

INSTAGRAM USER ANALYTICS PROJECT REPORT

Project Description

This project aims to analyse the Instagram users. The entire analysis involves marketing analysis as well as investor metrics. The analysis will be done in MySQL workbench using SQL queries. The database consist of 7 tables that provide data related to the user engagement with the Instagram platform in terms of posting photos, liking posts and ads, following and followers of other users, tags used etc.

Approach

The database is created first and then the database is imported to MySQL workbench for analysis.

Step 1: Understanding Data

To understand the magnitude and quality of the data, ‘users’ table is first analyzed.

Query used:

```
SELECT * FROM users
```

order by created_at

Output

MySQL Workbench interface showing the results of a SQL query. The query is: `SELECT * FROM users order by created_at`. The results are displayed in a table with columns: id, username, and created_at. The table shows 10 rows of data, including users like Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71, Jordyn_Jacobson2, Nia_Haag, Rafael_Hickle2, Aurelie71, Chaz27, and Bethany20. The output also shows the execution time and duration for each query step.

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
71	Nia_Haag	2016-05-14 15:38:50
40	Rafael_Hickle2	2016-05-19 09:51:26
58	Aurelie71	2016-05-31 06:20:57
88	Chaz27	2016-06-02 21:40:10
91	Bethany20	2016-06-02 23:31:53
26	Josianne_Friesen	2016-06-07 12:47:01
39	Kelsi26	2016-06-08 17:48:08

Output:

#	Time	Action	Message	Duration / Fetch
1	08:05:02	show tables	7 row(s) returned	0.016 sec / 0.000 sec
2	08:05:02	SELECT * FROM users LIMIT 0, 1000	100 row(s) returned	0.031 sec / 0.000 sec
3	08:06:10	show tables	7 row(s) returned	0.000 sec / 0.000 sec
4	08:06:10	SELECT * FROM users LIMIT 0, 50000	100 row(s) returned	0.000 sec / 0.000 sec
5	13:29:16	show tables	7 row(s) returned	0.079 sec / 0.000 sec
6	13:29:16	SELECT * FROM users order by created_at LIMIT 0, 50000	100 row(s) returned	0.000 sec / 0.000 sec

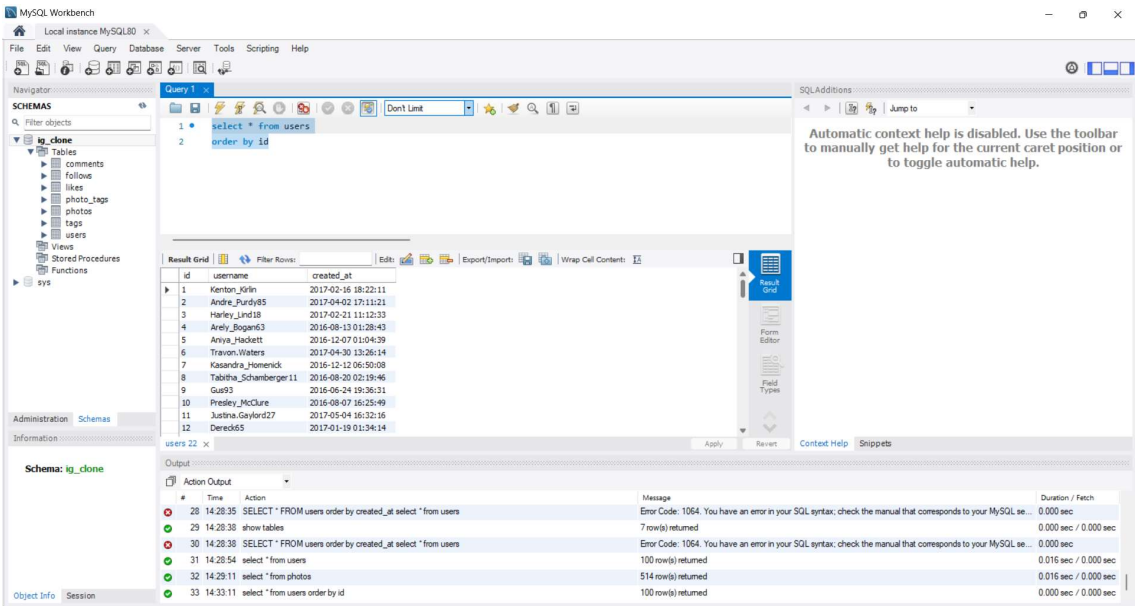
The output reveals that the data that is available for analysis including the users who created their Instagram profiles in 2016 and 2017.

