

## Project Description:

This project aims to analyse user interactions on Instagram to extract actionable insights that can drive business growth. The data analysis focuses on user engagement and behaviour, providing the product, marketing, and development teams with critical information to enhance the platform's features and user experience.

## Approach:

I began by thoroughly understanding the questions posed by the management team. Using MySQL Workbench, I wrote SQL queries to retrieve the necessary data from the Instagram database. The approach involved:

1. **Data Extraction:** Writing and executing SQL queries to extract relevant user data.
2. **Data Analysis:** Identifying patterns, such as the oldest users, users who never posted, the most popular hashtags, and the most active days on the platform.
3. **Insight Generation:** Summarizing the extracted data to answer the specific questions posed by the marketing and investor teams.

## 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.  
Your Task: Identify the five oldest users on Instagram from the provided database.
  - For this task I'll fetch data from users table in ascending order based on created\_at field then I'll limit the row count to 5 as we only required 5 oldest user information.

The screenshot shows a database management interface with a dark theme. On the left, a 'SCHEMAS' sidebar lists 'companydb' and 'ig\_clone'. The 'ig\_clone' schema is expanded, showing 'Tables', 'Views', 'Stored Procedures', 'Functions', and 'sys'. The main area displays a SQL query in 'Query 2':

```
1 USE IG_CLONE;
2
3 SELECT
4     *
5 FROM
6     USERS
7 ORDER BY USERS.CREATED_AT ASC
8 LIMIT 5;
```

Below the query editor, a 'Result Grid' shows the first five rows of the 'USERS' table. The columns are 'id', 'username', and 'created\_at'. The data is as follows:

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

2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

Your Task: Identify users who have never posted a single photo on Instagram.

- For this task I will use 'NOT IN' operator with sub query. This will help me to fetch only those users whose userid is not present in photos table.
- Alternative way would be to use left join this also fetch same data.( SELECT u.\* FROM users u LEFT JOIN photos p ON u.id = p.user\_id WHERE p.user\_id IS NULL)

SQL Query:

```

10 SELECT
11 *
12 FROM
13 USERS
14 WHERE
15 USERS.ID NOT IN (SELECT
16 PHOTOS.USER_ID
17 FROM
18 PHOTOS);

```

Result Grid:

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

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

Your Task: Determine the winner of the contest and provide their details to the team.

- I will be using multiple joins with users, photos and likes tables. To get total count on photo I will use "count" on user\_id field present in likes table. Now as we need username, I will use group by to create group for username and photo\_id. To get most liked photo I will use order by on total\_likes and limit the result to 1 with limit clause as we need only one winner.
- Alternative approach would be to use with clause to create temp table for count of likes for each photos, sort data in descending to total likes then limit the result to 1. Now join that temp table with users and photos table to get username and photo id. (WITH TOTAL\_LIKES AS (SELECT COUNT(USER\_ID) AS TOTAL\_LIKES ,PHOTO\_ID FROM LIKES GROUP BY PHOTO\_ID ORDER BY TOTAL\_LIKES DESC LIMIT 1) SELECT U.USERNAME,P.ID,L.TOTAL\_LIKES FROM USERS U,PHOTOS P, TOTAL\_LIKES L WHERE U.ID=P.USER\_ID AND P.ID=L.PHOTO\_ID;)

The screenshot shows a database management interface. On the left, a sidebar lists the database structure: 'companydb' (expanded) contains 'ig\_clone' (expanded), which includes 'Tables' (comments, follows, likes, photos, photo\_tags, tags, users), 'Views', 'Stored Procedures', and 'Functions'. The main area displays a SQL query:

```
19
20 SELECT
21     U.USERNAME,
22     P.ID AS PHOTO_ID,
23     COUNT(L.USER_ID) AS TOTAL_LIKES
24 FROM
25     USERS U
26 JOIN
27     PHOTOS P ON U.ID = P.USER_ID
28 JOIN
29     LIKES L ON P.ID = L.PHOTO_ID
30 GROUP BY
31     P.ID, U.USERNAME
32 ORDER BY
33     TOTAL_LIKES DESC
34 LIMIT 1;
```

Below the query, the 'Result Grid' shows the following data:

