

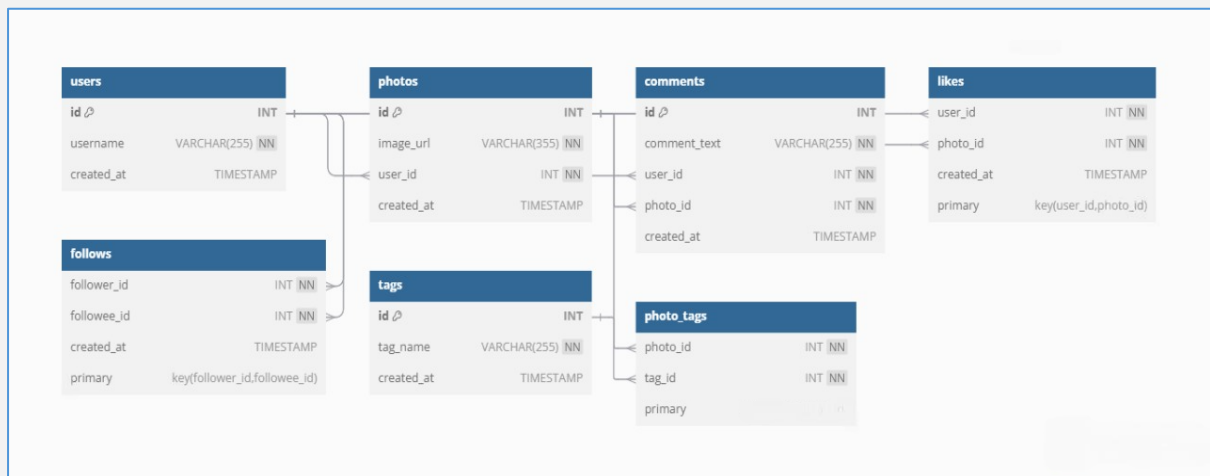
Instagram User Analytics

1. Project Description

The aim of this project is to analyse Instagram user interactions and engagement data using SQL. User analysis involves tracking how users engage with a digital product, such as software application or a mobile app. By deriving specific questions posed by the management team, the analysis provides insights into user behaviour, platform trends, and engagement patterns. The findings will support informed decision-making for marketing campaigns, product improvement and investor reports.

2. Approach

Step 1. Database Creation: The database **ig_clone** was created using the CREATE DATABASE command, followed by tables with user data using CREATE TABLE command. Then filled the tables with data using INSERT INTO table(values) command.



The database created includes the following tables:

1. Users: Contains details about Instagram users like user_id, username, created_at as timestamp indicating when the account was created.
2. Photos: Stores the Information about photos uploaded by users like user_id, image_url, user_id, created_at as timestamp indicating when photo was uploaded.
3. Comments: Stores the Information about comments made by users on photos like id , comment_text, user_id, photo_id, created_at.
4. Likes: Stores the information about the photos liked by users like user_id, photo_id, created_at.
5. Follows: This table tracks follower relationships between users who follow each other on platform like follower_id, followee_id, created_at.
6. Tags: Stores information about tags like id, tag_name, created_at.
7. Photo tags: This table tracks relationship between photos and tags and stores information like photo_id, tag_id.

Step 2. Analysis: Queries were written to find the answers to the specific tasks mentioned for marketing analysis and investor metrics. Results from the queries were analysed to derive actionable insights.

3. Tech-Stack Used

- MySQL Workbench 8.0: Used for writing and executing SQL queries due to its user-friendly interface and advanced query features.
- SQL: Used for querying the Instagram user database.
- Relational database, as it is suitable for structured data.

4. Insights

A) Marketing Analysis:

1. **Problem Statement:** To Reward the most loyal users, find the 5 users who have been using the platform for the longest time.

Query:

```
select * from users
order by created_at
limit 5;
```

Output:

	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

Insights:

- The five oldest users with id's 80, 67, 63, 95, 38 are identified based on their account creation dates. All these accounts were created in May,2016.
 - Users with the longest time on the platform can be rewarded with loyalty incentives to enhance retention.
2. **Problem Statement:** The team wants to encourage inactive users by sending them promotional emails to post their 1st photo. Identify users who have never posted a single photo on Instagram.

