INSTAGRAM USER ANALYTICS

MIRRA.G

PROJECT DESCRIPTION:

Instagram User Analytics is a comprehensive data analysis project aimed at providing valuable insights into user behaviour, engagement patterns, and platform performance on Instagram. Through a series of tasks, this project explores various aspects of Instagram usage, including user demographics, activity levels, content preferences, and potential issues such as fake accounts or bot activity.

The project encompasses a range of tasks, each designed to address specific inquiries from different stakeholders, including marketing teams, investors, and partner brands. By leveraging data from the Instagram database, the project offers actionable insights and recommendations to support decision-making and strategic planning.

APPROACH:

In the Instagram User Analytics project, SQL queries are pivotal for extracting insights from the database. It begins with understanding the schema and relationships, followed by breaking down tasks and planning queries accordingly. With SQL, data extraction is executed through tailored queries, leveraging functions for analysis. Results are then interpreted to derive actionable insights, while optimization ensures query efficiency. This streamlined approach maximizes the use of SQL for generating valuable insights across all project tasks.

TECH STACK USED:

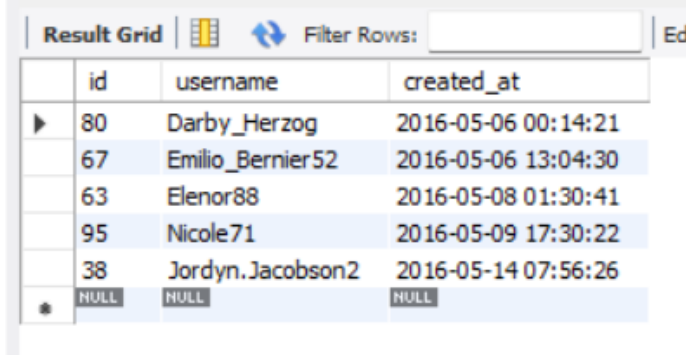MySQL Workbench 8.0 CE

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 id, username, created_at
   FROM users
   ORDER BY created_at
   LIMIT 5;
   ```

   OUTPUT:

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

   ANALYSIS:

   The output of this query provides information about the five oldest users on the platform based on their registration dates. It allows you to identify the users who joined the platform earliest. Hence, we have identified 5 oldest users to reward as the most loyal users.

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
   FROM users u
   LEFT JOIN photos p ON u.id = p.user_id
   WHERE p.id IS NULL;
   ```

OUTPUT:

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

68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

ANALYSIS:

The output of this query provides a list of users who have not posted any photos on the platform. These users may be inactive or may prefer to engage with the platform in other ways, such as liking or commenting on posts. This data can be used to encourage the inactive users to start posting by sending them promotional emails.

3. *Contest Winner Declaration:* **The team has organized a contest where the user with the most likes on a single photo wins.**
   *Your Task:* **Determine the winner of the contest and provide their details to the team.**

QUERY:

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

OUTPUT:

username	photo_id	image_url	total_likes
Zack_Kemmer93	145	https://jarret.name	48

The user with the most likes on a single photo is considered the winner of the contest. This information is crucial for determining which user's content has resonated the most with the audience and has garnered the highest level of engagement. By identifying the user with the highest number of likes on a single photo, the team can declare them as the contest winner.
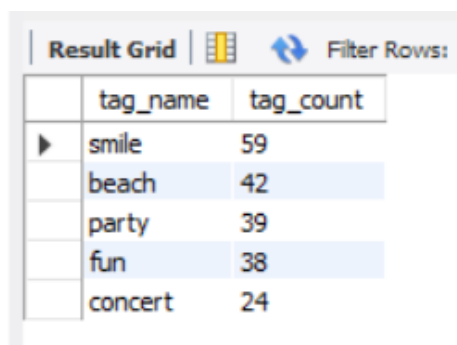
4. *Hashtag Research:* **A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.**
   Your Task: **Identify and suggest the top five most commonly used hashtags on the platform.**

QUERY:

SELECT t.tag_name, COUNT(*) AS tag_count
FROM tags t
JOIN photo_tags pt ON t.id = pt.tag_id
GROUP BY t.tag_name
ORDER BY tag_count DESC
LIMIT 5;

OUTPUT:

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

ANALYSIS:

