bogan63	76	99
5	Aniya_Hackett	76	99
6	Travon.Waters	76	99
93	Willie_Leuschke	76	99
8	Tabitha_Schamberger11	76	99
9	Gus93	76	99
10	Presley_McClure	76	99
11	Justina.Gaylord27	76	99
12	Dereck65	76	99

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

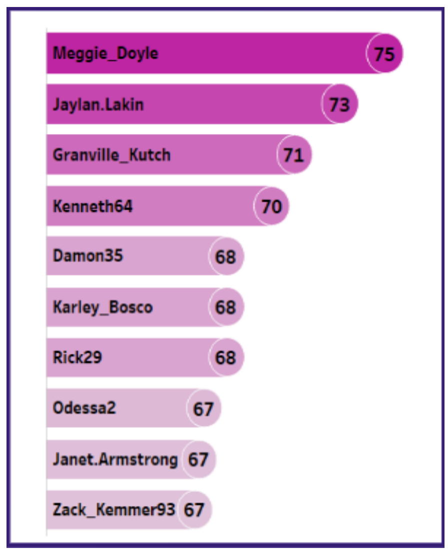
Ans - To calculate average engagement rate we have used user_summary view and calculated average likes per photo and average comment per photo by dividing them by total no. of photos. And then calculating average Engagement rate for individual user and user rank function to rank them from highest to lowest.

<u>QUERY</u>

```
SELECT
    UserID,
    username,
    Total_Photos,
    Total_Likes,
    Total_Comments,
    ROUND((Total_Likes + Total_Comments) / Total_Photos,2) AS Average_Engagement_Rate,
    RANK() OVER( ORDER BY (Total_Likes+Total_Comments)/Total_Photos DESC) AS Top_Users_Rank
FROM user_summary;
```

<u>OUTPUT</u>

UserID	username	Total_Photos	Total_Likes	Total_Comments	Average_Engagement_Rate	Top_Users_Rank
55	Meggie_Doyle	1	41	34	75.00	1
73	Jaylan.Lakin	1	38	35	73.00	2
48	Granville_Kutch	1	37	34	71.00	3
22	Kenneth64	1	39	31	70.00	4
94	Damon35	1	40	28	68.00	5
87	Rick29	4	140	132	68.00	5
69	Karley_Bosco	1	36	32	68.00	5
18	Odessa2	1	36	31	67.00	8
43	Janet.Armstrong	5	180	154	66.80	9
52	Zack_Kemmer93	5	182	151	66.60	10
46	Malinda_Streich	4	145	121	66.50	11
82	Aracely.Johnsto...	2	72	61	66.50	11
44	Seth46	4	146	118	66.00	13



XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

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

Ans - This was derived using users table and then eliminating all user_id from it that were present in the likes table.

QUERY

```
SELECT
    id AS UserID, username AS Users_Never_Liked_Any_Post
FROM users
WHERE
    id NOT IN (SELECT DISTINCT user_id
        FROM likes);
```

OUTPUT

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
86	Delfina_VonRuede...
88	Clint27
89	Jessyca_West
90	Esmeralda.Mraz57

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

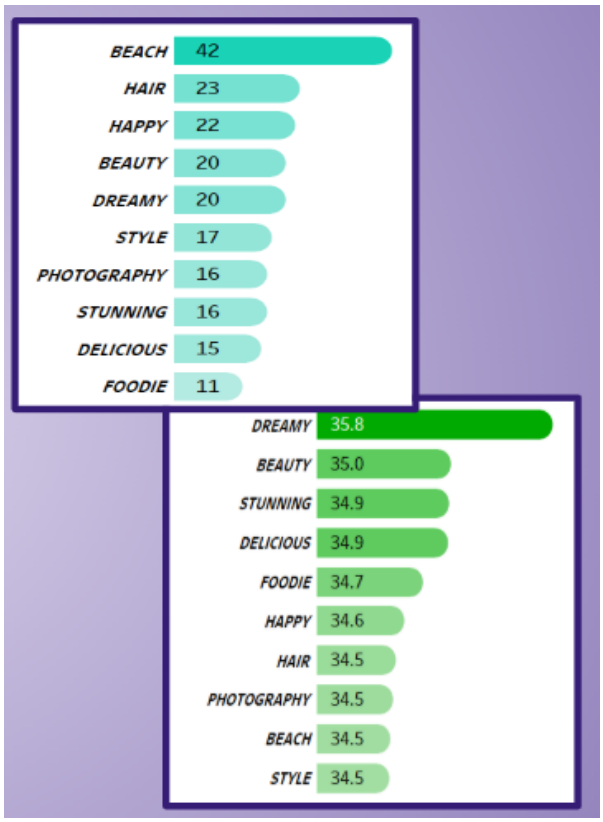
Ans - Insights from this table can help leverage user-generated content to create more personalized and engaging ad campaigns. (Table taken from Objective Q12)

QUERY

```
WITH tags_avg_likes AS (SELECT
    a.tag_id, ROUND(AVG(Total_likes),2) AS Avg_Likes,
    count(a.photo_id) as Times_Tag_Used
FROM photo_tags AS a
    INNER JOIN
    (SELECT  photo_id, COUNT(user_id) AS Total_likes
    FROM likes
    GROUP BY photo_id) b
    ON a.photo_id = b.photo_id
    GROUP BY tag_id
)

SELECT  id AS tag_id, tag_name, Avg_Likes, Times_Tag_Used
FROM tags a
    EFT JOIN
    tags_avg_likes b ON id = tag_id
ORDER BY Avg_Likes DESC;
```

