Drive Link for the project-
https://drive.google.com/drive/folders/1u--jSDPNbTsI9YWRUGsomWZpzcYkIv6d

Project by –

Mohd Faiz

faiz.mohd340@gmail.com

Instagram User Analytics

Project Description

This project will go through some of the essential questions the marketing team has when launching new campaigns and assisting investors in making decisions about future investments by letting them know whether Instagram is still working well or is becoming obsolete, much like Facebook.

Project Approach

In order to answer the questions given by management, SQL was used. Using SQL, we created the database using the raw data provided to us for this project. Once, the database was set up, we performed various operations like selecting, sorting, joining tables, etc. to get the insights we needed.

Tech Stack Used

MySQL Workbench v8.0.32 was used to query the database.

Project Insight

Marketing

The marketing team requests information so they may launch some campaigns, and those insights are given.

1. **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.

SQL Query:

```
-- 5 oldest users of Instagram from the database provided

SELECT
        *
FROM
        ig_clone.users
ORDER BY
        created_at
LIMIT 5
```

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

Result-

Darby_Herzog (80), Emilio_bernier52 (67), Elenor88 (63), Nicole71 (95), Jordyn.Jacobson2 (38), are the 5 oldest users of Instagram.

2. **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.

SQL Query-

```
-- Users who have never posted a single photo on Instagram

SELECT
        u_table.id, username
FROM
        ig_clone.users AS u_table
LEFT JOIN
        ig_clone.photos AS p_table
        ON u_table.id = p_table.user_id
WHERE
        p_table.user_id IS null
```

5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

Result-

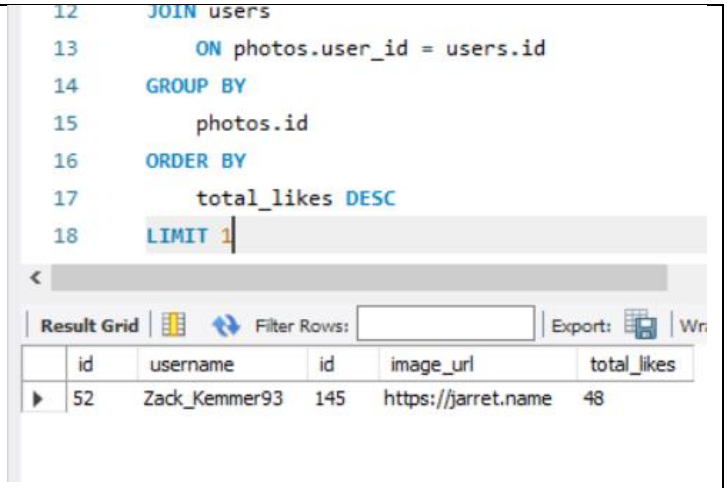
The ids and usernames are provided in the table of the users who have never posted a single photo so that we can send reminder mails to post their 1st photo.

3. **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest and will be declared winner.

SQL Query-

```
-- User who is having the most likes on a
single photo

SELECT
        users.id,
  username,
  photos.id,
  photos.image_url,
  COUNT(*) AS total_likes
FROM
        photos
JOIN likes
  ON likes.photo_id = photos.id
JOIN users
  ON photos.user_id = users.id
GROUP BY
        photos.id
ORDER BY
        total_likes DESC
LIMIT 1
```



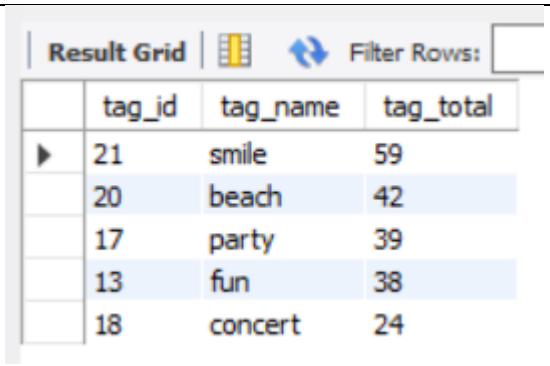
Result-

With 48 likes, user Zack Kemmer93 (52) has received the most likes for his photo with the id=145.

4. **Hashtag Researching:** A partner brand wants to know which hashtags to use in the post to reach the most people on the platform

SQL Query-

```
-- TOP 5 most commonly used hashtags on Instagram
SELECT
   tag_id,
   tag_name,
   COUNT(*) AS tag_total
FROM
          photo_tags
JOIN tags
          ON tags.id = photo_tags.tag_id
GROUP BY
          photo_tags.tag_id
ORDER BY
          tag_total DESC
LIMIT 5
```



Result Grid | Filter Rows:

tag_id	tag_name	tag_total
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

Result- The top 5 commonly used hashtags on Instagram are – **smile, beach, party, fun & concert**.

