Name: Shubham Aaba Savale
Project: Instagram User Analytics

Project description:
Firstly, what I reduced is my role during the project. So what I understood is that I would be analyzing the data provided such as user engagement and interactions with the app. Based on my analysis the various teams within the company will be utilizing the information gathered by me. So there were various tasks asked to carry out on the dataset provided.
Approach:

I started working on the project by first drawing out a plan for what I'll be doing and what queries I need. Then after outlining the plan I understood that i need to work on each query and since the data was vast, though may not be that vast but was still bigger in nature to showcase. So idecided to limit the responses that the queries will provide to 5. Then I understood what was asked and then wrote queries on a paper.and then started working on the software.

Tech-stack used:

Since we were taught SQL on MySQL Workbench I too, used the MySQL Workbench 8.0 CE. It is easy to use and I experienced that it is a great software for a learner and is very easy to use. Hence, I used the MySQL Workbench.

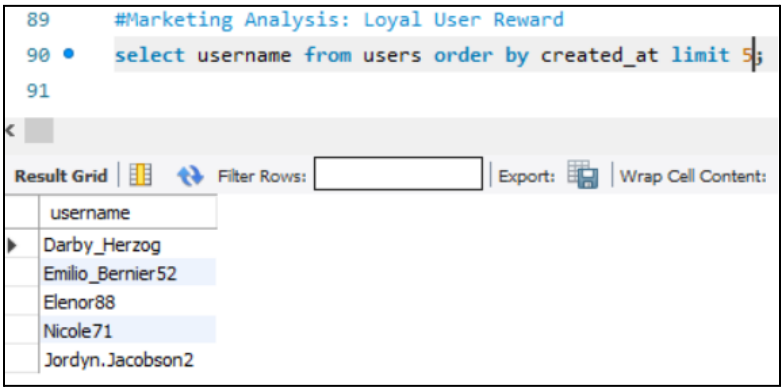
Insights:

The project was in two parts as follows.

A. Market Analysis:

The insights that I found are as follows:

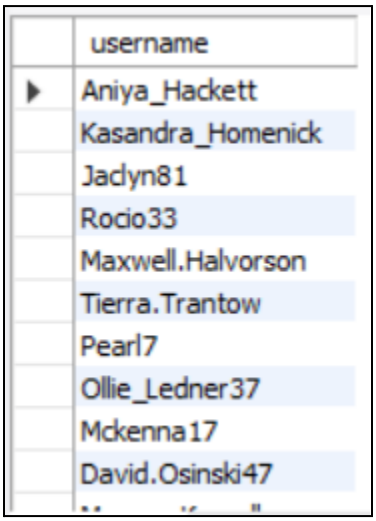
1. We were supposed to find the five oldest users on instagram from the provided database. So the Query I used and the output I got is as below:-



2. The second task was titled Inactive User Engagement. Inthis the query I used was:

```
#Marketing Analysis: Inactive User Engagement
select username from users left join photos on users.id=photos.user_id where photos.id is null;
```

and I got 26 such users that haven't posted a single photo.



username
Aniya_Hackett
Kasandra_Homenick
Jadyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47

username
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67

username
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

3. The 3rd task asks us to determine the winner of the contest that the team has organized where the user with most likes on its photo wins. So my approach was to take the users with most liked photos and show them. I considered the ranking of only Top winner so limited my response to 1. I also saw that the top 3 has 2 users with same likes so if we want top 3 we will need to tie for 2nd rank and there will 4 winners for 3 ranks. But I considered only the winner of the contest. Hence, the query I finally used is as follows:

```
#Marketing Analysis: Contest Winner Declaration
select username, photos.id, photos.image_url, count(likes.user_id) as total from photos
inner join likes on likes.photo_id=photos.id
inner join users on photos.user_id=users.id
group by photos.id order by total desc
limit 1;
```

And, the winner is:

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

4. The next task was to carry out hashtag research and to find out the top 5 most commonly used hashtags on the platform. So, query used is :

```
#Marketing Analysis: Hashtag Research
select tags.tag_name, count(*) as total from photo_tags join tags on photo_tags.tag_id=tags.id
group by tags.id order by total desc limit 5;
```

And the most commonly used top 5 hashtags are as follows:

tag_name	total
smile	59
beach	42
party	39
fun	38
concert	24

5. The day we would like to launch any marketing campaign would be the day that most people would see, similarly here we had to find the day which had the most users registered on it so for this the query ued is:

```
#Marketing Analysis: Ad Campaign Launch
select dayname(created_at) as day, count(*) as total from users group by day order by total desc limit 1;
```

and the output we got is thursday as per the above query -

day	total
Thursday	16

B. Investor Metrics:

1. Here we had to find the metrics useful for the investor such as the user engagement and the first query was to find the average number of posts per user and also provide with the total number of photos on the instagram divided by the total users, which is nothing but the average posts per the user. So the query I used is as follows:

```
#Investor Metrics: User Engagement
SELECT COUNT(*) FROM photos;
SELECT COUNT(*) FROM users;
SELECT (SELECT COUNT(*) FROM photos) / (SELECT COUNT(*) FROM users) AS avg;
```

In this I checked first the number of users and the number of posts which were found out to be 257 and 100 and the average is 257

2. Here we were asked to identify the fake users who have liked every single photo on the platform. So we saw that certain there are 13 users which have done this and they have been identified by user id as well as the username as well by 2 queries, which are as follows:

```
#Investor Metrics: Bots & Fake Accounts
select user_id, count(*) as num_likes from likes group by user_id having num_likes = (select count(*) from photos);
select u.username, count(*) as num_likes from users u join likes l on u.id = l.user_id group by u.id having num_likes = (select count(*) from photos);
```

And the users with user id and user name are as follows

user_id	num_likes
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

username	num_likes
Aniya_Hackett	257
Jaclyn81	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

Result:
I have learnt SQL to a level that i can start working on the projects of my own and try newer databases. I am proficient with the smaller queries. But the lengthier ones take time and I'm sure with given time I would be able to do it more quickly and grow my skills of Structured Query Language. Also, I got familiar with the MySQL Workbench and now I;m able to use it efficiently.