Instagram User Analytics

Project Description:

This project involves analysing user interactions and engagement patterns in the Instagram app using SQL in MySQL Workbench to provide valuable insights. The insights derived from this analysis can be used by various teams within the business including marketing, product, and development, in making informed decisions to enhance user engagement and detect fake accounts.

The project is mainly divided into 2 areas:

- 1. Marketing Analysis: This includes finding loyal users, inactive users, contest winners in order to increase user engagement. Also determining popular hashtags, best day of week to launch adds.
- 2. Investor Metrics: This involves evaluating user engagement for the app by calculating average post per user. Also identifying potential bot accounts.

Approach:

We used SQL in MySQL Workbench to analyse Instagram user data and extract valuable insights for the marketing and product teams.

The provided database consists of the following tables:

1. users

```
id (INT, PRIMARY KEY, AUTO_INCREMENT)
username (VARCHAR)
created_at (TIMESTAMP)
```

2. photos

```
id (INT, PRIMARY KEY, AUTO_INCREMENT)
image_url (VARCHAR)
user_id (INT, FOREIGN KEY)
created_at (TIMESTAMP)
```

3. likes

```
user_id (INT, FOREIGN KEY).
photo_id (INT, FOREIGN KEY)
created at (TIMESTAMP)
```

4. comments

```
id (INT, PRIMARY KEY, AUTO_INCREMENT)
comment_text (VARCHAR)
user_id (INT, FOREIGN KEY)
photo_id (INT, FOREIGN KEY)
created at (TIMESTAMP)
```

5. follows

```
follower_id (INT, FOREIGN KEY)
followee_id (INT, FOREIGN KEY)
created at (TIMESTAMP)
```

6. **tags**

```
id (INT, PRIMARY KEY, AUTO_INCREMENT)
tag_name (VARCHAR, UNIQUE)
created at (TIMESTAMP)
```

7. photo_tags

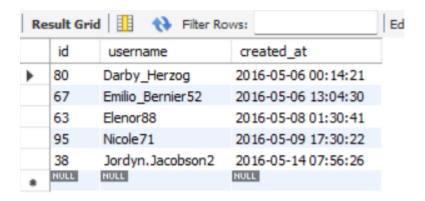
```
photo_id (INT, FOREIGN KEY) – ID of the tagged photo.
tag_id (INT, FOREIGN KEY) – ID of the tag.
```

The approach involves steps:

I. Marketing analysis

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

```
1 SELECT id, username, created_at
2 FROM users
3 ORDER BY created_at ASC
4 LIMIT 5;
5
```



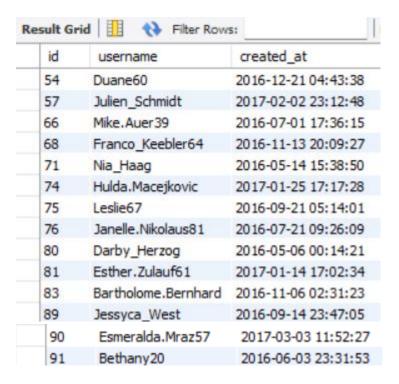
2. Inactive User Engagement: Identify users who have never posted a single photo on Instagram.

```
FROM users u

LEFT JOIN photos p ON u.id = p.user_id

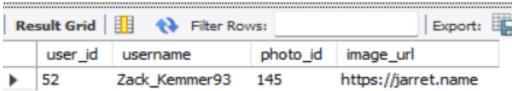
WHERE p.id IS NULL;
```

	id	username	created_at
١	5	Aniya_Hackett	2016-12-07 01:04:39
	7	Kasandra_Homenick	2016-12-12 06:50:08
	14	Jadyn81	2017-02-06 23:29:16
	21	Rocio33	2017-01-23 11:51:15
	24	Maxwell.Halvorson	2017-04-18 02:32:44
	25	Tierra.Trantow	2016-10-03 12:49:21
	34	Pearl7	2016-07-08 21:42:01
	36	Ollie_Ledner37	2016-08-04 15:42:20
	41	Mckenna 17	2016-07-17 17:25:45
	45	David.Osinski47	2017-02-05 21:23:37
	49	Morgan.Kassulke	2016-10-30 12:42:31
	53	Linnea59	2017-02-07 07:49:34



