Anushka Gharage

Project 2 - Instagram User Analytics

Project Description

- The purpose of this project was to analyze user interactions and engagement with the Instagram app in order to provide valuable insights that can help the various teams within Instagram to make informed decisions about the future direction of the Instagram app.
- The primary focus was on answering specific questions related to user behaviour, contest winners, hashtag popularity, and potential bot activity.

Approach:

The approach towards the project to analyze the data and find the answers to the questions is:

1. **Creating Database:** We initiated the project by creating a clone of Instagram's database with the help of MySQL. This allowed us to have a realistic dataset to work with.
2. **SQL Queries:** We used SQL queries to extract the necessary information from the database to answer the specific questions posed.
3. **Data Interpretation:** The results obtained from SQL queries were interpreted to derive insights. For instance, we identified contest winners based on the most likes on a single photo, determined popular hashtags, and identified potential bot activity.

Tech-Stack Used:

- **MySQL Workbench:** We chose MySQL Workbench for this project due to its robust capabilities for managing and querying relational databases. It provided a user-friendly interface for SQL development and analysis.

- **MySQL Database:** We used MySQL as the relational database management system to store and manipulate the provided Instagram data.

Insights:

- **Contest Winner:** The contest winner was identified as the user with the most likes on a single photo. This insight can be valuable for recognizing and rewarding engaged users.
- **Popular Hashtags:** The top five most commonly used hashtags on the platform were identified. This information can guide content strategy and advertising efforts, as these hashtags reflect user interests.
- **Bot Detection:** Potential bot activity was detected by identifying users who have liked every single photo on the site. Further investigation may be needed to confirm their status.
- **User Registration Insights:** We determined the day of the week when most users register on Instagram. This can inform the scheduling of ad campaigns for maximum reach and engagement.
- **User Engagement:** User information such as who have never posted a single photo, who have been using the platform for the longest time, average number of posts per user, etc. These insights provide if users are still active and posting and encourage inactive users to start posting by sending promotional emails.

Result:

Through this project, we achieved the following:

- Identified contest winners based on user engagement data.
- Discovered popular hashtags to inform content and marketing strategies.

- Flagged potential bot activity for further investigation.
- Provided insights on the most common day for user registrations, facilitating better ad campaign scheduling.
- Observed daily user engagement on the Instagram app.

SQL Tasks -

A) Marketing Analysis:

1. **Loyal User Reward:** Identify the five oldest users on Instagram from the provided database.

```
3   # TASK 1 : Identify the five oldest users on Instagram from the provided database
4   SELECT username, created_at
5   FROM users
6   ORDER BY created_at ASC
7   LIMIT 5;
```

Result Grid

username	created_at
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

2. **Inactive User Engagement:** Identify users who have never posted a single photo on Instagram.

```
9   # Identify users who have never posted a single photo on Instagram
10  SELECT u.username
11  FROM users u
12  LEFT JOIN photos p ON u.id = p.user_id
13  WHERE p.id IS NULL;
```

Result Grid

username
Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

3. **Contest Winner Declaration:** Determine the winner of the contest and provide their details to the team.

```
15    # Determine the winner of the contest and provide their details to the team
16 •  SELECT u.username, p.id AS photo_id, COUNT(l.user_id) AS like_count
17    FROM users u
18    JOIN photos p ON u.id = p.user_id
19    JOIN likes l ON p.id = l.photo_id
20    GROUP BY u.username, p.id
21    ORDER BY like_count DESC
22    LIMIT 1;
```

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

username	photo_id	like_count
Zack_Kemmer93	145	48

4. **Hashtag Research:** Identify and suggest the top five most commonly used hashtags on the platform.

```
24    # Identify and suggest the top five most commonly used hashtags on the platform
25 •  SELECT tag_name, COUNT(photo_id) AS use_count
26    FROM tags
27    JOIN photo_tags ON id = tag_id
28    GROUP BY tag_name
29    ORDER BY use_count DESC
30    LIMIT 5;
```

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

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

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

Limit to 1000 rows

```
32    # Determine the day of the week when most users register on Instagram
33 •  SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS registration_count
34    FROM users
35    GROUP BY registration_day
36    ORDER BY registration_count DESC
37    LIMIT 1;
```

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

registration_day	registration_count
Thursday	16

B) Investor Metrics:

1. **User Engagement:** 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.

```
39  # Calculate the avg. no. of posts per user on instagram
40  SELECT COUNT(*) / COUNT(DISTINCT user_id) AS average_posts_per_user
41  FROM photos;
42  # Provide total no. of photos on Instagram divided by the total no. of users
43  SELECT COUNT(*) AS total_photos, (SELECT COUNT(*) FROM users) AS total_users,
44         COUNT(*) / (SELECT COUNT(*) FROM users) AS photos_per_user
45  FROM photos;
```

Result Grid

average_posts_per_user
3.4730

```
39  # Calculate the avg. no. of posts per user on instagram
40  SELECT COUNT(*) / COUNT(DISTINCT user_id) AS average_posts_per_user
41  FROM photos;
42  # Provide total no. of photos on Instagram divided by the total no. of users
43  SELECT COUNT(*) AS total_photos, (SELECT COUNT(*) FROM users) AS total_users,
44         COUNT(*) / (SELECT COUNT(*) FROM users) AS photos_per_user
45  FROM photos;
```

Result Grid

total_photos	total_users	photos_per_user
257	100	2.5700

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

```
47  # Identify users (potential bots) who have liked every single photo on the site
48  SELECT DISTINCT u.username
49  FROM users u
50  JOIN photos p ON u.id = p.user_id
51  WHERE NOT EXISTS (SELECT 1
52                    FROM photos p1
53                    LEFT JOIN likes l ON p1.id = l.photo_id AND l.user_id = u.id
54                    WHERE p1.id = p.id AND l.user_id IS NULL);
```

Result Grid

username
Andre_Purdy85
Travon.Waters
Tabitha_Schamberger11
Gus93
Justina.Gaylord27
Dereck65
Alexandro35
Billy52
Annalise.McKenzie16
Odessa2
Hailee26
Kenneth64
Josianne.Friesen
Darwin29
Dario77
Kaley9
Irwin.Larson
Yvette.Gottlieb91
Lennie_Hartmann40

username
Kelsi26
Maya.Farrell
Janet.Armstrong
Seth46
Malinda_Streich
Gerard79
Zack_Kemmer93
Ressie_Stanton46
Elenor88
Adelle96
Kathryn80
Colten.Harris76
Aracely.Johnston98
Alysa22
Rick29
Frederik_Rice
Damon35
Nicole71
Keenan.Schamberger60