Query used:

select * from users

order by id

Output



100 users data is available for analysis

Step 2: Data Analysis & Driving Insights

SQL Tasks :

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.

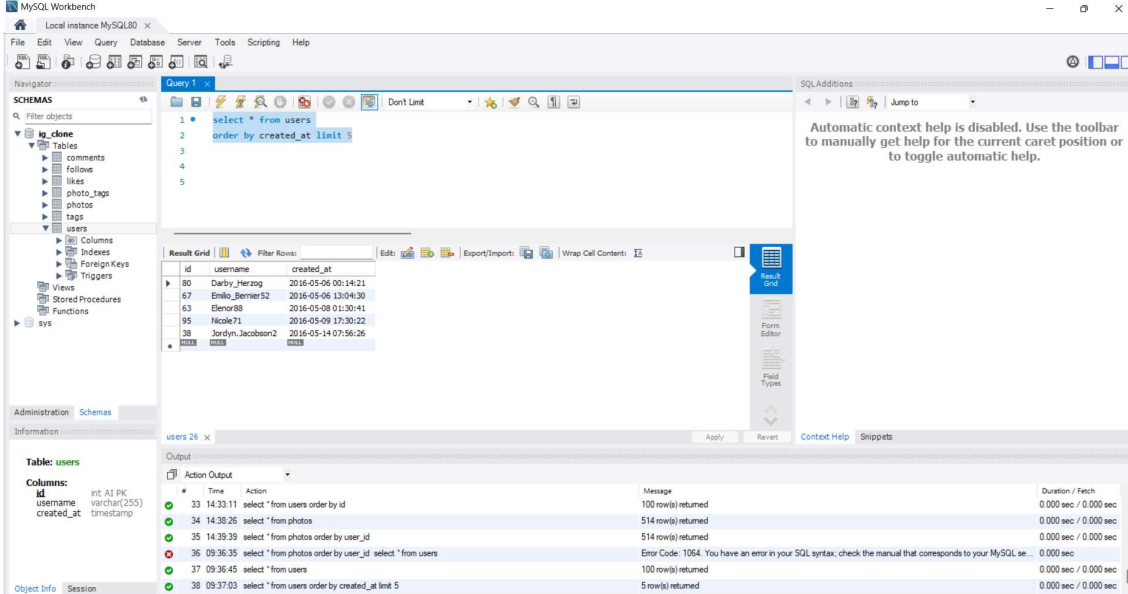
This can be determined by ordering the users in the ascending order of Instagram profile created date and then by displaying first 5 users data.

Query

select * from users

order by created_at limit 5

Output



The screenshot shows the MySQL Workbench interface. The 'Query' tab is active, displaying the following SQL query:

```
1 select * from users
2 order by created_at limit 5
3
4
5
```

The 'Result Grid' shows the results of the query, displaying the first 5 users ordered by their creation date:

id	username	created_at
80	Darby_Hierzog	2016-05-06 00:14:21
67	Emilio_Biermer52	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

The 'Output' tab is also visible, showing the execution log with the following entries:

Time	Action	Message	Duration / Fetch
33 14:33:11	select * from users order by id	100 row(s) returned	0.000 sec / 0.000 sec
34 14:38:26	select * from photos	514 row(s) returned	0.000 sec / 0.000 sec
35 14:39:39	select * from photos order by user_id	514 row(s) returned	0.000 sec / 0.000 sec
36 09:36:35	select * from photos order by user_id select * from users	Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds to your MySQL se...	0.000 sec
37 09:36:45	select * from users	100 row(s) returned	0.000 sec / 0.000 sec
38 09:37:03	select * from users order by created_at limit 5	5 row(s) returned	0.000 sec / 0.000 sec

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.

Inactive users can be found out by joining users table with photos table. To perform this task, left join of users table and photos table is necessary to identify those user ids that carry null values in the photos table and such user ids with null values indicate that they have been inactive users without posting any photos in Instagram.

