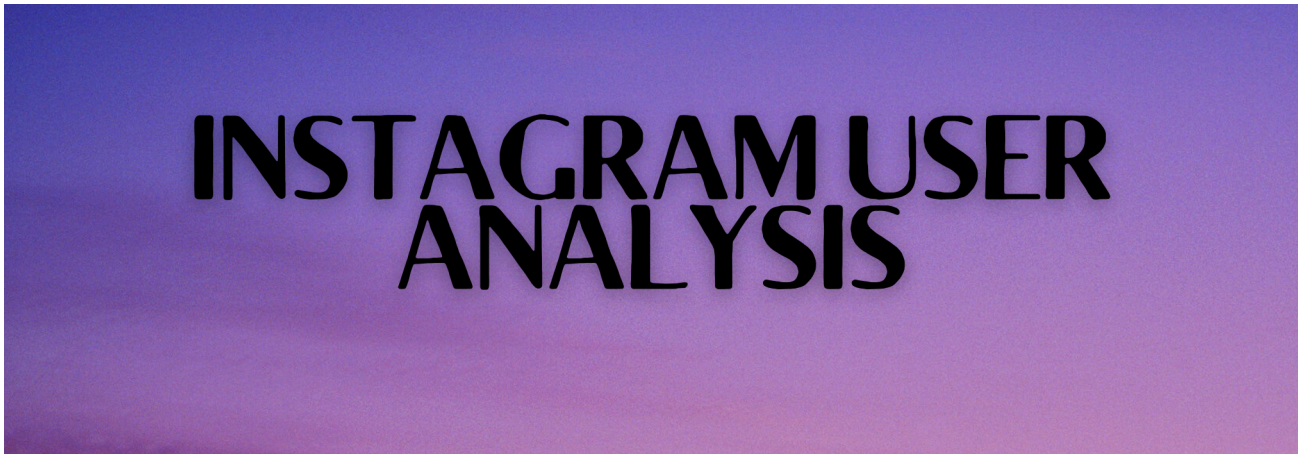


Project Description:

The Project is on Instagram User analysis, which is the process where we to track on how users engage and interact with the interface. From which business insights can be derived for marketing, product and development teams so that the business can either conduct new campaigns accordingly, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.

In order to further proceed with the product team, we are asked to provide insights asked by management team.

Below are the tasks that we were asked by the management:

Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

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
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
3. **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
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 commonly used hashtags on the platform
5. **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

B) Investor Metrics: Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

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
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).

Approach:

- **First step is to copy the dataset in MySQL by creating a new database 'ig_clone'.**
- **Check all the tables and columns within the table.**
- **For every task we are assigned, we break the task and then merge and query from the database to get the solution to the problem question.**

For example:

Take, Task 6 where we need to query how many times does an average user posts on IG.

Here first we check the table 'photos'

```
select * from photos;
```

And then find total photos posts per user .

```
select count(user_id) as avg_users
from photos
group by user_id;
```

And then finally come to finding Average of them.

```
select avg(num_users)
from
(select count(user_id) as num_users
from photos
group by user_id) as t1;
```

TASK 1: FINDING THE FIVE OLDEST USERS TO REWARD THEM

```
mysql> select * from users
    -> order by created_at
    -> 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 |
+----+------------------+---------------------+
5 rows in set (0.00 sec)
```

TASK 2: FINDING THE USERS WHO NEVER POSTED A SINGLE PHOTO TO SEND A NOTIFICATION

```
mysql> select users.id from users
    -> left join photos
    -> on photos.user_id = users.id
```

```
   -> where photos.user_id is null;
+----+
| id |
+----+
|  5 |
|  7 |
| 14 |
| 21 |
| 24 |
| 25 |
| 34 |
| 36 |
| 41 |
| 45 |
| 49 |
| 53 |
| 54 |
| 57 |
| 66 |
| 68 |
| 71 |
| 74 |
| 75 |
| 76 |
| 80 |
| 81 |
| 83 |
| 89 |
| 90 |
| 91 |
+----+
26 rows in set (0.00 sec)
```

TASK 3: IDENTIFY THE WINNER WITH MOST LIKES ON A SINGLE PHOTO

```
mysql> select likes.photo_id, photos.user_id, users.username  from likes
    -> join photos on likes.photo_id = photos.id
    -> join users on photos.user_id = users.id
    -> group by photo_id
    -> order by count(photo_id) desc
    -> limit 1;
+----------+---------+---------------+
| photo_id | user_id | username      |
+----------+---------+---------------+
|      145 |      52 | Zack_Kemmer93 |
+----------+---------+---------------+
1 row in set (0.01 sec)
```

