Instagram User Analytics

trainity

PROJECT #2

Submitted by:

Anurag Changmai

Data Analytics Trainee

16/06/2023

Project Description

User analytics form a core part of every digital product, be it a mobile application, software, or website. It enables a wider understanding of user behavior based on interaction with the product, which provides valuable insights, and consequently, helps in effective decision-making.

In this project, a dataset of users on the Instagram platform is utilized. This dataset is sorted to produce specific output to obtain user interactions (posts, likes, comments, etc.) and derive meaningful insights.

These insights are then analyzed from a marketing and investment point-of-view. In doing so, parameters like user loyalty, user engagement, hashtag usage, and bot identification are found. These insights will help in user targeting, deciding what ad campaigns to run and when, and eliminating bots for a leaner, improved marketing and investment strategy.

Approach

This project starts by first obtaining a sample dataset of users and their interactions on the Instagram platform.

This data is then cleaned and stored in a database called *ig_*clone, using multiple tables such as *comments, follows, likes, photo_tags, photos, tags,* and *users*.

Once this is done, SQL queries are executed on these tables to obtain specific data as per requirements.

Tech-stack used

In the execution of this project, the following software was used:

1. Oracle MySQL Workbench 8.0 v8.0.33

   → It was used to run SQL queries for creating the database, performing operations on the tables within the database, and obtaining desired outputs.

Insights

In the following slides, insights on the following points are obtained:

From a Marketing POV:

1. Identifying the 5 oldest users on the Instagram platform based on the available dataset to reward them for their loyalty

2. Identifying users who have not posted a single photo on the Instagram platform to send promotional emails reminding them to start posting

3. Identifying the user who gets the most likes on a single photo to be declared the winner of a new contest

4. Identifying the 5 most used hashtags on the Instagram platform so as to use them for expanded reach

5. Identifying the day(s) of the week when most users register on the Instagram platform in order to schedule an ad campaign accordingly

Insights (cont.)

From an Investor's POV:

1. Identifying average posts per user on the Instagram platform to understand if users are active

2. Identifying fake accounts/bots to clean up and streamline target audience

1. Identifying the 5 oldest users

SQL QUERY

RESULTS

```
1 •    USE ig_clone;
2 •    SELECT *
3      FROM users
4      ORDER BY created_at ASC
5      LIMIT 5;
```

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

The results show that users 'Darby_Herzog', 'Emilio_Bernier52', 'Elenor88', 'Nicole71', and 'Jordyn.Jacobson2' are the oldest users in that order on the platform

2. Identifying users who have not posted a single photo

SQL QUERY

```
1 •  USE ig_clone;
2 •  SELECT *
3    FROM users
4    WHERE users.id NOT IN
5          (SELECT user_id
6          FROM photos);
```

2. Identifying users who have not posted a single photo (Cont.)

RESULTS

The results show the user IDs, usernames, and profile creation dates of **26** users who have not posted a single photo on the platform

id	username	created_at
5	Aniya_Hackett	2016-12-07 01:04:39
7	Kasandra_Homenick	2016-12-12 06:50:08
14	Jaclyn81	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	Mckenna17	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
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
68	Franco_Keebler64	2016-11-13 20:09:27
71	Nia_Haag	2016-05-14 15:38:50
74	Hulda.Macejkovic	2017-01-25 17:17:28
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
80	Darby_Herzog	2016-05-06 00:14:21
81	Esther.Zulauf61	2017-01-14 17:02:34
83	Bartholome.Bernhard	2016-11-06 02:31:23
89	Jessyca_West	2016-09-14 23:47:05
90	Esmeralda.Mraz57	2017-03-03 11:52:27
91	Bethany20	2016-06-03 23:31:53

3. Identifying the user with the most likes on a single photo

SQL QUERY

```sql
1   USE ig_clone;
2   SELECT
3       users.id AS user_id,
4       users.username,
5       users.created_at AS profile_created_on,
6       photos.id AS photo_id,
7       photos.image_url AS photo_url,
8       photos.created_at AS photo_created_on,
9       COUNT(photo_id) AS no_of_likes
10  FROM photos
11  JOIN likes ON likes.photo_id = photos.id
12  JOIN users ON users.id = photos.user_id
13  GROUP BY photos.id
14  ORDER BY no_of_likes DESC
15  LIMIT 1;
```