Query

```
select * from users left join photos on photos.user_id = users.id
```

where photos.image_url is null order by users.username;

Output

The screenshot shows the MySQL Workbench interface. The query editor contains the following SQL query:

```
1 select * from users left join photos on photos.user_id = users.id
2 where photos.image_url is null order by users.username;
3
4
```

The Results grid displays the output of the query, showing columns: id, username, created_at, id, image_url, user_id, and created_at. The data is sorted by username. The first few rows are:

id	username	created_at	id	image_url	user_id	created_at
5	Aniya_Hackett	2016-12-07 01:04:39				
83	Bertholome.Bernhard	2016-11-06 02:31:23				
91	Bethany20	2016-06-03 23:31:53				
80	Devivy_Herzog	2016-05-06 00:14:21				
45	David.Cornel47	2017-02-05 21:23:37				
54	Duane60	2016-12-21 04:43:38				
90	Esmeralda.Mraz57	2017-03-03 11:52:27				
81	Esther_Zulauf61	2017-01-14 17:02:34				
68	Franco_Kiebler54	2016-11-13 20:09:27				
74	Hulda.Macejovic	2017-01-25 17:17:28				
14	Jacyn81	2017-02-06 23:29:16				
76	Janelle.Nikolaus81	2016-07-21 09:26:09				

The Output tab shows the execution log with the following entries:

#	Time	Action	Message	Duration / Fetch
41	09:43:31	select * from photos	514 row(s) returned	0.000 sec / 0.000 sec
42	12:52:36	select * from users left join photos on photos.user_id = users.id where photos.image_url = is null order by users...	Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds to your MySQL se...	0.000 sec
43	12:53:01	select * from users left join photos on photos.user_id = users.id where photos.image_url = is null order by users...	Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds to your MySQL se...	0.000 sec
44	14:22:56	select * from users u left join photos p on p.user_id = u.id where p.image_url is null order by u.username	Error Code: 1054. Unknown column 'users.username' in 'order clause'	0.000 sec
45	14:24:40	select * from users u left join photos p on p.user_id = u.id where p.image_url is null order by u.username	25 row(s) returned	0.000 sec / 0.000 sec
46	14:27:44	select * from users left join photos on photos.user_id = users.id where photos.image_url is null order by users...	25 row(s) returned	0.016 sec / 0.000 sec

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.

The details of the winner should include the username, user id, likes count and the photo that won the contest.

Answer

On examining the 'likes' table, user id, photos and corresponding likes are recorded but the username is not available. Hence, it is necessary to use inner join on users table and the likes table for precise analysis.

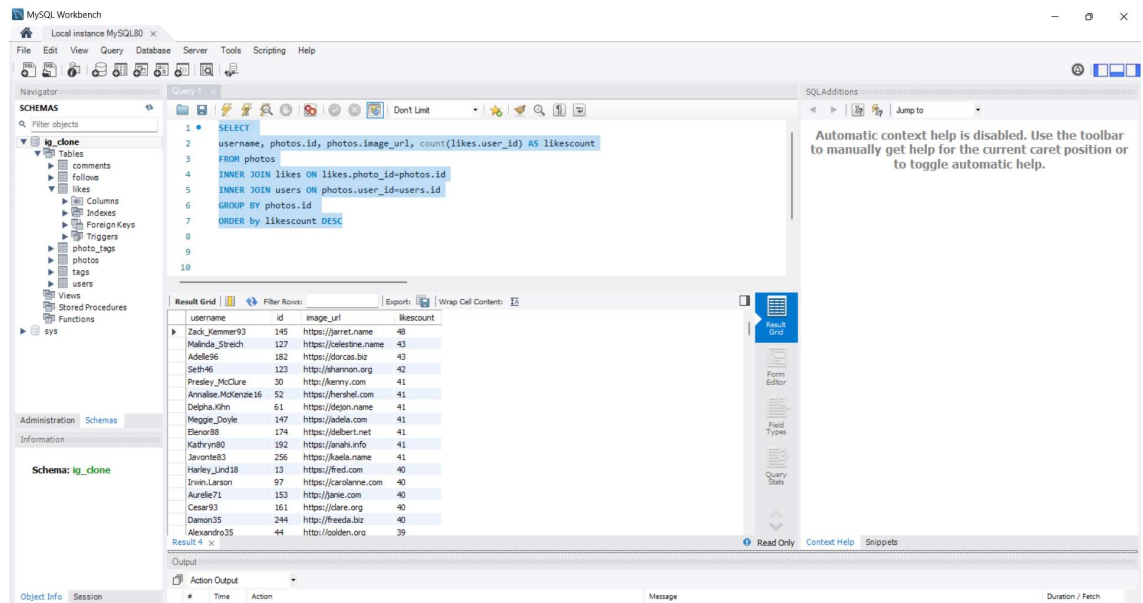
To determine the winner, it is necessary to find the total likes for each user's photos. Then, the table should be arranged based on photo id in the descending order of total likes so that the first output will be the user who got maximum likes for the posted photos.