OUTPUT

tag_id	tag_name	Avg_Likes	Times_Tag_Used
10	dreamy	35.75	20
8	beauty	34.95	20
9	stunning	34.94	16
7	delicious	34.93	15
6	foodie	34.73	11
12	happy	34.59	22
15	hair	34.52	23
2	photography	34.50	16
20	beach	34.48	42
14	style	34.47	17
21	smile	34.46	59
18	concert	34.38	24
13	fun	34.24	38
11	lol	34.21	24
1	sunset	34.21	19
19	drunk	34.05	19
17	party	33.92	39
5	food	33.83	24
3	sunrise	33.76	17
16	fashion	33.68	19
4	landscape	33.59	17



Suggestions

- By identifying popular hashtags to see what themes or topics users are engaging with. This insight allows us to create ads that resonate with these themes, increase engagement.
- Can 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.
- By tracking trending hashtags, we can keep ads aligned with current trends, ensuring they're fresh and engaging for the audience.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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 - There can be potential correlation between user activity levels and specific content types. By analyzing user activity levels and their interaction with different content types, we can gather valuable insights that help:

- Customise content creation strategies to match user preferences.
- Create targeted campaigns to boost engagement.
- Leverage content types that align with high user activity for more effective advertising.

However, the current database is limited to photos and does not cover other post types. Hence, with the provided database, it would be difficult to gain actual insights into the correlation between user activity levels and specific content types.

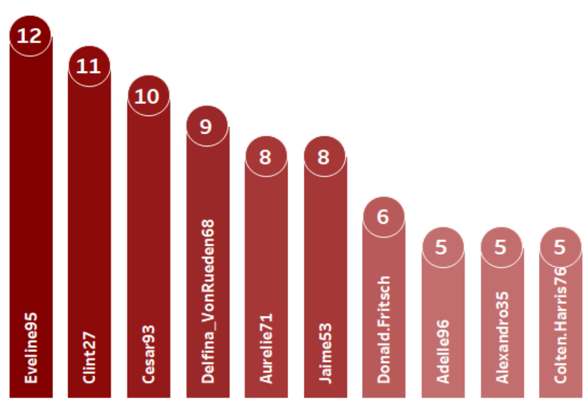
QUERY

```sql
select Username, Total_Photos,
Total_Photos+Total_Likes+Total_Comments+Total_Tags+
Posts_Liked+Comments_Made+Total_Followers+Total_Followings
as Engagement_Score
from user_summary
order by Total_Photos desc
limit 10;
```

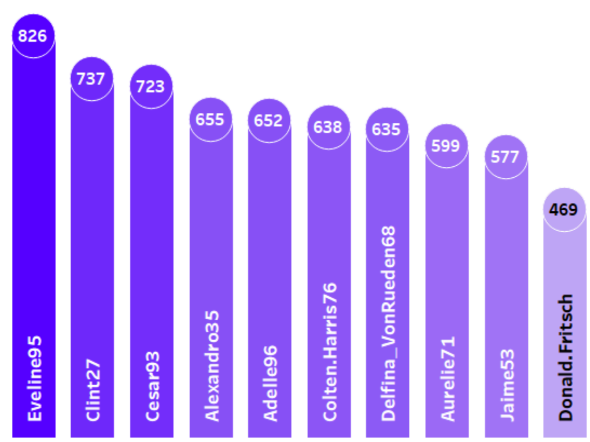
OUTPUT

Username	Total_Photos	Engagement_Score
Eveline95	12	862
Clint27	11	774
Cesar93	10	749
Delfina_VonRueden68	9	661
Jaime53	8	600
Aurelie71	8	627
Donald.Fritsch	6	490
Kenton_Kirlin	5	410
Mariano_Koch3	5	398
Kathryn80	5	637

Total Photos



Engagement Score



xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

10. Calculate the total number of likes, comments, and photo tags for each user.
Ans - We have directly consumed user_summary view and selected needed columns to be displayed.

QUERY

```
SELECT
    UserID,
    username,
    Total_Photos,
    Total_Likes,
    Total_Comments,
    Total_Tags
FROM
    user_summary;
```

OUTPUT

UserID	username	Total_Photos	Total_Likes	Total_Comments	Total_Tags
1	Kenton_Kirlin	5	168	142	18
2	Andre_Purdy85	4	127	119	13
3	Harley_Lind18	4	132	117	7
4	Arely_Bogan63	3	106	77	2
5	Aniya_Hackett	0	0	0	0
6	Travon.Waters	5	173	139	8
7	Kasandra_Homenick	0	0	0	0
8	Tabitha_Schamberger11	4	137	119	13
9	Gus93	4	130	126	11
10	Presley_McClure	3	105	90	10
11	Justina.Gaylord27	5	166	147	12
12	Dereck65	4	140	117	2
13	Alexandro35	5	181	148	7
14	Jaclyn81	0	0	0	0
15	Billy52	4	129	115	4
16	Annalise.McKenzie16	4	137	126	4

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans - Created CTE

Used users table as base table.

Left joined likes table grouped at user_id & Month_Year, to calculate total likes.