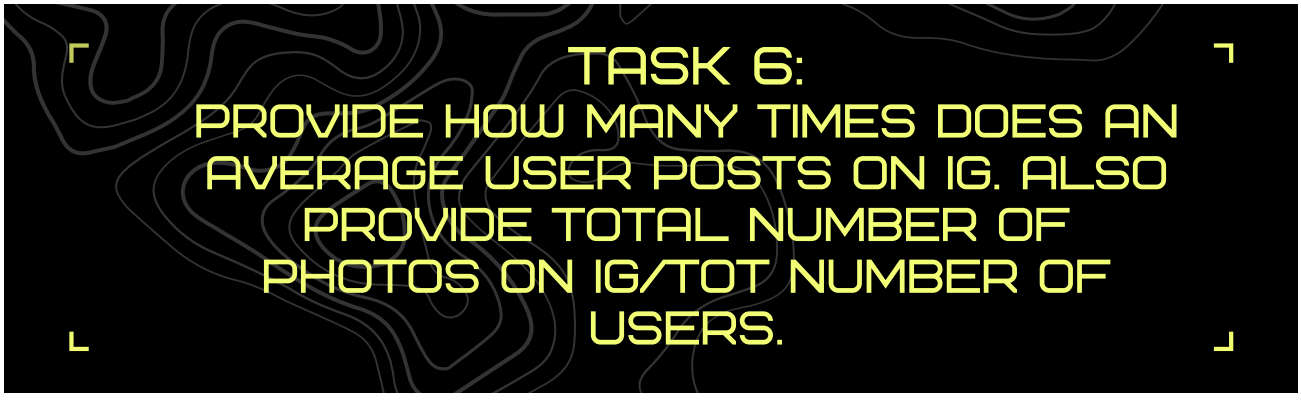
TASK 4: IDENTIFY AND SUGGEST TOP FIVE MOST USED HASTAGS ON THE PLATFORM

```
mysql> select tags.id, tags.tag_name, count(photo_tags.tag_id) as no_of_tags
    -> from tags
    -> join photo_tags
    -> on tags.id = photo_tags.tag_id
```

```
   -> group by  photo_tags.tag_id
   -> order by no_of_tags desc
   -> limit 5;
+----+----------+------------+
| id | tag_name | no_of_tags |
+----+----------+------------+
| 21 | smile    |         59 |
| 20 | beach    |         42 |
| 17 | party    |         39 |
| 13 | fun      |         38 |
| 18 | concert  |         24 |
+----+----------+------------+
5 rows in set (0.00 sec)
```


```
mysql> select weekday(created_at), count(*) from users
   -> group by weekday(created_at)
   -> order by count(*) desc;
+----------------------+----------+
| weekday(created_at) | count(*) |
+----------------------+----------+
|                    3 |       16 |
|                    6 |       16 |
|                    4 |       15 |
|                    1 |       14 |
|                    0 |       14 |
|                    2 |       13 |
```

```
|                5 |      12 |
+--------------------+----------+
7 rows in set (0.00 sec)
```




```
mysql> select avg(num_users) from
    -> (select count(user_id) as num_users from photos group by user_id) as t1;
+---------------+
| avg(num_users) |
+---------------+
|        3.4730 |
+---------------+
1 row in set (0.00 sec)

mysql> select num_photos, num_users, num_photos/num_users AS num_photos_per_user
    -> from (
    -> select
    -> (select count(*) from photos) as num_photos,
    -> (select count(*) from users) as num_users ) A;
+------------+-----------+--------------------+
| num_photos | num_users | num_photos_per_user |
+------------+-----------+--------------------+
|        257 |       100 |             2.5700 |
+------------+-----------+--------------------+
1 row in set (0.07 sec)
```

TASK 7:
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).

```
mysql> select distinct(likes.user_id), count(likes.user_id) as fake_user from likes
    -> join photos on photos.id = likes.photo_id
    -> group by likes.user_id
    -> having fake_user = (select count(*) from photos);
+---------+-----------+
| user_id | fake_user |
+---------+-----------+
|       5 |       257 |
|      14 |       257 |
|      21 |       257 |
|      24 |       257 |
|      36 |       257 |
|      41 |       257 |
|      54 |       257 |
|      57 |       257 |
|      66 |       257 |
|      71 |       257 |
|      75 |       257 |
|      76 |       257 |
|      91 |       257 |
+---------+-----------+
13 rows in set (0.02 sec)
```

Insights:

- **People mostly register their accounts on Tuesday and Friday so IG can launch an AD campaign on a day before to attract more users.**
- **0.26 % or one fourth of users have not posted a single photo**
- **Out of the top five hashtags, we can see the trend that most people are social and outgoing and like to post photos when they go outside rather than indoors.**
- **Data shows on an average a user posts 3 or 4 photos on IG.**
- **0.13 % are fake users .**