Along with the username of the winner, user id, likes count and the url of the photo that won maximum likes should also be displayed.

Given below is the code and the output screenshot

Code

```
SELECT
username, photos.id, photos.image_url, count(likes.user_id) AS likescount
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 likescount DESC
```



- 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.

Answer

Top most used 5 hashtags need to be determined for which it is necessary to investigate photo_tags table as well as tags table.

On investigating the tags table, it lists down the various hashtags used by various users along with the tag id. On investigating the photo_tags table, the photo on which the tags are used are listed along with the corresponding tag id. Hence, for precise analysis, it is necessary to join tags table with photo_tags table.

Code

```
SELECT
```

```
tags.tag_name,
```

```
count(*) AS total
```

```
FROM photo_tags
```

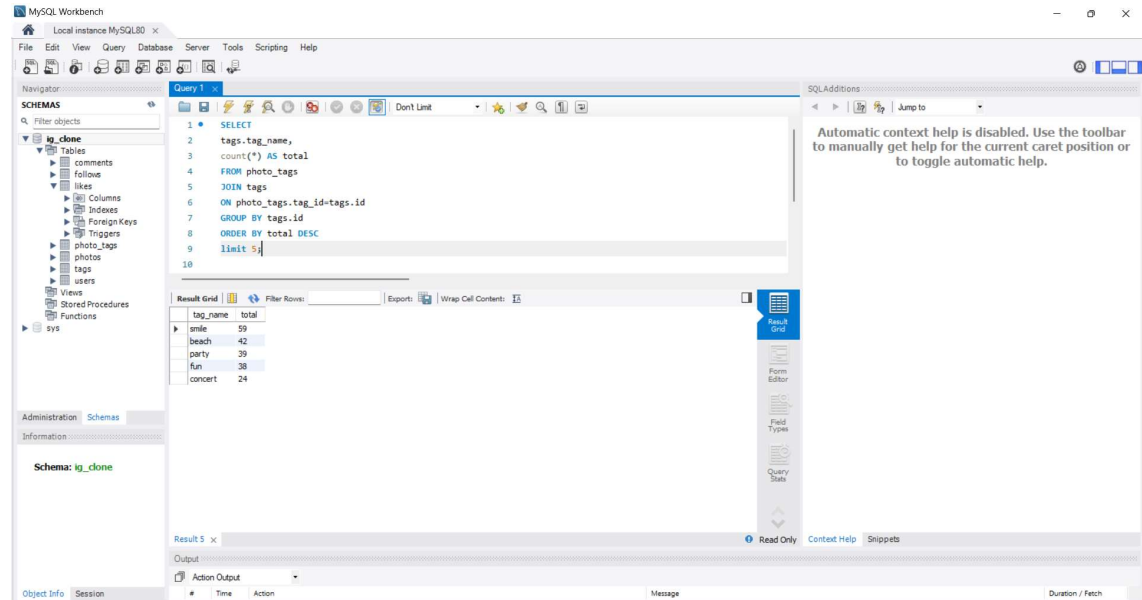
```
JOIN tags
```

```
ON photo_tags.tag_id=tags.id
```

```
GROUP BY tags.id
```

ORDER BY total DESC

limit 5;



4. **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.

Answer

On investigating the users table for analysing profile creation date, the day is not mentioned. It is not given on which day of the week users register on Instagram. Hence, it is necessary to convert the created_at date to day using DAYNAME().