Then Left joined the comments table grouped at user_id & Month_Year, to calculate total comments.

Finally, used rank function to rank users based on total_engagement(Total_Likes+Total_Comments) over a month.

Since data is available only for a single month. Only one month's ranking is available in the output.

Displaying top 20 users.

QUERY

```
with Posts_Liked as (select user_id, count(*) as Engagement , DATE_FORMAT(created_at,'%y-%c') as MY from likes
group by user_id, MY ),

Comments_Made as (select user_id, count(*) as Engagement , DATE_FORMAT(created_at,'%y-%c') as MY from comments
group by user_id, MY ),

Total_Likes as (select b.user_id, count(*) as Engagement,  DATE_FORMAT(created_at,'%y-%c') as MY from likes a
left join photos b
on a.photo_id=b.id
group by b.user_id, MY ),

Total_Comments as (select b.user_id, count(*) as Engagement,  DATE_FORMAT(created_at,'%y-%c') as MY from comments a
left join photos b
on a.photo_id=b.id
group by b.user_id, MY ),

Total_Engagement as (select * from Posts_Liked
union all
select * from Comments_Made
union all
select * from Total_Likes
union all
select * from Total_Comments)

select MY, username, sum(Engagement) as Total_Engagement,
rank() over(partition by MY order by sum(Engagement) desc) as Monthly_Top_Users
from Total_Engagement a
left join users b
on a.user_id=b.id
group by MY, user_id;
```

OUTPUT

MY	username	Total_Engagement	Monthly_Top_Users
25-1	Eveline95	749	1
25-1	Clint27	660	2
25-1	Cesar93	646	3
25-1	Delfina_VonRueden68	558	4
25-1	Aurelie71	522	5
25-1	Maxwell.Halvorson	514	6
25-1	Ollie_Ledner37	514	6
25-1	Aniya_Hackett	514	6
25-1	Janelle.Nikolaus81	514	6
25-1	Duane60	514	6
25-1	Mckenna17	514	6
25-1	Jaclyn81	514	6
25-1	Leslie67	514	6
25-1	Nia_Haag	514	6
25-1	Mike.Auer39	514	6
25-1	Rocio33	514	6

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

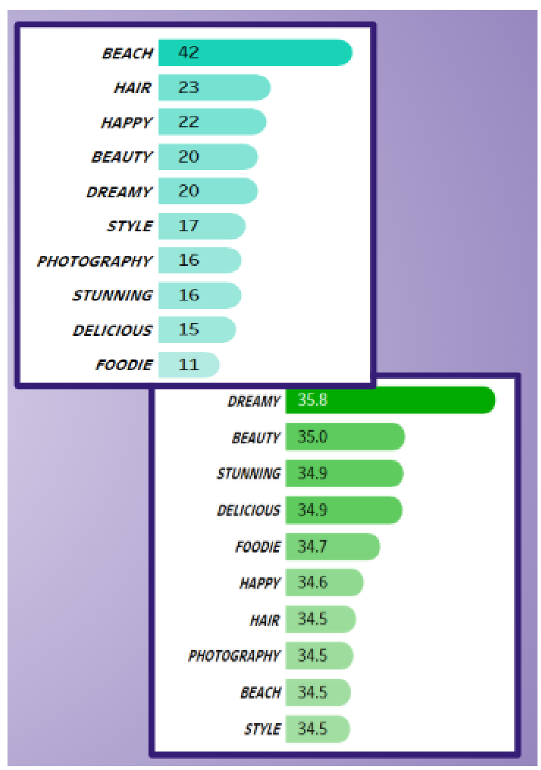
12. Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.
Ans - Created a CTE to find average likes for each hashtags named tags_avg_likes.
We have a user photo_tags table as base table and left join it with an intermediate table which is an aggregation of likes table that gives total likes for each photo.
Performed left join on photo_id and then calculated average of total likes by grouping them on tag_id, also calculated no. of times each tag is used using count function.
And finally we have performed a join of this CTE on the tags table to extract tag_name and average likes for it. Sorted result in descending order of avg likes.

QUERY

```
WITH tags_avg_likes AS (SELECT
    a.tag_id, ROUND(AVG(Total_likes),2) AS Avg_Likes, count(a.photo_id) as Times_Tag_Used
FROM photo_tags AS a
        INNER JOIN
    (SELECT  photo_id, COUNT(user_id) AS Total_likes
    FROM likes
    GROUP BY photo_id) b ON a.photo_id = b.photo_id
GROUP BY tag_id
)

SELECT  id AS tag_id, tag_name, Avg_Likes, Times_Tag_Used
FROM tags a
     LEFT JOIN
    tags_avg_likes b ON id = tag_id
ORDER BY Avg_Likes DESC;
```

OUTPUT

tag_id	tag_name	Avg_Likes	Times_Tag_Used
10	dreamy	35.75	20
8	beauty	34.95	20
9	stunning	34.94	16
7	delicious	34.93	15
6	foodie	34.73	11
12	happy	34.59	22
15	hair	34.52	23
2	photography	34.50	16
20	beach	34.48	42
14	style	34.47	17
21	smile	34.46	59
18	concert	34.38	24
13	fun	34.24	38
11	lol	34.21	24
1	sunset	34.21	19
19	drunk	34.05	19
17	party	33.92	39
5	food	33.83	24
3	sunrise	33.76	17
16	fashion	33.68	19
4	landscape	33.59	17



xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans - Condition for this problems are -

i) Both users must be following each other.

ii) And that user must be a followee first then a follower for a given pair of users.

i.e. user_id of followee with earliest date for given pair.
Logically inner self joining follows table on
a.followee_id = b.follower_id
AND b.followee_id = a.follower_id
AND a.created_at < b.created_at;
Since in this dataset, created_at has the same value for all records.
The condition a.created_at < b.created_at does not hold true even once.
hence, No output is displayed.

