Analyzing Instagram User Data Using SQL

Introduction:

In this project , I assumed the role of a data analyst working with Instagram's project team. My task was to analyze user interactions and engagement on the platform , providing valuable insights for marketing strategies , product development , and investor reports. This involved writing and executing SQL queries to extract meaningful data from s simulated Instagram database.

Project objectives

1. Marketing Analysis :
    - Reward the most loyal users.
    - Engage inactive users.
    - Determine the contest winner.
    - Research the most popular hash tags.
    - Find the best day to launch ad campaigns.

2. Investor Metrics :
    - Asses user management.
    - Identify potential bots or fake accounts.

Database Schema Overview

This project utilized a database with the following tables:

- users: Stores user information.
- photos : Contains details about user photos.
- comments: tracks comments on photos.
- likes: Records user likes on photos.
- follows: Captures follower-followers relationships.
- tags: Maintains unique tags used on the platform.
- photo_tags: Connects photos with their respective tags.

Key SQL Queries and Snapshots :

1. Loyal User Reward :

```
89    #Loyal User Reward
90 •  SELECT * FROM users ORDER BY created_at ASC LIMIT 5;
```

Result Grid | Filter Rows: | Edit: | Export

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

Identified the first five users who joined the Instagram, highlighting the most loyal users based on their registration dates.

2. Inactive User Engagement :

```
#Inactive user engagement
SELECT DISTINCT
    u.id AS 'User id',
    u.username AS 'User Name',
    p.id AS 'User ID Photo'
FROM users u LEFT JOIN photos p ON u.id=p.user_id
WHERE p.id IS NULL;
```

Result Grid | Filter Rows:

User id	User Name	User ID Photo
57	Julien_Schmidt	NULL
66	Mike.Auer39	NULL
68	Franco_Keebler64	NULL
71	Nia_Haag	NULL
74	Hulda.Macejkovic	NULL
75	Leslie67	NULL
76	Janelle.Nikolaus81	NULL
80	Darby_Herzog	NULL
81	Esther.Zulauf61	NULL
83	Bartholome.Bernhard	NULL
89	Jessyca_West	NULL
90	Esmeralda.Mraz57	NULL
91	Bethany20	NULL

Identified users who have never posted a photo , providing a target list for re – engagement campaigns.

3. Contest Winner Declaration :

```sql
#Contest Winner Declaration
SELECT
    u.id AS user_id,
    u.username,
    u.created_at AS user_created_at,
    p.id AS photo_id,
    p.image_url,
    p.created_dat AS photo_created_at,
    p.id AS photo_id,
    COUNT(l.user_id) AS count_userID
FROM
    photos p
JOIN
    users u ON p.user_id = u.id
```

```sql
LEFT JOIN
    likes l ON p.id = l.photo_id
GROUP BY
    p.id, u.id, u.username, u.created_at, p.image_url, p.created_dat
ORDER BY
    count_userID DESC
LIMIT 1;

#Hashtag Research
SELECT
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

user_id	username	user_created_at	photo_id	image_url	photo_created_at	photo_id	count_userID
52	Zack_Kemmer93	2017-01-01 05:58:22	145	https://jarret.name	2024-08-19 17:48:11	145	48

Determined the contest winner by indentifying the photo with the most likes .

4. Hash Tag Research :

```sql
#Hashtag Research
SELECT
    t.id,
    t.tag_name,
    COUNT(*) AS usage_count
FROM
    photo_tags pt
JOIN
    tags t ON pt.tag_id = t.id
GROUP BY
    t.id, t.tag_name
ORDER BY
    usage_count DESC
LIMIT 5;
```

Result Grid | Filter Rows:

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

Result 4

Suggested the top five most popular hash tags to improve the brand reach on the platform .

5. Ad Campaign Launch :

```sql
#Ad Campaign Launch
SELECT DAYNAME(created_at) AS day_of_week, COUNT(*) AS registrations
FROM users
GROUP BY day_of_week
ORDER BY registrations DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

day_of_week	registrations
Thursday	16

Identified the best day to the week to launch ad campaign based on user registration trends.

Investor Metrics:

6. User Engagement for Investors :

```
144     #User Engagement
145 •   SELECT AVG(photo_count) AS avg_photos_per_user
146     FROM (
147         SELECT user_id, COUNT(id) AS photo_count
148         FROM photos
149         GROUP BY user_id
150     ) AS user_photos;
```

Result Grid | Filter Rows: | Export: | Wrap Ce

avg_photos_per_user
3.4730

Provided insights into user engagement by calculating the average number of posts per user.

7. Identifying Bots & Fake Accounts :

```
#Bots & Fake Accounts
SELECT u.id, u.username
FROM users u
WHERE u.id IN (
    SELECT l.user_id
    FROM likes l
    GROUP BY l.user_id
    HAVING COUNT(DISTINCT l.photo_id) = (SELECT COUNT(*) FROM photos)
);
```

Result Grid | Filter Rows:

id	username
5	Aniya_Hackett
14	Jaclyn81
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

Detected potential bot accounts by identifying users who liked every single photo on the platform.

CONCLUSION :

This project allowed me to apply SQL skills to analyze real world data offering insights that could directly influence business decisions . I learned how to structure queries for various use cases , how to think critically about the data-driven strategies , and how to present findings that are both actionable and relevant to different stakeholders.

TOOLS USED :

1. SQL
2. MySQL Workbench