TRAINITY PROJECT 02

Instagram User Analytics

Project Description:

As a data analyst for Instagram, my goal is to analyze user interactions and engagement with the app using SQL. I have to derive insights from the given dataset to assist various teams within the business that are working on the app. The project will use SQL and MySQL Workbench. These insights will help make informed decisions about the future direction of the Instagram app.

Approach:

Creating Database: Created and inserted the values in the database using the SQL queries provided by the dataset attached in the MySQL database using MySQL Workbench 8.0

Performing analysis and extracting insights: Used SQL to perform analysis and answer the questions asked. Insights are gathered from the tables in the database using SQL queries

Tech Stack Used:

- MySQL Workbench 8.0

Insights:

A) **Marketing Analysis:**

1. **Loyal User Reward**

   The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

   Your Task: Identify the five oldest users on Instagram from the provided database.

   Query:

   select * from ig_clone.users
   order by created_at asc
   limit 5;

```
1     #Identify the five oldest users on instagram from the provided database
2 •   select * from ig_clone.users
3     order by created_at asc
4     limit 5;
```

Output: The 5 oldest users of Instagram from the given database are:

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. **Inactive User Engagement**

   The team wants to encourage inactive users to start posting by sending them promotional emails.

   Your Task: Identify users who have never posted a single photo on Instagram.

   Query:

   SELECT u.id, u.username, p.image_url
   FROM users u
   LEFT JOIN photos p
   ON u.id = p.user_id
   WHERE p.image_url IS NULL;

```
#Identify users who have never posted a single photo on Instagram
SELECT u.id, u.username, p.image_url
FROM users u
LEFT JOIN photos p
ON u.id = p.user_id
WHERE p.image_url IS NULL
```

Output: The users who have never posted a single photo on Instagram

id	username	image_url
5	Aniya_Hackett	NULL
7	Kasandra_Homenick	NULL
14	Jadyn81	NULL
21	Rocio33	NULL
24	Maxwell.Halvorson	NULL
25	Tierra.Trantow	NULL
34	Pearl7	NULL
36	Ollie_Ledner37	NULL
41	Mckenna17	NULL
45	David.Osinski47	NULL
49	Morgan.Kassulke	NULL
53	Linnea59	NULL
54	Duane60	NULL
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
90	Esmeralda.Mraz57	NULL
91	Bethany20	NULL

3. **Contest Winner Declaration**

   The team has organized a contest where the user with the most likes on a single photo win.

   Your Task: Determine the winner of the contest and provide their details to the team.

   Query:

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

```sql
#Determine the winner of the contest and provide their details to the team.
SELECT p.user_id, count(l.photo_id) AS num_of_likes, l.photo_id, u.username
FROM users u
JOIN photos p
ON u.id = p.user_id
JOIN likes l
ON p.id = l.photo_id
GROUP BY p.user_id,l.photo_id, u.username
ORDER BY num_of_likes DESC
LIMIT 1;
```

Output: Details of the winner of the contest are:

user_id	num_of_likes	photo_id	username
52	48	145	Zack_Kemmer93

4. **Hashtag Research**

A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

Query:

SELECT t.id, t.tag_name AS hashtags, count(*) AS num_of_hashtags
FROM tags t
JOIN photo_tags p
ON t.id = p.tag_id
GROUP BY t.id,t.tag_name
ORDER BY num_of_hashtags DESC
LIMIT 5;

```sql
#Identify and suggest the top five most commonly used hashtags on the platform.
SELECT  t.id, t.tag_name AS hashtags, count(*) AS num_of_hashtags
FROM tags t
JOIN photo_tags p
ON t.id = p.tag_id
GROUP BY t.id,t.tag_name
ORDER BY num_of_hashtags DESC
LIMIT 5;
```

Output: Top 5 most commonly used hashtags on the platform are

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

5. **Ad Campaign Launch**

The team wants to know the best day of the week to launch ads.

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

Query:

SELECT dayname(created_at) AS day_of_week, count(dayname(created_at)) AS num_of_users_registered
FROM users
GROUP BY dayname(created_at)
ORDER BY count(dayname(created_at)) DESC;

```
# Determine the day of the week when most users register on Instagram
SELECT dayname(created_at) AS day_of_week, count(dayname(created_at)) AS num_of_users_registered
FROM users
GROUP BY dayname(created_at)
ORDER BY count(dayname(created_at)) DESC
```

Output: Day of the week most users register on is Thursday or Sunday. So the best day of the week to launch an ad campaign could be either Thursday or Sunday.

day_of_week	num_of_users_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

B) Investor Metrics:

1. **User Engagement**

Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

Your Task: 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.

Query:

select
(select count(*) from photos) / (select count(*) from users) as avg_photos_per_user;

```
#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
(select count(*) from photos) / (select count(*) from users) as avg_photos_per_user;
```

Output: Average number of posts per user on Instagram are:

avg_photos_per_user
2.5700

2. **Bots and Fake Accounts**

Investors want to know if the platform is crowded with fake and dummy accounts.

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

Query:

SELECT username, count(*) AS Total_likes_per_user
FROM users
JOIN likes
ON users.id = likes.user_id
GROUP BY likes.user_id
HAVING Total_likes_per_user = (SELECT count(*) FROM photos);

```
#Identify users (potential bots) who have liked every single photo on the site,
#as this is not typically possible for a normal user
SELECT username, count(*) AS Total_likes_per_user
FROM users
JOIN likes
ON users.id = likes.user_id

GROUP BY likes.user_id
HAVING Total_likes_per_user =  (SELECT count(*) FROM photos);
```

Output: These are the users (bots) who have liked every single photo on the site

username	Total_likes_per_user
Aniya_Hackett	257
Jaclyn81	257
Rocio33	257
Maxwell.Halvorson	257
Ollie_Ledner37	257
Mckenna17	257
Duane60	257
Julien_Schmidt	257
Mike.Auer39	257
Nia_Haag	257
Leslie67	257
Janelle.Nikolaus81	257
Bethany20	257

Conclusion:

- The marketing team can reward the most loyal users
- Emails can be sent to the inactive users to encourage them into posting more
- The winner of the contest for the user with the most likes on a single photo can be declared
- Promotions can be done using popular hashtags
- The ad campaign can be scheduled for the most active day according to the analysis.
- This analysis can improve user engagement which can then help the growth of the company
- Bots and fake accounts can now be removed to provide better experience

Result:

I have done data analysis using SQL queries to extract insights from the database. Through this I uncovered valuable insights about the platform. I used the basics and advanced concepts of SQL. This project provided valuable project experience with MySQL.