The output reveals the top five hashtags most frequently used on Instagram, offering insights into prevalent themes and user engagement patterns. These hashtags signify popular topics, enabling marketers to align content strategies accordingly for increased visibility and audience engagement. Additionally, they provide opportunities for timely and relevant content creation, allowing brands to capitalize on trending discussions and enhance their competitive edge.
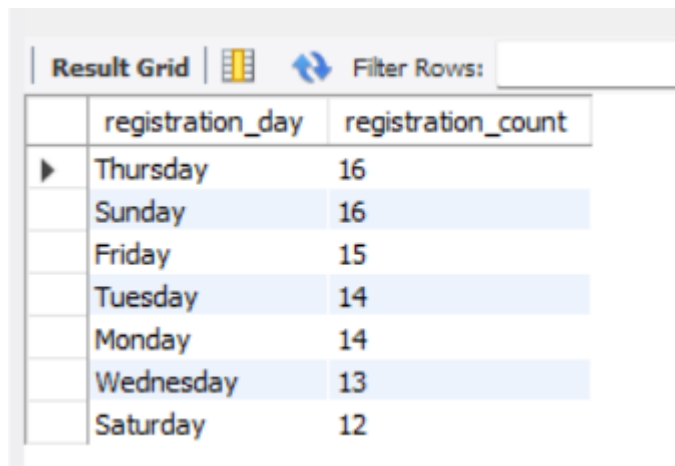
5. *Ad Campaign Launch:* **The team wants to know the best day of the week to launch ads.**
   Your Task: **Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.**

   QUERY:

   SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS registration_count
   FROM users
   GROUP BY registration_day
   ORDER BY registration_count DESC;

   OUTPUT:

registration_day	registration_count
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

   ANALYSIS:

   Based on the analysis of user registration data, it's evident that the distribution of registrations is maximum on Thursdays and Saturdays. By using this identification, the team can gain insights into user behaviour and preferences, which can inform the scheduling of ad campaigns on Instagram.

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.**
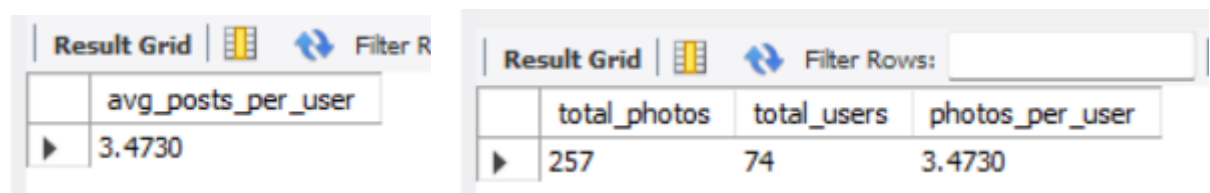
-- Calculate the average number of posts per user
SELECT COUNT(*) / COUNT(DISTINCT user_id) AS avg_posts_per_user
FROM photos;

-- Provide the total number of photos divided by the total number of users
SELECT COUNT(*) AS total_photos,
    COUNT(DISTINCT user_id) AS total_users,
    COUNT(*) / COUNT(DISTINCT user_id) AS photos_per_user
FROM photos;

OUTPUT:

avg_posts_per_user
3.4730

total_photos	total_users	photos_per_user
257	74	3.4730

ANALYSIS:

This query assesses user engagement on Instagram by calculating the average posts per user and the ratio of total photos to total users. A high average suggests active user participation, while discrepancies between averages may indicate varied user activity levels.
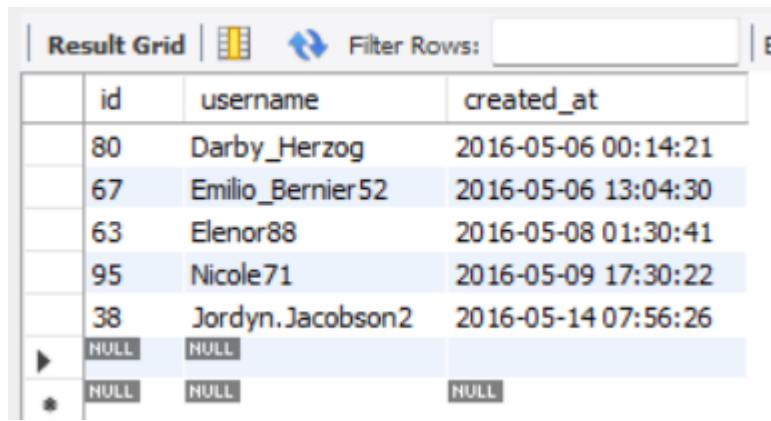
2. *Bots & Fake Accounts:* **Investors want to know if the platform is crowded with fake and dummy accounts.**
   Your Task: **Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.**

QUERY:

```
SELECT user_id
FROM users
WHERE id NOT IN (
    SELECT DISTINCT user_id
    FROM photos
    WHERE id NOT IN (
        SELECT photo_id
        FROM likes
    )
);
```

This query identifies potential bots or fake accounts on the platform by finding users who have liked every photo. The presence of such users may indicate automated or fraudulent activity, highlighting the need for further investigation and mitigation strategies to maintain platform integrity and user trust.

RESULT:

Hence, we have implemented all the tasks given as a part of the Instagram User Analytics project and provided the queries and outputs along with the analysis.