

Instagram USER ANALYSIS

By-Yash Sangwan

Instagram
Instagram, Inc.

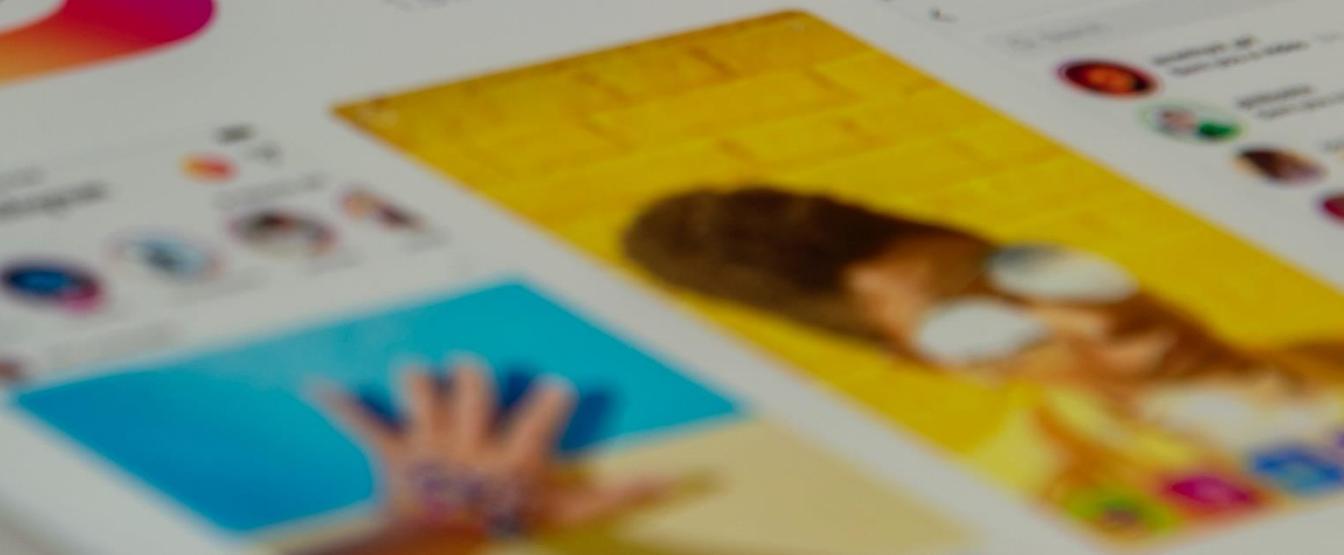
GET

In-App
Purchases

4.7 ★★★★☆
1.5M Ratings

12+

Age



Instagram

Content



Project overview

Tech-Stack

Approach

Insights

Result

With the aid of Data Analytics and SQL, we shall observe how several significant Instagram queries are resolved under this project. We'll be utilizing an Instagram database clone called ig clone. We will explore some insights using SQL that benefit the marketing team in the following ways:

- Rewarding most loyal users
- Reward inactive users to start posting
 - Declaring contest winner
 - Hashtag Researching
 - Launch AD campaign

and will help the investors in the following ways:

- Users Engagement
- Bots & fake accounts

These will help the business related team to get the maximum of the user data in the Instagram

Overview

Let's see how we handled various instances now using the **simplest queries** we could. The next section is separated into segments of examples, followed by their **MYSQL query-based solutions**. My primary method of solving any problem was to first determine which tables in the entire database would be required to find the precise solution, then, if there were multiple approaches, I used the **hit trial method** to determine which query gave the best solution, and finally, we would execute our query and determine how we could make it **more precise** so that we could give the team the precise solution.

Approach

MYSQL Workbench 8.0 CE is the programme that we have employed throughout our project to carry out all of the queries. The **benefits** of using this software include its **ease of use**, the **features** offered by the editor, and, of course, the organization of numerous databases and their tables in a structured manner, which makes it simpler to quickly lookout for a database and the tables inside it. Last but not least, its **error recognition** mechanism is also very helpful in identifying the query's problem.



Tech-Stack

Insights



Let's continue to examine the queries utilized in the various sections and scenarios of the project that was assigned to us.

PART A (Marketing)

1

- **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.
- **Our task:** Find the 5 oldest users of the Instagram from the database provided

Query 1 x

1 • Select * from users order by created_at
2 LIMIT 5;
3

Result Grid Filter Rows: Edit: Export/Import: Wrap Cell Content

| | 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 |



PART A (Marketing)

2

- **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.
- **Our Task:** Find the users who have never posted a single photo on Instagram

Query 1 x

1 • SELECT * FROM users WHERE NOT EXISTS
2 (SELECT
3 DISTINCT(user_id)
4 FROM
5 photos WHERE photos.user_id = users.id);

Result Grid Filter Rows: Edit: Export/Import: Wrap C

| | id | username | created_at |
|---|----|--------------------|---------------------|
| > | 5 | Aniya_Hackett | 2016-12-07 01:04:39 |
| | 7 | Kassandra_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 | 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 |



PART A (Marketing)

3
a)

- **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
- **Our Task:** Identify the winner of the contest and provide their details to the team

```
SQL File 2* x photo_tags tags photos
1 • SELECT
2     l.photo_id, COUNT(l.user_id) numoflikes, p.user_id
3 FROM
4     likes l,
5     photos p
6 WHERE
7     l.photo_id = p.id
8 GROUP BY l.photo_id
9 ORDER BY numoflikes DESC
10 LIMIT 1;
```

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

| photo_id | numoflikes | user_id |
|----------|------------|---------|
| 145 | 48 | 52 |



PART A (Marketing)

3

b)