QUERY

```
SELECT
    a.followee_id AS user_id
FROM
    follows a
        INNER JOIN
    follows b ON a.followee_id = b.follower_id
        AND b.followee_id = a.follower_id
        AND a.created_at < b.created_at;
```

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Subjective Questions

*** Please Refer to SQL File for ALL the Related Queries used to answer these 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 -

APPROACH -

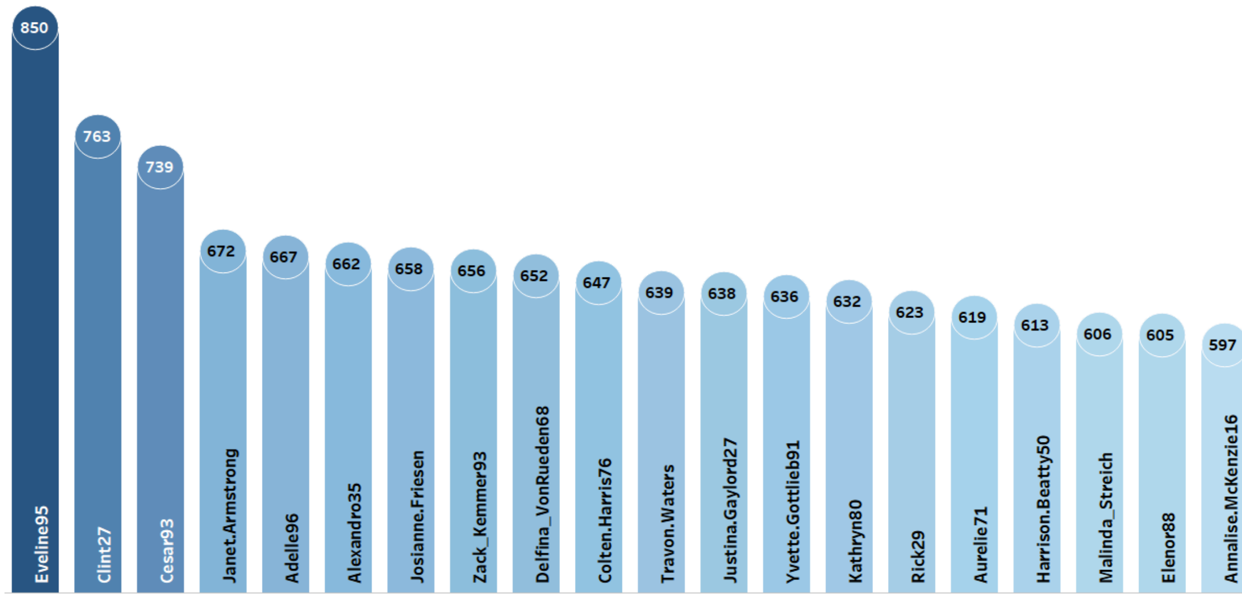
Using user_summary view and considering (Total_Likes, Total_Comments, Total_Tags, Posts_Liked, Comments_Made, Total_followers, Total_followings) these as parameters for engagement and activity levels to find most loyal & valuable users. Below attached list shows the most loyal & valuable users.

QUERY

```
SELECT  *,
    (Total_Likes + Total_Comments + Total_Tags + Posts_Liked + Comments_Made + Total_followers + Total_followings) AS Total_Engagement
FROM user_summary
WHERE
    Total_Photos > (SELECT  AVG(Total_Photos)
        FROM user_summary)
        AND Total_Likes > (SELECT  AVG(Total_Likes)
        FROM user_summary)
        AND Total_Comments > (SELECT  AVG(Total_Comments)
        FROM user_summary)
ORDER BY Total_Engagement DESC
LIMIT 20;
```

OUTPUT

UserID	Username	Total_Photos	Total_Likes	Total_Comments	Total_Tags	Posts_Liked	Comments_Made	Total_followers	Total_followings	Total_Engagement
23	Eveline95	12	420	329	24	0	0	77	0	850
88	Clint27	11	361	299	26	0	0	77	0	763
59	Cesar93	10	338	308	16	0	0	77	0	739
43	Janet.Armstrong	5	180	154	5	86	72	76	99	672
65	Adelle96	5	179	142	15	96	60	76	99	667
13	Alexandro35	5	181	148	7	93	58	76	99	662
26	Josianne.Friesen	5	168	141	11	94	69	76	99	658
52	Zack_Kemmer93	5	182	151	7	85	56	76	99	656
86	Delfina_VonRueden68	9	285	273	17	0	0	77	0	652
78	Colten.Harris76	5	177	143	9	83	60	76	99	647
6	Travon.Waters	5	173	139	8	82	62	76	99	639



RECOMMENDATIONS

They can be rewarded in following ways-

- **Exclusive Features:** Provide early access to new features, such as filters or analytics tools.
- **Badges:** Assign loyalty badges to their profiles, enhancing social status.
- **Discounts or Partnerships:** Offer discounts on sponsored products or partner with them for exclusive deals.
- **Recognition:** Highlight top users in community announcements or create "Top User" leaderboards.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans -

APPROACH -

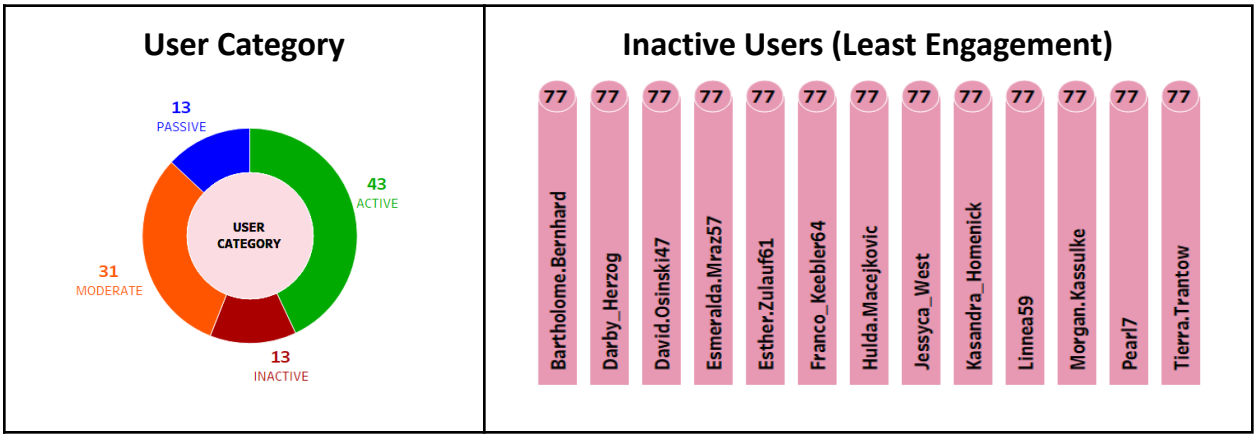
Inactive users can be identified as those with minimal engagement (Total_Likes, Total_Comments, Total_Tags, Posts_Liked, Comments_Made, Total_followers, Total_followings) . Users with least total_engagement are considered inactive.
Using user_summary view to find the list of such users.

QUERY

```
SELECT
    *,
    (Total_Likes + Total_Comments + Total_Tags + Posts_Liked + Comments_Made + Total_followers + Total_followings) AS Total_Engagement
FROM
    user_summary
ORDER BY Total_Engagement
LIMIT 20;
```

OUTPUT

UserID	Username	Total_Photos	Total_Likes	Total_Comments	Total_Tags	Posts_Liked	Comments_Made	Total_followers	Total_followings	Total_Engagement
25	Tierra.Trantow	0	0	0	0	0	0	77	0	77
80	Darby_Herzog	0	0	0	0	0	0	77	0	77
34	Pearl7	0	0	0	0	0	0	77	0	77
74	Hulda.Macejkovic	0	0	0	0	0	0	77	0	77
7	Kasandra_Homenick	0	0	0	0	0	0	77	0	77
68	Franco_Keebler64	0	0	0	0	0	0	77	0	77
45	David.Osinski47	0	0	0	0	0	0	77	0	77
49	Morgan.Kassulke	0	0	0	0	0	0	77	0	77
53	Linnea59	0	0	0	0	0	0	77	0	77
90	Esmeralda.Mraz57	0	0	0	0	0	0	77	0	77
89	Jessyca_West	0	0	0	0	0	0	77	0	77
83	Bartholome.Bernhard	0	0	0	0	0	0	77	0	77
81	Esther.Zulauf61	0	0	0	0	0	0	77	0	77



RECOMMENDATIONS

The following strategies could be suggested for inactive users

- Send notifications about friends' recent activities or anniversaries of their account creation.
- Introduce challenges or streak rewards to encourage regular activity.
- Offer discounts, free features, or additional storage for becoming active again.
- Engage users through surveys to understand why they became inactive and address their concerns.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans -

<u>APPROACH -</u>

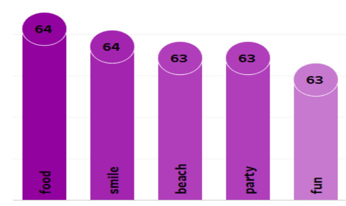
Using an approach similar to objective-12, and additionally using comments table as well. We have identified hashtags with highest engagement rates on the basis of average engagement(total_likes+Total_Comments). Attaching the list of top hashtags.