After that the total number of user registrations that happen each day needs to be counted and then we need to determine the best days for AD Campaign.

Logically thinking, the best day for AD campaign is a holiday combined with the day with maximum user registrations.

Given below is the code and the output

Code

SELECT

dayname(created_at) AS day, count(*) AS totalregistrations

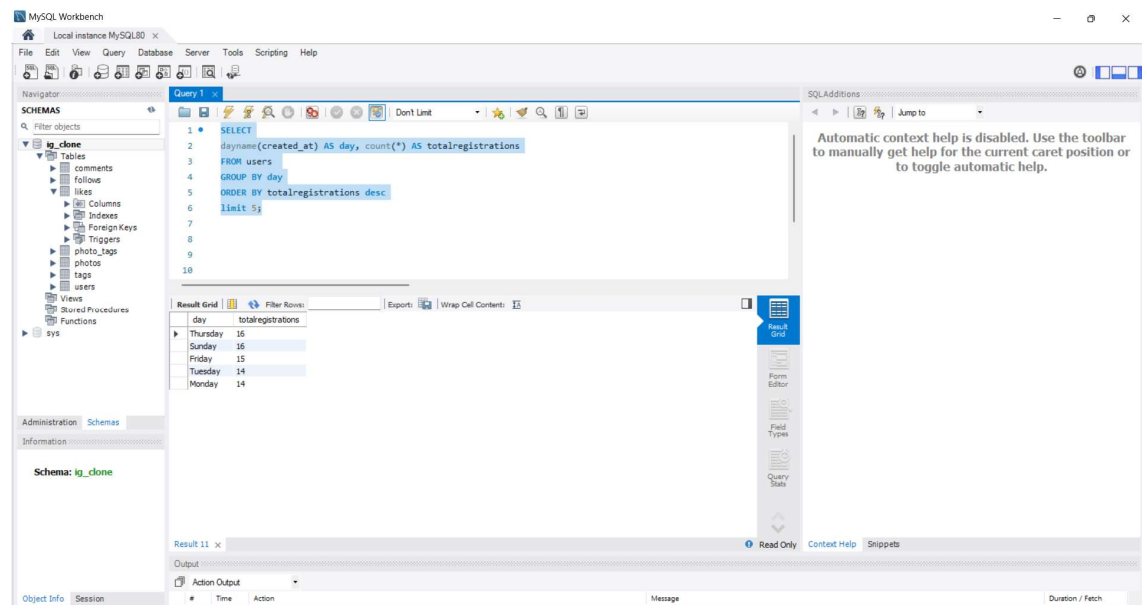
FROM users

GROUP BY day

ORDER BY totalregistrations desc

limit 5;

Output



The analysis shows that Sunday and Thursday are the days on which maximum people register in Instagram. However, Sunday is the best day for Ad Campaign because on Sundays people are relaxed and have time to click on Ads and shop.

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.

Answer

To analyse the active users in Instagram, it is necessary to determine the photos posted per user in the Instagram.