3. Identifying the user with the most likes on a single photo (Cont.)

RESULTS

	user_id	username	profile_created_on	photo_id	photo_url	photo_created_on	no_of_likes
▶	52	Zack_Kemmer93	2017-01-01 05:58:22	145	https://jarret.name	2023-06-13 21:21:15	48

The results show that user 'Zack_Kemmer93' with user ID 52 posted a photo with ID 145 on 13/06/2023 at 9:21 pm that has received 48 likes, the highest on the platform

4. Identifying the 5 most used hashtags

SQL QUERY

```
1 •   USE ig_clone;
2 •   SELECT photo_tags.tag_id,
3             tags.tag_name,
4             COUNT(tag_id) AS total_use
5     FROM photo_tags
6     JOIN tags
7     ON photo_tags.tag_id = tags.id
8     GROUP BY photo_tags.tag_id
9     ORDER BY total_use DESC
10    LIMIT 5;
```

RESULTS

tag_id	tag_name	total_use
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

The results show that 'smile', 'beach', 'party', 'fun', and 'concert' are the 5 most used hashtags in that order on the platform

5. Identifying the busiest days of the week in terms of user registration

SQL QUERY

RESULTS

```sql
1 •  SELECT DAYNAME(created_at) AS day_of_the_week,
2           COUNT(*) AS frequency
3      FROM users
4      GROUP BY day_of_the_week
5      ORDER BY frequency DESC;
```

day_of_the_week	frequency
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

The results show that most users register on the platform on **Thursdays** and **Sundays**

6. Identifying average posts per user

SQL QUERY

```
1 •  SELECT COUNT(DISTINCT photos.id) AS total_posts,
2            COUNT(DISTINCT users.id) AS total_users,
3            ROUND((COUNT(DISTINCT photos.id) / COUNT(DISTINCT users.id)),2) AS average_posts_per_user,
4            COUNT(DISTINCT photos.user_id) AS total_active_users,
5            ROUND((COUNT(DISTINCT photos.id) / COUNT(DISTINCT photos.user_id)),2) AS average_posts_per_active_user
6      FROM photos
7      LEFT JOIN users
8      ON users.id = photos.user_id OR users.id != photos.user_id;
```

6. Identifying average posts per user (Cont.)

RESULTS

	total_posts	total_users	average_posts_per_user	total_active_users	average_posts_per_active_user
▶	257	100	2.57	74	3.47

The results show that there are a total of **257 posts** on the platform. With **100 total users**, this gives an average of **2.57 posts per user**.

However, these 100 users also contain **inactive users** (i.e. users who have not posted a single photo) and **fake accounts/bots**.

Hence, considering the total **active users** (i.e. users who have posted at least one photo), which is **74**, the **average number of posts per user** is **3.47**.

7. Identifying fake accounts/bots

SQL QUERY

```sql
1  WITH bot_finder AS (SELECT user_id, COUNT(*) AS like_count
2                         FROM likes
3                         GROUP BY user_id
4                         HAVING like_count = 257
5                         ORDER BY like_count DESC)
6  SELECT users.username as bots
7  FROM users
8  JOIN bot_finder
9  ON users.id = bot_finder.user_id
```

7. Identifying fake accounts/bots (Cont.)

RESULTS

The results show that **13 users** have liked every photo on the platform (**257 photos**).

Considering that any normal user would not be able to do so, these **13 users** have been marked as **fake accounts/bots**.

bots
▶ Aniya_Hackett
Jaclyn81
Rocio33
Maxwell.Halvorson
Ollie_Ledner37
Mckenna17
Duane60
Julien_Schmidt
Mike.Auer39
Nia_Haag
Leslie67
Janelle.Nikolaus81
Bethany20

Result

By identifying active users and cleaning up fake accounts/bots, the management team can now focus on targeting the right users. Next, they can run campaigns on specific days coupled with the usage of specific hashtags to ensure wider reach.

Moreover, by finding and rewarding loyal users as well as users who secure a high number of likes on their posts, the campaign could be further publicized. It will also ensure continued patronage by existing users while also securing more users, who are attracted by the fact that the platform looks after its regular users. This could, in turn, lead to more posts by more users, thereby increasing the average posts-per-user values.

This project has been very helpful in understanding basic and advanced SQL concepts, as well as the workings of relational databases. Moreover, it provides a peek into how product teams work at organizations, how they handle copious amounts of data, and the valuable insights they produce.