<u>QUERY</u>

```
WITH tags_engagement AS (SELECT
    a.tag_id,
    SUM(Total_Likes) AS Total_Likes,
    SUM(Total_Comments) AS Total_Comments,
    ROUND(AVG(Total_Likes), 2) AS Avg_Likes,
    ROUND(AVG(Total_Comments), 2) AS Avg_Comments,
    COUNT(a.photo_id) AS Times_Tag_Used
FROM
    photo_tags AS a
        INNER JOIN
    (SELECT   photo_id, COUNT(user_id) AS Total_Likes
    FROM likes
    GROUP BY photo_id) b ON a.photo_id = b.photo_id

    INNER JOIN
    (SELECT   photo_id, COUNT(user_id) AS Total_Comments
    FROM  comments
    GROUP BY photo_id) c ON a.photo_id = c.photo_id
GROUP BY tag_id
)

SELECT
    id AS tag_id, tag_name, Total_Likes,
    Total_Comments, Avg_Likes, Avg_Comments,
    (Total_Likes + Total_Comments) AS Total_Engagement,
    (Avg_Likes + Avg_Comments) AS Avg_Engagement,
    Times_Tag_Used
FROM tags a
     LEFT JOIN
    tags_engagement b ON id = tag_id
WHERE (Total_Likes + Total_Comments) > (SELECT  AVG(Total_Likes + Total_Comments)
        FROM tags_engagement)
ORDER BY Avg_Engagement DESC;
```

OUTPUT

tag_id	tag_name	Total_Likes	Total_Comments	Avg_Likes	Avg_Comments	Total_Engagement	Avg_Engagement	Times_Tag_Used
5	food	812	727	33.83	30.29	1539	64.12	24
21	smile	2033	1725	34.46	29.24	3758	63.70	59
20	beach	1448	1216	34.48	28.95	2664	63.43	42
17	party	1323	1151	33.92	29.51	2474	63.43	39
13	fun	1301	1089	34.24	28.66	2390	62.90	38



RECOMMENDATIONS

Prioritize topics : Creating more posts around popular hashtags or high-engagement topics can help attract more followers and keep existing ones engaged.

Targeted Ad-Campaign : Audience is more likely to connect with ads that are related to interesting and well-liked subjects, which will boost ad performance and conversion rates.

Tailored Recommendations: Use insights from these hashtags to tailor content recommendations and suggestions for users, enhancing their overall experience and increasing engagement.

Trending Content Series: Develop a content series based on these popular hashtags. Example, a weekly #sunset or #party post could keep followers engaged and looking forward to new content.

Influencer Collaborations: Partner with influencers who frequently use these hashtags and have high engagement rates. For example, a collaboration with a food influencer under #delicious or #foodie can drive targeted traffic and engagement.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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 -

APPROACH -

While the provided schema lacks demographic data, trends can still be analyzed based on posting times. Engagement trends across Hour and Day can be calculated by grouping likes & comments by hour(created_at) & dayname(created_at)

QUERY

```
with cte as (select * from likes
union all
select user_id, photo_id, created_at from comments)

select dayname(created_at) as Day, count(*) as total_activity from cte
group by dayname(created_at)
order by Day desc;

select hour(created_at) HH, count(*) as total_activity from cte
group by hour(created_at)
Order by HH Desc;
```

OUTPUT

Best Days to post

Day	total_activity
Wednesday	16270

Best Time to Post

HH	total_activity
23	16270

INSIGHTS & RECOMMENDATIONS-

- **Optimal Posting Day:** Wednesday is the most suitable day for posting.
- **Best Time to Post:** Around **11 PM** shows the highest user activity.
- **Maximum Engagement Window:** Posts made during this period have a **higher chance of receiving more reach and reactions**.
- **User Activity Insight:** Data indicates that most users interact with content during this time frame.
- **Recommendation:** Schedule important posts, promotions, or announcements for **Wednesdays at 11 PM** to maximize visibility and engagement.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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 -

APPROACH

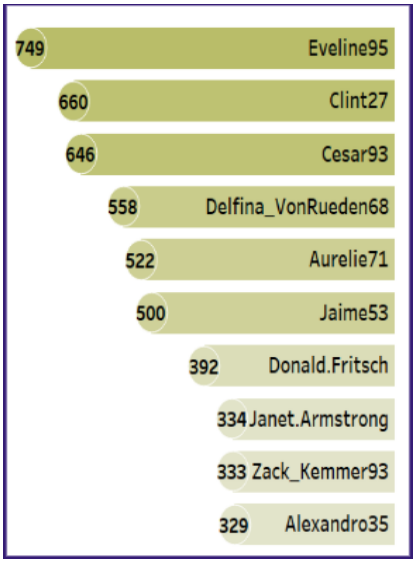
Using user_summary view to find Total_Engagement(**total_Likes+Total_Comments**) and Total_Followers. On the basis of these Parameters, the top users can be considered ideal **candidates for influencer marketing campaigns.**

QUERY

```
SELECT
    UserID,
    Username,
    (Total_Likes + Total_Comments) AS Total_Engagement,
    Total_Followers
FROM
    user_summary
ORDER BY Total_Engagement DESC , Total_Followers DESC
LIMIT 10;
```

OUTPUT

UserID	Username	Total_Engagement	Total_Followers
23	Eveline95	749	77
88	Clint27	660	77
59	Cesar93	646	77
86	Delfina_VonRueden68	558	77
58	Aurelie71	522	77
29	Jaime53	500	77
77	Donald.Fritsch	392	77
43	Janet.Armstrong	334	76
52	Zack_Kemmer93	333	76
13	Alexandro35	329	76



RECOMMENDATIONS

- **Collaboration Strategy:** Approach them with clear objectives for the campaign. Offer opportunities like product collaborations, sponsored content, or exclusive access to events or features. Give them creative freedom within content guidelines to maintain authenticity.
- **Build Relationship with Influencer :** Build long-term relationships with influencers rather than one-off campaigns for sustained engagement. Consider offering affiliate programs or ambassador roles.
- **Compensation & Incentives:** Provide fair monetary compensation or product gifting.
- **Tracking & Amplification:** Set clear KPIs(Key Performance Indicator) to measure success. Amplify influencer content across your brand's channels for maximum reach.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans -

APPROACH -

Dividing Users in 4 different categories on the basis of behavior and engagement.