- **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
- **Our Task:** Identify the winner of the contest and provide their details to the team

```
SQL File 2* x photo_tags tags photos
Limit to 1000 rows
1 • SELECT
2     l.photo_id, COUNT(l.user_id) numoflikes, p.user_id
3 FROM
4     likes l,
5     photos p
6 WHERE
7     l.photo_id = p.id
8 GROUP BY l.photo_id
9 ORDER BY numoflikes DESC
10 LIMIT 1;
```

| photo_id | numoflikes | user_id |
|----------|------------|---------|
| 145 | 48 | 52 |



PART A (Marketing)

4

- **Hashtag Researching :** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
- **Our Task :** Identify and suggest the top 5 most used hashtags on the platform

```
1 • SELECT
2     pt.tag_id,
3     COUNT(photo_id) tags_use_num,
4     t.tag_name,
5     t.created_at
6 FROM
7     photo_tags pt,
8     tags t
9 WHERE
10    pt.tag_id = t.id
11 GROUP BY pt.tag_id
12 ORDER BY tags_use_num DESC LIMIT 5;
```

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

| tag_id | tags_use_num | tag_name | created_at |
|--------|--------------|----------|---------------------|
| 21 | 59 | smile | 2023-01-27 22:43:48 |
| 20 | 42 | beach | 2023-01-27 22:43:48 |
| 17 | 39 | party | 2023-01-27 22:43:48 |
| 13 | 38 | fun | 2023-01-27 22:43:48 |
| 18 | 24 | concert | 2023-01-27 22:43:48 |



PART A (Marketing)

5
a)

- **Launch AD Campaign** : The team wants to know, which day would be the best day to launch Ads.
- **Our Task:** What day of the week do most users register on? Provide insights on when to schedule an ad campaign

SQL File 2* users

```
1 • SELECT
2     COUNT(username) number_of_registrations,
3     DAYNAME(created_at) day_of_week
4 FROM
5     users
6 GROUP BY day_of_week
7 ORDER BY number_of_registrations DESC
8 LIMIT 1;
```

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

| number_of_registrations | day_of_week |
|-------------------------|-------------|
| 16 | Thursday |



PART A (Marketing)

5
b)

- **Launch AD Campaign** : The team wants to know, which day would be the best day to launch Ads.
- **Our Task:** What day of the week do most users register on? Provide insights on when to schedule an ad campaign

```
SQL File 2" X users
Limit to 1000 rows
1 • SELECT
2     COUNT(id) AS traffic, HOUR(created_at) AS peakhr
3 FROM
4     users
5 GROUP BY peakhr
6 ORDER BY traffic DESC
7 LIMIT 1;
```

| traffic | peakhr |
|---------|--------|
| 9 | 17 |



PART B (Investors Metrics)

1

- **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts
- **Our Task :** Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

SQL File 2* photos users

1 • SELECT

2 COUNT(p.id) AS totalposts,

3 COUNT(DISTINCT (u.id)) AS totalusers,

4 FLOOR(COUNT(p.id) / COUNT(DISTINCT (p.user_id))) AS avgpost_by_a_user,

5 FLOOR(COUNT(p.id) / COUNT(DISTINCT (u.id))) AS post_per_user

6 FROM

7 users u

8 LEFT JOIN

9 photos p ON u.id = p.user_id;

10

Result Grid Filter Rows: Export: Wrap Cell Content:

| | totalposts | totalusers | avgpost_by_a_user | post_per_user |
|--|------------|------------|-------------------|---------------|
| | 257 | 100 | 3 | 2 |



PART B (Investors Metrics)

2

- **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts
- **Our Task:** Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

SQL File 2 x likes

```
1 • SELECT
2     u.* , COUNT(u.id) likes_by_a_user
3     FROM
4         users u
5             INNER JOIN
6                 likes l ON u.id = l.user_id
7             GROUP BY u.id
8             HAVING likes by a user = (SELECT COUNT(*) FROM photos);
```

Result Grid | Filter Rows: Export: Wrap Cell Content: □

| | id | username | created_at | likes_by_a_user |
|---|----|--------------------|---------------------|-----------------|
| ▶ | 5 | Aniya_Hackett | 2016-12-07 01:04:39 | 257 |
| | 14 | Jadyn81 | 2017-02-06 23:29:16 | 257 |
| | 21 | Rocio33 | 2017-01-23 11:51:15 | 257 |
| | 24 | Maxwell.Halvorson | 2017-04-18 02:32:44 | 257 |
| | 36 | Ollie_Ledner37 | 2016-08-04 15:42:20 | 257 |
| | 41 | Mckenna17 | 2016-07-17 17:25:45 | 257 |
| | 54 | Duane60 | 2016-12-21 04:43:38 | 257 |
| | 57 | Julien_Schmidt | 2017-02-02 23:12:48 | 257 |
| | 66 | Mike.Auer39 | 2016-07-01 17:36:15 | 257 |
| | 71 | Nia_Haag | 2016-05-14 15:38:50 | 257 |
| | 75 | Leslie67 | 2016-09-21 05:14:01 | 257 |
| | 76 | Janelle.Nikolaus81 | 2016-07-21 09:26:09 | 257 |
| | 91 | Bethany20 | 2016-06-03 23:31:53 | 257 |



Results

Achievements

- By doing this project I got to know about many functions in SQL.
- I learned how to refine our outputs in a better way.
- Both conceptual and practical understanding of queries.

How this projects helped me?

- An increase in confidence after using what we've learned in practical situations.
- Learned about many scenarios and instances that were examined utilizing data analytics at major digital firms like Instagram or Meta, etc.

*and this is the end of
our project on..*

Instagram user analysis

THANK YOU.