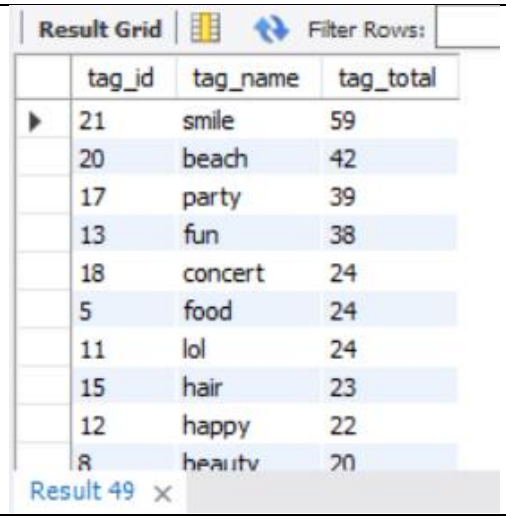
Note- However, when we don't put any limit in the query then we see that "food, lol" is also used the same number of times as "concert", i.e., 24 times.

Hence, we can say that the most commonly used hashtags on Instagram used are - **smile, beach, party, fun, concert, food & lol.**

SQL Query without limit-

```
-- TOP 5 most commonly used hashtags on Instagram

SELECT
   tag_id,
   tag_name,
   COUNT(*) AS tag_total
FROM
          photo_tags
JOIN tags
          ON tags.id = photo_tags.tag_id
GROUP BY
          photo_tags.tag_id
ORDER BY
          tag_total DESC
```



Result Grid | Filter Rows:

tag_id	tag_name	tag_total
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24
5	food	24
11	lol	24
15	hair	23
12	happy	22
8	beauty	20

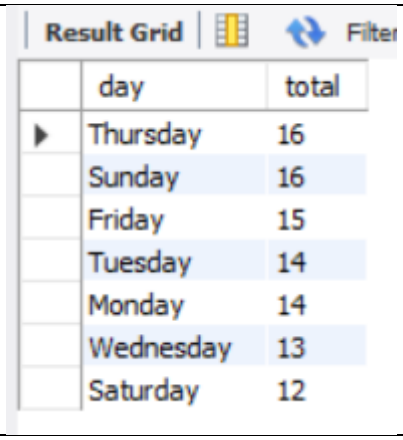
Result 49

5. **Launch AD Campaign:** The team wants to know which day would be the best day to launch Ads.

SQL Query-

```
-- What day of the week most people register on

SELECT
   DAYNAME(created_at) AS day,
   COUNT(*) AS total
FROM
        users
GROUP BY
        day
ORDER BY
        total DESC
```



day	total
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Result- The two days when people are most likely to register are **Thursday** and **Sunday**. The start of an AD campaign can be planned for one of two days to have the best chance of success.

Investor Metrics

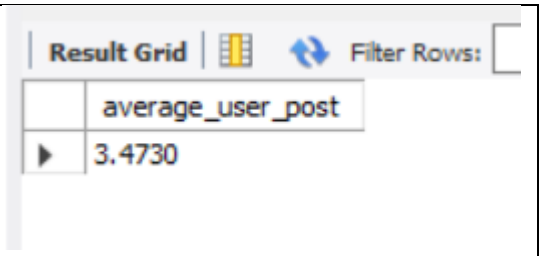
Investors want to know if Instagram is doing well and isn't going out of business like Facebook. They want to evaluate the app on the following criteria, and answers to their questions are provided.

1. **User Engagement:** Are user still active and post on Instagram or they are making fewer posts.

SQL Query-

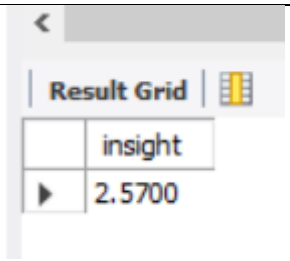
```
-- Average user posts on Instagram

SELECT
        AVG(posts_per_user.total_posts_per_user) AS
average_user_post
FROM
        (SELECT
                username,
                COUNT(photos.id) AS total_posts_per_user
        FROM
                users
        JOIN photos
                ON users.id = photos.user_id
        GROUP BY
                username) AS posts_per_user
```




```
-- total number of photos on Instagram/total number of
users

SELECT (SELECT COUNT(*)FROM photos)/(SELECT COUNT(*)
FROM users) as insight
```



Results-

The average user posts on Instagram is 3.4730

total number of photos on Instagram/total number of users = 2.57

2. **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts.

SQL Query-

```
-- Users (bots) who have liked every single photo on the site

SELECT
        likes.user_id,
   users.username,
   COUNT(*) total_likes_per_user
FROM
        likes
JOIN users
        ON likes.user_id = users.id
GROUP BY
        likes.user_id
HAVING
        total_likes_per_user = (SELECT COUNT(*) total_photos FROM photos)
```

Result Grid | Filter Rows: | Export:

user_id	username	total_likes_per_user
5	Aniya_Hackett	257
14	Jaclyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

Results- Since it would be extremely difficult for a user to like each and every picture on Instagram, we are classifying them as bots. Running the above-mentioned query has returned a list of all such users (bots).

Project Conclusion

While analyzing the dataset provided, several meaningful insights were discovered that could not have been discovered by manually searching the dataset for insights. These insights can help investors and marketing teams make better judgments in the future, saving them a good amount of time and money.

We could also leverage the MySQL tool and got a little more experienced in using the tool and also injecting different types of queries to look for insights.