Instagram User Analytics

Project Description
This is the user analysis for the product team Instagram to provide insights on the questions asked by the management team.

Following are the questions of teams, to whom we have to provide a detailed report:

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

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

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

- **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.
  Task: Identify the winner of the contest and provide their details to the team

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

- **Launch AD Campaign**: The team wants to know, which day would be the best day to launch ADs.
  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

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

- **Bots & Fake Accounts**: The investors want to know if the platform is crowded with fake and dummy accounts
  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
Since the dataset is already provided which is clean and 3NF normalized. I explored the schema of all the provided tables in the Mysql workbench and also understood how the tables are connected with each other. so that SQL queries can be executed easily.
Some of the tasks also require some extra information hence there is the need of creating some extra columns like weekday, total post, etc.

Tech-Stack Used
The software used for the analysis is *MySQL Workbench 8.0 CE*.
Since the dataset provided is already written in MySQL, using Workbench is quite easy to execute the queries. Apart from this, Workbench is easy to learn and beginner friendly, making it my first choice for analysis.

Insights

Following are all the insights that we got from the dataset after executing different queries.

- **Rewarding Most Loyal Users:**
  Find the 5 oldest users of Instagram from the database provided

  ```
  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

- **Remind Inactive Users to Start Posting:**
  Find the users who have never posted a single photo on Instagram.

  ```
  select id, username from users
  where id not in (
          select distinct user_id from photos
  );
  ```

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

- **Declaring Contest Winner**::
  Identify the winner of the contest and provide their details to the team (who gets the most likes on a single photo)

```
select * from users
where id = (
        select user_id from photos
        where id = (
        select photo_id
        from likes
        group by photo_id
        order by count(user_id) desc
        limit 1
        )
);
```

id	username	created_at
52	Zack_Kemmer93	2017-01-01 05:58:22

- **Hashtag Researching**:
  Identify and suggest the top 5 most commonly used hashtags on the platform

```
select id, tag_name from tags
where id in (
        select tag_id from (
                select  tag_id, count(photo_id) as tag_count
                from photo_tags
                group by tag_id
                order by tag_count desc
                limit 5
    ) top_tag
);
```

id	tag_name
21	smile
20	beach
17	party
13	fun
18	concert

- **Launch AD Campaign**:
  What day of the week do most users register on? Provide insights on when to schedule an ad campaign.

```
select weekday(created_at) as day, count(id) as new_user_count
from users
group by day
order by new_user_count desc;
```

Note: 0 = Monday, 1 = Tuesday, 2 = Wednesday, 3 = Thursday, 4 = Friday, 5 = Saturday, 6 = Sunday.

day	new_user_count
3	16
6	16
4	15
1	14
0	14
2	13
5	12

- **User Engagement**:
  Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram / total number of users

```
select round(avg(photo_count)) as 'average_post'
from (
        select user_id, count(id) as photo_count
        from photos
        group by user_id
)temp;
```

average_post
3

```
select count(id) as Total_photos from photos;
select count(id) as Total_users from users;
```

```
select (select count(id) as total_photos from photos) /
(select count(id) as total_users from users) as 'Total photos / Total users';
```

Total photos / Total users
2.5700

- **Bots & Fake Accounts**:
  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).

```
select id, username as BOT_Names from users
where id in (
        select user_id
        from likes
        group by user_id
        having count(photo_id) = (
                select count(id) as Total_Photos from photos)
);
```

id	BOT_Names
5	Aniya_Hackett
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
36	Ollie_Ledner37
41	Mckenna17
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
71	Nia_Haag
75	Leslie67
76	Janelle.Nikolaus81
91	Bethany20

Result:
By completing the project, I am feeling more confident in my SQL knowledge. It really helped me to brush up on my concepts related to Sub-queries and Aggregate functions.
It also helped me to understand the table schema and how normalization can better help to understand the dataset.

Name: Umesh Sugara
Task: 2