USERNAME	PHOTO_ID	TOTAL_LIKES
Zack_Kemmer93	145	48

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

Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

- For this task I will be using group by and count to get total count of tag\_id for each tag\_id then join photo\_tag table with tag table to get the name of the tag. Order the total count in descending order limit the result to 5 to get most 5 used hashtags.

```
36 SELECT
37     tag_id, tag_name, COUNT(tag_id) total_count
38 FROM
39     photo_tags,
40     tags
41 WHERE
42     id = tag_id
43 GROUP BY tag_id
44 ORDER BY total_count DESC
45 LIMIT 5;
```

100% 6:41

Result Grid Filter Rows: Search Export: Fet

	tag_id	tag_name	total_count
	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.

Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

- I will be using DAYNAME to extract the name of the day from created\_at field as we required the name of the day, user count to sum up the number of users who created their account on that day. Use order by to sort the count in descending order.

```
47 SELECT
48     DAYNAME(CREATED_AT) AS DAY_OF_WEEK,
49     COUNT(*) AS USER_COUNT
50 FROM
51     USERS
52 GROUP BY
53     DAY_OF_WEEK
54 ORDER BY
55     USER_COUNT DESC;
```

100% 15:55

Result Grid Filter Rows: Search Export:

	day_of_week	user_count
	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.  
Your Task: 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.
  - For first task I have divided total number of photos to total distinct users who posted on the Instagram which will give the average number of posts by user.
  - For second task I have to divide total number of post with total number of user present on Instagram irrespective of their posting status for that I used right join to include those users as well from users table who has not posted any photos.

```

57
58 SELECT
59     COUNT(P.ID) / COUNT(DISTINCT (P.USER_ID)) AS AVG_POST_PER_USER
60 FROM
61     PHOTOS P;
62

```

100% 14:65

Result Grid Filter Rows: Search Export:

AVG_POST_PER_USER
3.4730

```

62 SELECT
63     COUNT(P.ID) / COUNT(U.ID) AS TOTAL_PHOTOS_DIV_TOTAL_USERS
64 FROM
65     PHOTOS P
66     RIGHT JOIN
67     USERS U ON U.ID = P.USER_ID;

```

100% 32:67

Result Grid Filter Rows: Search Export:

TOTAL_PHOTOS_DIV_TOTAL_USERS
0.9081

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

Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

- For this task I have used group by and having clause to filter out the user who has likes all the photos present in photos table. As I am using group by clause I cannot use where condition to filter the data, so I have used having clause for filtration.

```
69 SELECT
70     USER_ID
71 FROM
72     LIKES
73 GROUP BY
74     USER_ID
75 HAVING
76     COUNT(PHOTO_ID) = (SELECT COUNT(ID) FROM PHOTOS);
77
```

100% 37:76

Result Grid Filter Rows: Search Export:

U...	
5	
14	
21	
24	
36	
41	
54	
57	
66	
71	
75	
76	
91	

## Tech-Stack Used:

- **MySQL Workbench** (version 9.0.1): Chosen for its robust database management and data visualization features, essential for executing complex SQL queries and analysing large datasets.
- **SQL**: Utilized to perform data extraction and manipulation, allowing for precise querying and efficient handling of Instagram's user data.



## Insights:

- **User Loyalty:** Identified the five oldest users on the platform, providing a basis for rewarding loyalty.
- **Engagement Patterns:** Found inactive users for targeted re-engagement campaigns and recognized the most popular hashtags to guide brand partnerships.
- **Contest Winner Declaration:** Zack\_Kemmer93 is the contest winner with 48 total like on his post.
- **Hashtag Research:** Five most common hashtags are smile, beach, party, fun and concert.
- **Optimal Ad Timing:** Determined the most active days for user registrations, helping to optimize ad campaign launches. Here Thursday and Sunday would be the best of the week for new AD launch.
- **User Engagement:** From the above output we can see that on an average user is posting 3-4 photos and around 10% of users have not posted anything yet.
- **Bots Detection:** Detected potential bots by identifying users who liked every photo, addressing concerns about fake accounts.

## Result:

The analysis provided the product team with clear insights that informed decisions on user engagement strategies, feature development, and marketing campaigns. The findings helped in identifying loyal users, optimizing ad spend, and improving user experience on Instagram, contributing to the platform's overall growth.