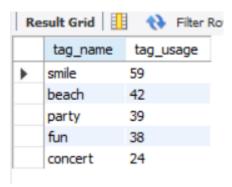
3. Contest Winner Declaration: Determine the winner of the contest and provide their details to the team.

```
11 •
       SELECT u.id AS user_id, u.username, p.id AS photo_id, p.image_url
12
       FROM photos p
       JOIN users u ON p.user_id = u.id
13
    14
          SELECT photo_id
16
          FROM likes
          GROUP BY photo id
17
          ORDER BY COUNT(*) DESC
18
          LIMIT 1
19
```



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

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



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.

```
SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS user_count
FROM users

GROUP BY registration_day

ORDER BY user_count DESC

LIMIT 1;

Result Grid Filter Rows:

registration_day user_count
```

II. Investor Metrics

Thursday

16

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.

```
34 •
      SELECT
           (COUNT(p.id) / COUNT(DISTINCT u.id)) AS avg_posts_per_user,
35
           COUNT(p.id) AS total_photos,
36
          COUNT(DISTINCT u.id) AS total users
37
       FROM users u
38
       LEFT JOIN photos p ON u.id = p.user_id;
39
40
Result Grid
                   Filter Rows:
                           total_photos
    avg_posts_per_user
                                          total_users
   2.5700
                          257
                                          100
```

- Counted the total number of photos from the **photos table**.
- Counted the total number of users from the users table.
- Computed the average posts per user using AVG() and division operations.
- 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.

```
SELECT l.user_id, u.username, COUNT(l.photo_id) AS total_likes

FROM likes l

JOIN users u ON l.user_id = u.id

GROUP BY l.user_id, u.username

HAVING total_likes = (SELECT COUNT(id) FROM photos);
```

	user_id	username	total_likes
١	5	Aniya_Hackett	257
	14	Jaclyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna 17	257
	54	Duane60	257
	57	Julien_Schmidt	257
	66	Mike. Auer 39	257
	71	Nia_Haag	257
	75	Leslie67	257
	76	Janelle.Nikolaus81	257
	91	Bethany20	257

- Identified potential bots who liked every single photo.
- Counted total likes per user from the likes table.
- Compared it with the total number of photos using HAVING COUNT(like_id) =
 (SELECT COUNT(*) FROM photos).
- Filtered users with suspiciously high engagement, suggesting automated behaviour.

Insights

- 1. **Loyal Users**: The oldest users on the platform registered in early May 2016. These users have been active for a long time, which suggests platform loyalty. Retaining such users could be crucial for engagement strategies.
- 2. **Inactive Users**: Several users have never posted a single photo, indicating a number of inactive users. Re-engagement campaigns, such as personalized emails or notifications, could encourage them to post.
- 3. **Contest Winner**: The user **Zack_Kemmer93** received the highest number of likes on a single post. Identifying such users can help in influencer collaborations and engagement-driven promotions.
- 4. **Popular Hashtags**: The most frequently used hashtags include *smile, beach, party, fun, and concert*. These insights help brands optimize their content by incorporating trending hashtags to increase reach.

- 5. **Best Day for Ad Campaigns**: *Thursday* had the highest number of user registrations. This suggests that launching ad campaigns midweek (especially on Thursdays) might attract the most engagement.
- 6. **User Engagement**: The average posts per user is **2.57**, and the total number of photos divided by users gives a similar figure. This suggests moderate user activity, indicating room for growth in user engagement.
- 7. **Potential Fake/Bot Accounts**: Several users have liked every single post, which is unusual behaviour. These accounts may be bots, requiring further investigation to maintain platform authenticity and trust.

Result

- This project provided valuable insights into user behavior, engagement patterns, and platform dynamics.
- We successfully identified the most loyal users, inactive users, and contest winners, helping in targeted marketing and engagement strategies.
- The hashtag research provided direct recommendations for brands to improve their content visibility.
- The analysis of the best registration day helps in optimizing ad campaign launches for maximum impact.
- The engagement metrics highlighted the average posting habits, useful for user retention strategies.
- Detecting potential bots ensures that the platform remains authentic and secure.
- Overall, this study delivers actionable insights that can drive platform growth, improve user interaction, and enhance marketing strategies.