Project Report: Instagram Clone Database Analysis

Project Description:

This project involves examining user activity and site usage via a mock-up Instagram-like database (ig_clone). The aim was to gather informative business insight for marketing and investor metrics using SQL queries. This involved discovering active users, top content, and usage habits.

Approach:

I first created the ig_clone database and added all the required tables and data. Then, I wrote SQL queries to answer each question in the case study. I used basic SQL commands like JOIN, GROUP BY, and COUNT to find useful information for marketing and business decisions.

Used Tech Stack

MySQL Workbench 8.x – I used this software to create tables, add data, and write queries. It's easy to use and great for working with databases.

SQL Language – I used SQL because it's the best language for working with databases and getting useful information from data.

Insights

A) Marketing Analysis:

1. Identify the five oldest users on Instagram from the provided database.

```
select * from users
order by created_at asc
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

2. Identify users who have never posted a single photo on Instagram.

```
select u.username, u.id as users_id
from users u
left join photos p on u.id = p.user_id
where p.id is null
order by u.id;
```

username	users_id
Aniya_Hackett	5
Kasandra_Homenick	7
Jaclyn81	14
Rocio33	21
Maxwell.Halvorson	24
Tierra.Trantow	25
Pearl7	34
Ollie_Ledner37	36
Mckenna17	41
David.Osinski47	45
Morgan.Kassulke	49
Linnea59	53
Duane60	54
Julien_Schmidt	57
Mike.Auer39	66
Franco_Keebler64	68
Nia_Haag	71
Hulda.Macejkovic	74

3. Determine the winner of the contest and provide their details to the team.

```
select photo_id,count(*) as like_count
from likes
group by photo_id
order by like_count Desc
limit 1;
```

photo_id	like_count
145	48

4. Identify and suggest the top five most commonly used hashtags on the platform

```
select t.tag_name,count(*) as tags_count
from tags t
join photo_tags pt on t.id = pt.tag_id
group by t.tag_name
order by tags_count desc
limit 5;
```

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

5. Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

```
select dayname(created_at) as day_name, count(*) as total_users_registered
from users
group by day_name
order by total_users_registered desc;
```

day_name	total_users_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

B) Investor Metrics:

1. Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

```
select count(*) / count(distinct user_id) as avg_post_per_user
from photos;
```

avg_post_per_user
3.4730

2.Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

```
select id, username
from users
where id in
 ( select user_id from likes
 group by user_id
 having count(user_id) = ( select count(id) from photos));
```

id	username
5	Aniya_Hackett
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson
36	Ollie_Ledner37
41	Mckenna17
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
71	Nia_Haag
75	Leslie67
76	Janelle.Nikolaus81
91	Bethany20
NULL	NULL

Result

- I learned how to write SQL queries and analyze data.
- I understood how to find useful information from a database.
- I learned about how users behave on social media platforms.
- This project helped me get started in data analytics.