Instagram User Analytics

Project Description:

As a data analyst, I collaborate with Instagram's product team to analyse interactions and engagement of the user towards the app to offer insightful analysis which can help to build better products and enhance the experience of the user. In user analysis, insights are drawn about how users interact with a digital product like software or mobile application and these insights are used by a different team to make the product better.

Its purpose is to get insights into the user’s interaction and engagement with Instagram which helps the marketing team decide when to launch a new ad campaign, helps the product team decide which features need to be built and modified and also helps the development team to improve the user interaction experience with the app.

Approach:

Firstly, download the MySQL Installer from the MySQL site and install it on the computer.

In MySQL Workbench, Create the database with the name ig_clone. In this database, create different tables such as Users, Photos, Comments, Likes, Follows, Tags, and Photo Tags. Then, add data to each of the tables using SQL Queries.

Use SQL queries to extract the information that is required from the Marketing and Investor team from that database.

Tech-Stack Used:

MySQL Workbench 8.0 – Version 8.0.36 build 3737333 CE (64 bits) is used in this project as it is free and open-source. It is an easy-to-use relational database management system.

Insights:

Summary from the extracting the information using SQL queries from the database ig_clone are:

1. Users with IDs (80, 67, 63, 95, 38) are among those who have been using Instagram for a long time.
2. Out of 100 members, 26 have not uploaded a single picture since their account was created.
3. The user who has ID-52 has received the most likes on just one photo.
4. Of the 21 hashtags, it is noted that six (smile, beach, party, fun, and concert) most often used hashtags.
5. The majority of users sign up on Thursdays and Sundays of the week.
6. The average number of posts per user is approx. 3.
7. Out of 100, 13 users were thought to be potential bots because they liked every post.

Some of the highlights for insights based on the **ig_clone** database are:

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.

```
16
17    /*  A.1 - Identify the five oldest users on Instagram from the provided database. */
18
19    SELECT * FROM users
20    ORDER BY created_at ASC
21    LIMIT 5;
22
23
```

00% 1:1

Result Grid | Filter Rows: Search | Edit: | Export/Import: | Fetch rows:

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.

```sql
/* A.2 - Identify users who have never posted a single photo on Instagram. */

SELECT users.id AS user_id, users.username, photos.id AS photo_id
FROM users
LEFT JOIN photos ON users.id = photos.user_id
WHERE photos.id IS NULL;
```

00% 1:35

Result Grid Filter Rows: Search Export:

user_id	username	photo_id
5	Aniya_Hackett	NULL
7	Kasandra_Homenick	NULL
14	Jaclyn81	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
91	Bethany20	NULL

3. **Contest Winner Declaration:** The team has organized a contest where the user with the most likes on a single photo wins.

```sql
/* A.3 - Determine the winner of the contest where the user with the most likes on a single photo wins. */

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

00% 11:43

Result Grid Filter Rows: Search Export: Fetch rows:

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

4. **Hashtag Research:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

```sql
/* A.4 - Identify and suggest the top five most commonly used hashtags on the platform. */

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

00% 26:53

Result Grid Filter Rows: Search Export: Fetch rows:

tag_id	tag_name	total_tags
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.

```sql
/* A.5 - Determine the day of the week when most users register on Instagram. */

SELECT DAYNAME(created_at) AS register_day, COUNT(*) AS register_count
FROM users
GROUP BY register_day
ORDER BY register_count DESC;
```

00% 30:68

Result Grid Filter Rows: Search Export:

register_day	register_cou...
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.

```
/* B.1 -  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 COUNT(photos.id) / COUNT(DISTINCT users.id) AS avg_posts_per_user,
       COUNT(photos.id) AS total_photos, COUNT(DISTINCT users.id) AS total_users,
       COUNT(photos.id) / COUNT(DISTINCT users.id) AS photo_user_ratio
FROM users
LEFT JOIN photos ON users.id = photos.user_id;
```

100% 1:81

Result Grid Filter Rows: Search Export:

avg_posts_per_user	total_photos	total_users	photo_user_ratio
2.5700	257	100	2.5700

2. **Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.

```
/* B.2 -  Identify users (potential bots) who have liked every single photo on the site,
as this is not typically possible for a normal user. */

SELECT users.username, likes.user_id , COUNT(likes.photo_id) AS total_liked
FROM likes
JOIN users ON users.id=likes.user_id
WHERE likes.photo_id
GROUP BY likes.user_id
HAVING total_liked = (SELECT COUNT(photos.id) FROM photos);
```

100% 38:71

Result Grid Filter Rows: Search Export:

username	user_id	total_liked
Aniya_Hackett	5	257
Jaclyn81	14	257
Rocio33	21	257
Maxwell.Halvorson	24	257
Ollie_Ledner37	36	257
Mckenna17	41	257
Duane60	54	257
Julien_Schmidt	57	257
Mike.Auer39	66	257
Nia_Haag	71	257
Leslie67	75	257
Janelle.Nikolaus81	76	257
Bethany20	91	257

Results:

While doing this project, I have learned about the fundamentals of SQL such as creating the database, creating the table, and inserting the table along with some SQL queries to extract insights from the relational database which helps to identify patterns and trends in user behaviours.

From this project, the Marketing team got to know about the user's behaviours towards the platform:

- Users who have been loyal and using this platform for a long time can receive the Loyal Users Rewards.
- They also got insights into inactive users so that they could send promotional emails to retain the users.
- The marketing team has organised a contest where the user with the most likes in a single photo will win to encourage the user to post to stay active.
- The partner brand got the details about the most popular hashtags so that they can reach most people.
- The marketing team got to know that Thursdays and Sundays are the days when they can schedule an ad campaign.

Investors got to know about users still using the platform and actively posting on Instagram. They also got to know that there are so many fake and dummy accounts.

Overall, this project helped us to derive some valuable insights to make the product better and enhance user interactions and engagement with the product.