Query

```
SELECT * FROM photos,users;
```

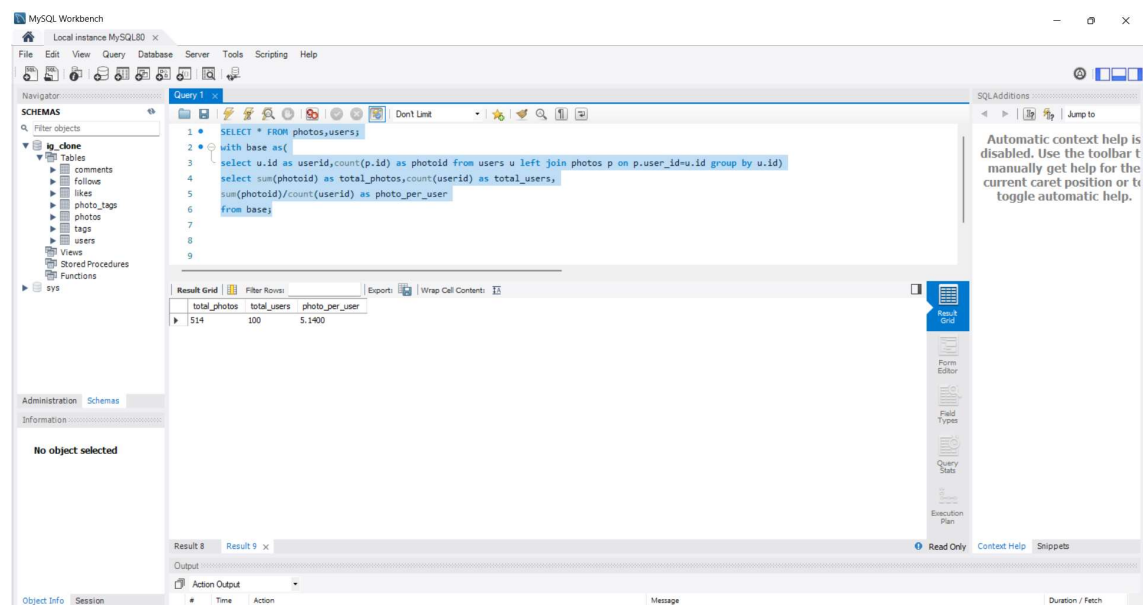
with base as(

```
select u.id as userid,count(p.id) as photoid from users u left join photos p on  
p.user_id=u.id group by u.id)
```

```
select sum(photoid) as total_photos,count(userid) as total_users,
```

```
sum(photoid)/count(userid) as photo_per_user
```

from base;



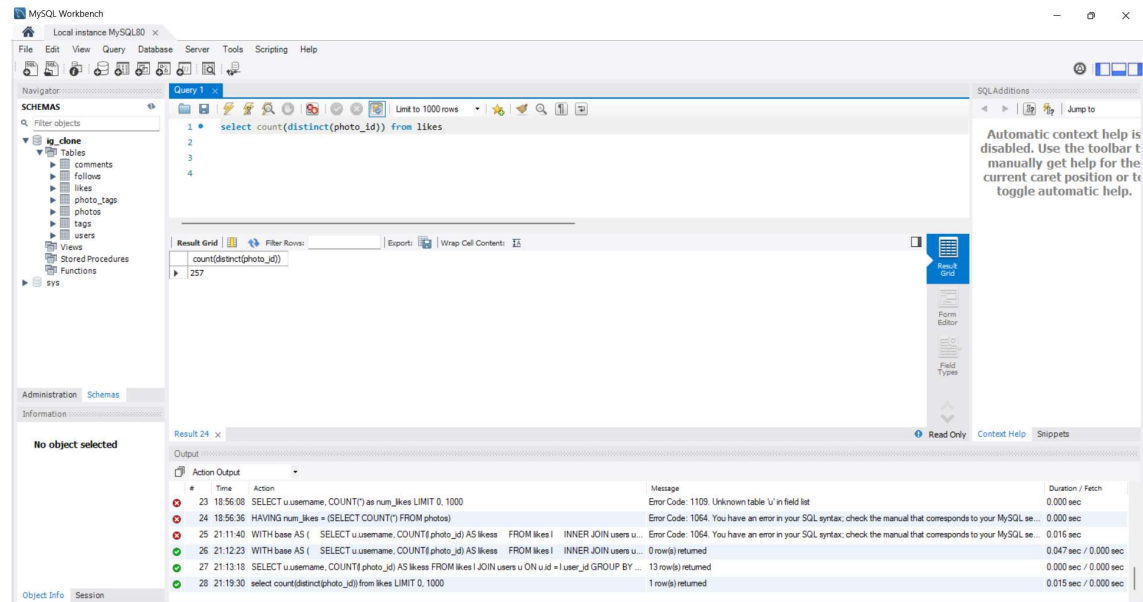
- Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.
Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Answer

To find the fake accounts, first the total count of distinct photo ids is determined. inner join of likes and users is required to count the number of likes per photo id by each username. This was done by 13 users. Hence, there are 13 fake accounts.

