Instagram User Analytics

SQL Fundamentals

PROJECT DESCRIPTION:

In this project i.e. Instagram user analytics, I provided support in the effort to produce business insights for the marketing, product, and by giving the data to the development teams working on this project using SQL (structured query language). I am utilising a variety of SQL queries during this process to obtain the relevant data. Based on the data obtained, I managed to figure out-

- The 5 oldest users of Instagram.
- The users who have never posted a single photo on Instagram.
- The user who gets the most likes on a single photo.
- The top 5 most commonly used hashtags on the platform.
- The day of the week most users register on and to provide insights on when to schedule an ad campaign.
- How often an average user posts on Instagram.
- To identify bot accounts.

APPROACH:

I initially examined the objective to find the relevant data that the team needed and then I imported the data into SQL and performed multiple queries to analyse the data and find the team's insights in requirement for commercial advantages.

TECH-STACK USED:

I carried out the project using **MySQL Workbench 8.0 CE.**

INSIGHTS:

I identify and clarify the main objective and by using SQL, I learn more about real-time analytics and obtain insights by applying various SQL commands on this project. I analysed the data obtained and provided information such as oldest users, inactive users, most liked photos, trending hashtags, day of the week most people register, user's average post and data on bot accounts which the team may use to raise the quality of Instagram.

RESULTS:

By the completion of the project, I had improved my SQL skills, learned how to work with real-world data and experienced how to approach and resolve problems.

Based on the data provided, I was able to achieve a number of results that the Instagram product team needs in order for the company to grow, which are listed below.

A) Marketing:

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

Task: Find the 5 oldest users of the Instagram from the database provided.

Query:

```
SELECT username, created_at
AS join_date
FROM ig_clone.users
ORDER BY join_date
LIMIT 5;
```

Result:

username	join_date
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

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

Task: Find the users who have never posted a single photo on Instagram.

Query:

```
SELECT u.username
FROM ig_clone.users u
LEFT JOIN
ig_clone.photos p
ON u.id = p.user_id
WHERE p.user_id IS NULL
ORDER BY
u.username;
```

Result:

username	username
Aniya_Hackett	Julien_Schmidt
Bartholome.Bernhard	Kasandra_Homenick
Bethany20	Leslie67
Darby_Herzog	Linnea59
David.Osinski47	Maxwell.Halvorson
Duane60	Mckenna17
Esmeralda.Mraz57	Mike.Auer39
Esther.Zulauf61	Morgan.Kassulke
Franco_Keebler64	Nia_Haag
Hulda.Macejkovic	Ollie_Ledner37
Jaclyn81	Pearl7
Janelle.Nikolaus81	Rocio33
Jessyca_West	Tierra.Trantow

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 now they wish to declare the winner.
Task: Identify the winner of the contest and provide their details to the team.

Query:

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

Result:

username	total_likes
Zack_Kemmer93	48

4. Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
Task: Identify and suggest the top 5 most commonly used hashtags on the platform.

Query:

```
SELECT t.tag_name,
COUNT(p.photo_id)
AS num_tags
FROM ig_clone.photo_tags p
```

```
INNER JOIN ig_clone.tags t
ON p.tag_id = t.id
GROUP BY tag_name
ORDER BY num_tags
DESC LIMIT 5
```

Result:

tag_name	tag_count
smile	59
beach	42
party	39
fun	38
concert	24

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

Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

Query:

```
SELECT DAYNAME(created_at) AS weekday,
COUNT(username) AS users_registered
FROM ig_clone.users
GROUP BY 1
ORDER BY 2 DESC
```

Result:

weekday	users_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Thursday and Sunday have the highest number of users registration.

Therefore it is ideal to schedule the ad campaign on said days.

B) Investor Metrics:

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

Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.

Query:

```
WITH CTE AS
        ( SELECT u.id AS user_id,
        COUNT(p.id) AS photo_id
        FROM ig_clone.users u
        LEFT JOIN ig_clone.photos p
        ON u.id = p.user_id
        GROUP BY u.id)
SELECT SUM(photo_id) AS total_photos,
COUNT(user_id) AS total_users,
SUM(photo_id) / COUNT(user_id) AS Photos_per_user
FROM CTE;
```

Result:

total_photos	total_users	Photos_per_user
257	100	2.5700

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

Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

Query:

```
SELECT u.id, u.username, COUNT(*) as Total_likes
FROM ig_clone.users u
JOIN ig_clone.likes l
on u.id = l.user_id
GROUP BY u.id
HAVING Total_likes = (SELECT COUNT(*)
FROM ig_clone.photos)
ORDER BY u.username;
```

Result:

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