I) Active users - Users with Total Photos, Likes & Comments > Avg of Photos, Likes & Comments per user.

II) Moderate Users - Users with Total Photos, Likes & Comments <= Avg of Photos, Likes & Comments per user.

III) Passive Users - Users with Total Photos = 0, but Posts_Liked > 0 or Comments_Made > 0.

IV) Inactive Users - Users with Total Photos = 0, Posts_Liked = 0 and Comments_Made = 0.

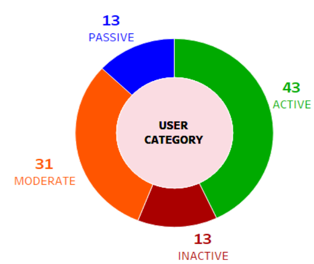
QUERY

```
with user_segment as  (select * , (Total_Likes + Total_Comments + Posts_Liked + Comments_Made + Total_Followers + Total_Followings)
 as Total_Engagement,
case
when Total_Photos = 0 and Posts_Liked = 0 and Comments_Made= 0 then "4"
when Total_Photos = 0 then "3"
when Total_Photos <= (SELECT AVG(Total_Photos) from user_summary)
and Total_Likes <= (SELECT AVG(Total_Likes) from user_summary)
AND Total_Comments <= (SELECT AVG(Total_Comments) from user_summary) then "2" else "1"
end as Segment from user_summary)

select Total_Photos, Total_Likes, Total_Comments, Posts_Liked, Comments_Made, Total_Followers, Total_Followings, Total_Engagement,
case when Segment = 1 then "Active User"
when Segment = 2 then "Moderate User"
when Segment = 3 then "Passive User"
when Segment = 4 then "Inactive User" end as "User_Category"
from user_segment
order by segment, Total_Engagement desc;

select count(*) User_Count,
case when Segment = 1 then "Active User"
when Segment = 2 then "Moderate User"
when Segment = 3 then "Passive User"
when Segment = 4 then "Inactive User" end as "User_Category"
from user_segment
group by segment
order by User_Count desc;
```

OUTPUT

Total_Photos	Total_Likes	Total_Comments	Posts_Liked	Comments_Made	Total_Followers	Total_Followings	Total_Engagement	User_Category
12	420	329	0	0	77	0	826	Active User
11	361	299	0	0	77	0	737	Active User
10	338	308	0	0	77	0	723	Active User
5	180	154	86	72	76	99	667	Active User
5	181	148	93	58	76	99	655	Active User
5	179	142	96	60	76	99	652	Active User
5	182	151	85	56	76	99	649	Active User
5	168	141	94	69	76	99	647	Active User
5	177	143	83	60	76	99	638	Active User
9	285	273	0	0	77	0	635	Active User
5	173	139	82	62	76	99	631	Active User
5	175	141	77	61	76	99	629	Active User
5	157	148	85	64	76	99	629	Active User

User_Count	User_Category
43	Active User
31	Moderate User
13	Passive User
13	Inactive User



INSIGHTS

- **Active Users (43)** make up the largest portion of the user base, contributing significantly to engagement.
- **Moderate Users (31)** show some interaction but have room for increased engagement.
- **Passive and Inactive Users (13 each)** represent a challenge, as they are less involved in interactions.
- High engagement is seen among users with **more photos and likes**, indicating visual content is a key driver.

RECOMMENDATIONS

- **Retarget Passive & Inactive Users** through personalized content, notifications, and engagement campaigns.
- **Encourage Moderate Users** with exclusive content or rewards to increase their interaction levels.
- **Leverage Active Users** by promoting user-generated content and incentivizing them to share more.
- **Optimize Posting Strategies** by analyzing engagement trends to post at peak activity times.
- **Introduce Community Challenges** to boost interactions across all user categories.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

7. **If data on ad campaigns (impressions, clicks, conversions) is available, how would you measure their effectiveness and optimize future campaigns?**
Ans - Analyzing a variety of indicators that reveal how well an ad campaign is reaching and interacting with its target audience is crucial for determining its efficacy. In addition to assessing the campaign's effectiveness, the goal is to use the data to improve future campaigns for better results.

<u>Measuring Campaign Effectiveness:</u>

1.Impressions: Good visibility is indicated by high impressions. However, effectiveness is not fully revealed by impressions alone. They must be examined alongside other metrics, such as conversions and clicks.

2.Clicks: A greater click-through rate indicates that the targeting, creativity, and ad content are appealing and relevant to the target group of people. A low rate of clicks could mean that the wrong people are seeing the advertisement or that it is not engaging.

3.Conversion Rate: A high conversion rate shows that people are motivated to take further action in addition to being interested enough to click on the advertisement. A low conversion rate can indicate problems with the offer, the landing page, or the user experience in general.

4.Engagement Metrics: Even if it doesn't result in conversions right away, high engagement shows that the audience finds the ad content compelling, which can raise brand awareness and loyalty.

<u>Optimizing Future Campaigns:</u>

1.Ad Testing: Test multiple versions of your advertisement to determine which works best (e.g., headlines, pictures, calls to action). Make future ads better by using the results. Put more effort into the ad variant that has a greater click-through rate or conversion rate.