Query:

```
11 • select * from users
12   where id not in(select user_id from photos);
```

Output:

	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
	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

Insights:

- There are total of 100 Users and out of that 26% users have never posted a single photo on Instagram.
- A significant number of users have never posted. Targeted email campaigns could encourage participation.

3. **Problem Statement:** The team started a contest and the user who gets the most likes on a single photo will the contest. Identify the winner of the contest and provide their details to the team to declare the winner.

Query:

```

15 • select username, ph.id,ph.image_url,
16     count(*) as Total
17     From photos ph
18     Inner Join likes l
19         on ph.id=l.photo_id
20     Inner Join users u
21         on u.id= ph.user_id
22     Group by ph.id
23     Order by total desc
24     Limit 1;

```

Output:

	username	id	image_url	Total
▶	Zack_Kemmer93	145	https://jarret.name	48

Insights:

- The user named “Zack_Kemmer93” with the most likes (48 likes) on a single photo with an id 145 is identified as the contest winner.

4. **Problem Statement:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people. Suggest the top five most commonly used hashtags on the platform.

Query:

```
with top_tags as
(select tag_id from photo_tags
 group by tag_id
 order by count(tag_id) desc
 limit 5)
select t.tag_name from top_tags
join tags t on top_tags.tag_id = t.id;
```

Output:

	tag_name
▶	smile
	beach
	party
	fun
	concert

Insights:

- The top 5 most commonly used hashtags on the platform are smile, beach, party, fun, and concert.
- Popular hashtags like #smile and #beach can be leveraged in campaigns could encourage participation.

5. **Problem Statement:** The team wants to know, which day would be the best day to launch Ads. Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

Query:

```

34 • select dayname(created_at) as days,
35         count(*) as registrations
36     from users
37     group by days
38     order by registrations desc;

```

Output:

	days	registrations
▶	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

Insights:

- The best days of the week for ad campaign are determined as Thursday and Sunday.
- The majority of user registrations occur on Thursday in Weekdays and on Sunday in weekends, making them ideal for launching campaigns.

B) Investor Metrics:

1. **Problem Statement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts. Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.

Query:

```

SELECT AVG(posts_per_user) AS avg_posts_per_user,
(SELECT COUNT(*) FROM photos) /
(SELECT COUNT(distinct id) FROM users) AS total_photos_per_user
FROM (
    SELECT user_id, COUNT(*) AS posts_per_user
    FROM photos
    GROUP BY user_id
) AS user_posts;

```

Output:

	avg_posts_per_user	total_photos_per_user
▶	3.4730	2.5700

Insights:

- Average posts per user is calculated as 3.47 to measure user engagement.
- Average posts per user show a healthy engagement trend.

2. **Problem Statement:** The investors want to know if the platform is crowded with fake and dummy accounts. Provide data on users (bots) who have liked every single photo on the site.

Query:

```
select username, count(*) as Number_of_likes from users
INNER JOIN likes
on likes.user_id= users.id
Group By likes.user_id
Having Number_of_likes = (select Count(*) from photos);
```

Output:

	username	Number_of_likes
▶	Aniya_Hackett	257
	Jadyn81	257
	Rocio33	257
	Maxwell.Halvorson	257
	Ollie_Ledner37	257
	Mckenna17	257
	Duane60	257
	Julien_Schmidt	257
	Mike.Auer39	257
	Nia_Haag	257
	Leslie67	257
	Janelle.Nikolaus81	257
	Bethany20	257

Insights:

- Identified 13 users (potential bots) who have liked every photo on the platform.
- Users who exhibit bot-like behaviour should be flagged for review.

6. Results

- From above data analysis, we have gained much important information about our users and their activities on Instagram.
- We now know that the top five oldest users accounts were all created in May,2016, there are a total of 257 photos, 100 users out of which 26 users are not active.
- Also, the most liked photo id is 145 with total of 48 likes, whereas smile and beach are the most commonly used hashtags.
- The marketing team now has a clear understanding of user behaviour to plan campaigns, it would be a wise strategy to launch ads on Thursdays and Sundays of the week, while investors are informed about platform activity and potential risks (fake accounts).
- The findings have reinforced the importance of data-driven decision-making for platform growth.