Query to find distinct photo id count

select count(distinct(photo_id)) from likes



Query to find the users/fake accounts

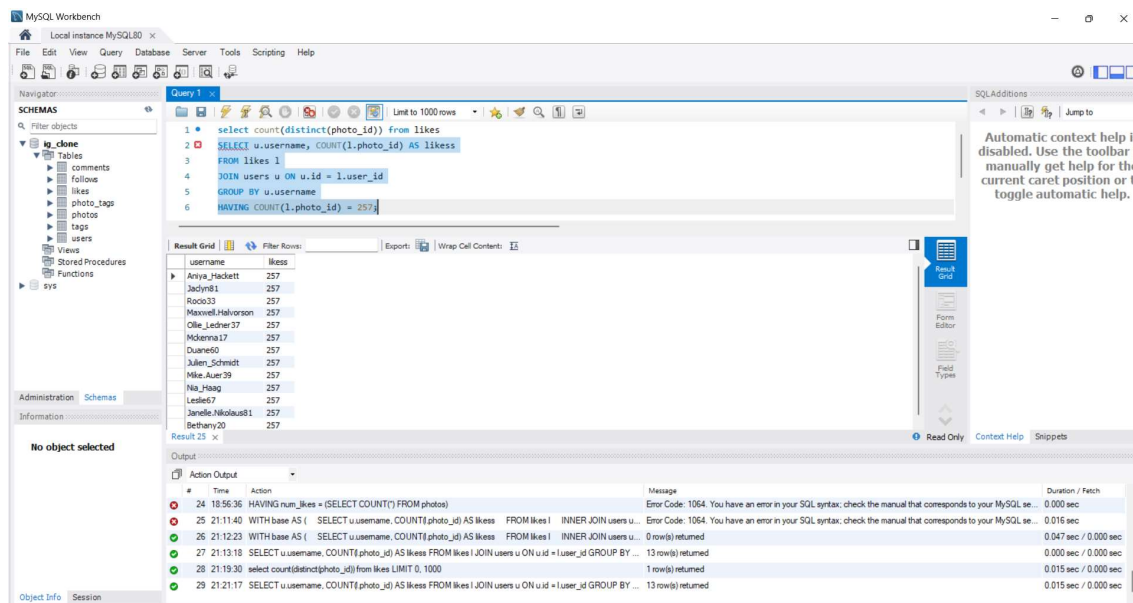
SELECT u.username, COUNT(l.photo_id) AS likess

FROM likes l

JOIN users u ON u.id = l.user_id

GROUP BY u.username

HAVING COUNT(l.photo_id) = 257;



Tech Stack Used

For this instagram user analytics project, MySQL workbench is used. It offers the facility to create and migrate database, create, execute and optimize sql queries and in addition, the facility to optimize query whenever required. MySQL workbench is also easy to use with user friendly SQL editor.

Insights

Instagram is a very popular social media platform which drags people into it as it has much more facilities than facebook or similar social media platforms can offer. In such a social media platform, it is important to analyze the data to identify various factors that can promote the platform better.

In this project, it is identified that a lot of users who have created their profile in Instagram are still inactive which could be due to their lack of knowledge on “how to post photos or videos or reels”. So, such users should be sent tutorial emails to help them use the platform well and become active in it. It can also promote the ad clicks and thereby, businesses.

Through the context to identify user with maximum photo likes, more and more users are promoted to be more active in the platform.

Through most popular hashtags identification, the businesses who want to launch ads can use these hashtags for obtaining maximum reach and engagement for their ads.

It is equally important to identify the user engagement rate. As per the available data, the average user engagement rate is 5.14. The more the platform is promoted, the better will be the user engagement rate. As the user engagement rate improves, the platform earns more and more.

As any social media platform will have fake profiles created, Instagram will also have fake accounts or profiles. It is extremely important to identify them. In this project, 13 fake accounts were detected.

Results

Through this project, I could identify 5 most used hashtags by the users in Instagram. On informing about these hashtags to the business profiles, they can use them in their ads which intends to increase the ad clicks and engagements which in turn increase the business at least by 50%. This is very significant in increasing Instagram platform's influence in the public as well as its revenue.

Another major finding is the user engagement rate. It is important to improve the user engagement rate to build more revenue for Instagram. In association with this, it is also important to find the factors that lead to low user engagement rate by analysing the data on reels created per user, videos posted per user, posts written per user, stories posted per user, time spent by each user in the platform etc. On performing these detailed analysis, more insights can be drawn.

In short, this entire project has personally improved my SQL skills, analytical thinking, problem solving skills and presentation skills.