2. Ad optimization : Regularly refresh ad creatives (images, videos, copy) to prevent ad fatigue and maintain user interest. Use insights from top-performing creatives to inform design and messaging for future campaigns.

3.Refining Targeting: Examine which audience segments (age, gender, geography, and interests) are reacting most favorably to your advertisements. Targeting should be modified to concentrate on high-performing segments.

4.Utilizing Retargeting: Show advertisements to people who have already interacted with your brand but haven't converted by using retargeting. Retarget those people with tailored advertisements to convince them to finish the conversion if data indicates a significant decrease following the initial engagement.

*We can make data-driven decisions to optimize future ads and guarantee more engagement, conversions, and overall return on investment by methodically evaluating the success of our ad campaigns using these metrics and tactics.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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 -

APPROACH

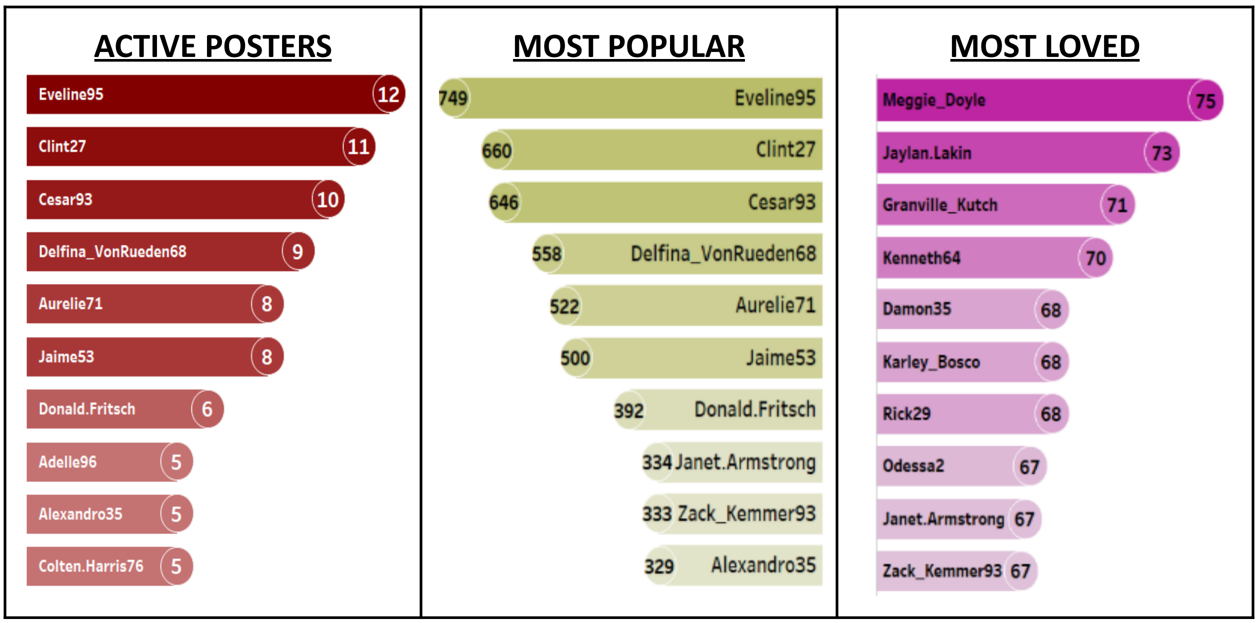
The key parameter to identify potential brand ambassadors or advocates who could help promote Instagram's initiatives or events can be -

Actively Posting - User with the most number of posts.

Most Popular - Users with engagement score(Likes+Comments+Followers)

Most Loved - Users with highest engagement rate. (Average of Likes & comments per post)

VISUALS *(Reference Obj Ques - 4, 6 & 9)*



INSIGHTS

- Eveline95,Clint27,Cesar93,Delfina_VonRueden68,Aurelie71,Jaime53,Donald.Fritsch are top 7 actively posting & Most Popular Users.
- Meggie_Doyle,Jaylan.Lakin,Granville_Kutch,Kenneth64,Damon35,Karley_Bosco,Rick 29 are the most loved Users.
- All these users are the potential community influencers & brand ambassadors for Meta on the basis of Activity and Popularity.

RECOMMENDATIONS

- Meta can approach them for Campaigns, Webinars and collaborations.
- In return users can be rewarded with cash prizes, gift cards and Merchandise - Recognising them as the most loyal and valuable Users.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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

Ans - If the objective and subjective questions weren't given, I would have approached this problem as per the following outline-

1. Data Health Check and Cleaning

- Check nulls and duplicates
- Remove duplicates
- Remove or replace nulls based on the count of null values

2. Data Exploration

- Analyse the attributes given in each table
- Identify how the tables (like photos and likes, etc.) are related to each other
- Explore and analyse engagement patterns & user activity trends (likes, comments, follows) using various SQL queries

3. Define Useful Metrics

- Engagement rates, posting trends, and follower counts.

4. Generate Insights

- Identify the top users using rank functions
- Identify active users, trends, and engagement drivers for rewarding them

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

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 -

We can use UPDATE command of MySQL on the **"User_Interactions" table, with the SET clause** on the"Engagement_Type" column to change all instances of "Like" to "Heart" to align with Instagram's terminology.

<u>Reference Query</u>

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

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx