Instagram Data Analysis Project

Project Overview

This project focuses on exploring user behavior on Instagram through detailed analysis of user-generated data. The primary objective is to uncover patterns and trends that reveal user preferences, activity levels, and overall engagement. SQL was used extensively for data handling and querying. The insights derived from this study aim to benefit marketing strategists and business analysts by providing a deeper understanding of the platform's user base.

Methodology

1. Database Setup: Created a relational database using MySQL to efficiently store and manage the Instagram data.

2. Data Ingestion: Imported datasets into the structured database.

3. Querying & Analysis: Leveraged SQL queries to perform comprehensive analysis and extract relevant insights.

Tools & Technologies

- Database Management: MySQL

- Query Language: SQL

Key Insights & SQL Queries

A) Marketing Strategies

1. Identifying Most Loyal Users

SQL Query:

```
SELECT username, created_at FROM users ORDER BY created_at ASC LIMIT 5;
```

Output:

Username	Created At
Darby_Herzog	06-05-2020 00:14
Emilio_Bernier52	06-05-2020 13:04
Elenor88	08-05-2020 01:30
Nicole71	09-05-2020 17:30
Jordyn.Jacobson2	14-05-2020 07:56

Insight:

These users were the earliest to join and can be considered for loyalty rewards.

2. Detecting Inactive Accounts

SQL Query:

```
SELECT u.username, COUNT(p.user_id) FROM users u LEFT JOIN photos p ON u.id =
p.user_id GROUP BY u.id HAVING COUNT(p.user_id) = 0;
```

Output:

List of users (truncated for brevity)

Aniya_Hackett, Kasandra_Homenick, Jaclyn81, Rocio33, Maxwell.Halvorson, etc.

Insight:

These users have registered but never posted. Target them for re-engagement.

3. Top Performer – Contest Winner

SQL Query:

```
SELECT u.username, u.id, COUNT(l.photo_id) FROM users u LEFT JOIN photos p ON u.id =
p.user_id LEFT JOIN likes l ON p.id = l.photo_id GROUP BY u.username, l.photo_id, u.id
ORDER BY COUNT(l.photo_id) DESC LIMIT 1;
```

Output:

Username	ID	Likes
Zack_Kemmer93	52	48

Insight:

This user had the most likes on a single post.

4. Popular Hashtags

SQL Query:

```
SELECT tag_name, COUNT(t.id) FROM photos p LEFT JOIN photo_tags pt ON p.id =
pt.photo_id LEFT JOIN tags t ON pt.tag_id = t.id GROUP BY t.id ORDER BY COUNT(t.id) DESC
LIMIT 5;
```

Output:

Tag Name	Count
smile	59
beach	42
party	39
fun	34
concert	24

Insight:

Most used hashtags—important for trend-based marketing.

5. Optimal Campaign Launch Days

SQL Query:

```
WITH cte AS (

SELECT CASE

WHEN WEEKDAY(created_at) = 0 THEN 'Monday'

WHEN WEEKDAY(created_at) = 1 THEN 'Tuesday'

WHEN WEEKDAY(created_at) = 2 THEN 'Wednesday'

WHEN WEEKDAY(created_at) = 3 THEN 'Thursday'

WHEN WEEKDAY(created_at) = 4 THEN 'Friday'

WHEN WEEKDAY(created_at) = 5 THEN 'Saturday'

WHEN WEEKDAY(created_at) = 6 THEN 'Sunday'

END AS weekday FROM users)

SELECT weekday, COUNT(*) FROM cte GROUP BY weekday;
```

Output:

Weekday	Accounts Created
Monday	14
Tuesday	14
Wednesday	13
Thursday	16
Friday	15
Saturday	12
Sunday	16

Insight:

Thursday and Sunday are the best days for launching campaigns.

B) Investor-Centric Metrics

6. User Engagement Statistics

SQL Query (Post Frequency):

```
WITH cte AS (SELECT u.id, COUNT(p.id) AS post_count FROM users u LEFT JOIN photos p
ON u.id = p.user_id GROUP BY u.id) SELECT post_count, COUNT(*) FROM cte GROUP BY
post_count;
```

Sample Output:

Posts	User Count
0	26
4	13
5	14
8	2
12	1

SQL Query (Average Posts):

```
WITH cte AS (SELECT u.id, COUNT(p.id) AS post_count FROM users u LEFT JOIN photos p
ON u.id = p.user_id GROUP BY u.id) SELECT AVG(post_count) FROM cte;
```

Output:

Average Posts per User: 2.57

Insight:

Most users post fewer than 6 times; engagement is moderate.

7. Suspicious Activity – Detecting Bots

SQL Query:

```
WITH cte AS (SELECT l.user_id, COUNT(*) AS like_count FROM likes l GROUP BY l.user_id)
SELECT c.user_id, u.username, c.like_count FROM cte JOIN users u ON c.user_id = u.id
WHERE like_count = 257;
```

Output (partial):

User ID	Username	Likes
5	Aniya	257
14	Jaclyn88	257
21	Rockom22	359

Insight:

These accounts liked all photos and may be bots or automated accounts.

Conclusion

Through structured SQL queries and data exploration, valuable insights were gathered regarding Instagram user behavior. This project enhanced technical proficiency in MySQL, particularly with advanced features like Common Table Expressions (CTEs) and date functions. The findings can be instrumental for marketing teams and investors aiming to make data-informed decisions.