Second Project on Instagram User Analytics

Project description-

The project is basically based on tracking the users engagement with digital platforms. There are various teams like digital marketing,development and product teams who track this records so that they can utilize this information for launching a new marketing campaign or adding new features to build for an app which will be beneficial for their business.

So, here I am working with product team of instragram and our marketing team wants to launch some Campaign so they need our help. So I have to provide some details about instragram users.

Marketing team wants to know –

- 5 oldest instragram users
- The user who have never been posted a single photo on instragram
- The user who gets most likes on a single photo
- Identify the top 5 most commonly used hashtags on the platform
- The day on which most users register on

Also investors wants to know some details about Instragram users.They want to know-

- How many times does average user posts on instragram & also total number of photos on Instagram/total number of users
- users (bots) who have liked every single photo on the site which is not possible for any normal user.

This is the overview of overall project.

So, I have to give answers all of these queries by creating a database given in dataset. By using SQL I have to perform the entire analysis.

First of all I need to put all the data given in the dataset in SQL workbench V.8.0.31 to create a database table ig_clone. Then I have to find all the answers given by investor & marketing team by running the Query in SQL. First I read the questions many times to understand what they exactly asking me to do. Then step

by step I have written the queries and taking the screenshot of all those queries and outputs and also saved the queries in one notepad.

First I found so many errors to write an query. This also means that my thinking process for written a query was wrong.so I need to shift my thought process in other direction to solve the problem. Then I have unlocked the solution video where I have seen the multiple approach to solve a query which was quite helpful. So from here I learned how to approach for solving a query in SQL.After getting the correct output I have taken the screenshot of the result and written all the queries in one notepad for the purpose of completing the project.

Approach-

First of all I read the Project 'Instagram User Analytics' many times and try to understand what exactly they want me to do.Then I have seen all the videos related to SQL fundamentals and try to understand the questions given by marketing team & investors. Then I have seen the video of how to install My SQL. After seen this I installed the My SQL V.8.0.31 step by step.Then I put all the data given in the dataset in SQL workbench V.8.0.31 to create a database table ig_clone.Then step by step I have written the queries. First I found so many errors to write an query. This also means that my thinking process for written a query was wrong.so I need to shift my thought process in other direction to solve the problem. After that I have successfully solved the 1st query. But I get stuck in 2nd query. I tried many times to solve but all are vain. Then I have unlocked the solution video where I have seen the multiple approach to solve a query which was quite really helpful. So from here I learned how to approach for solving a query in SQL.Then I have done the hand on practice to write the correct queries and taking the Screenshot of the Output.They are-

1st task- Find the 5 oldest users of the Instagram from the database provided

Query- **SELECT id,username,created_at**

FROM

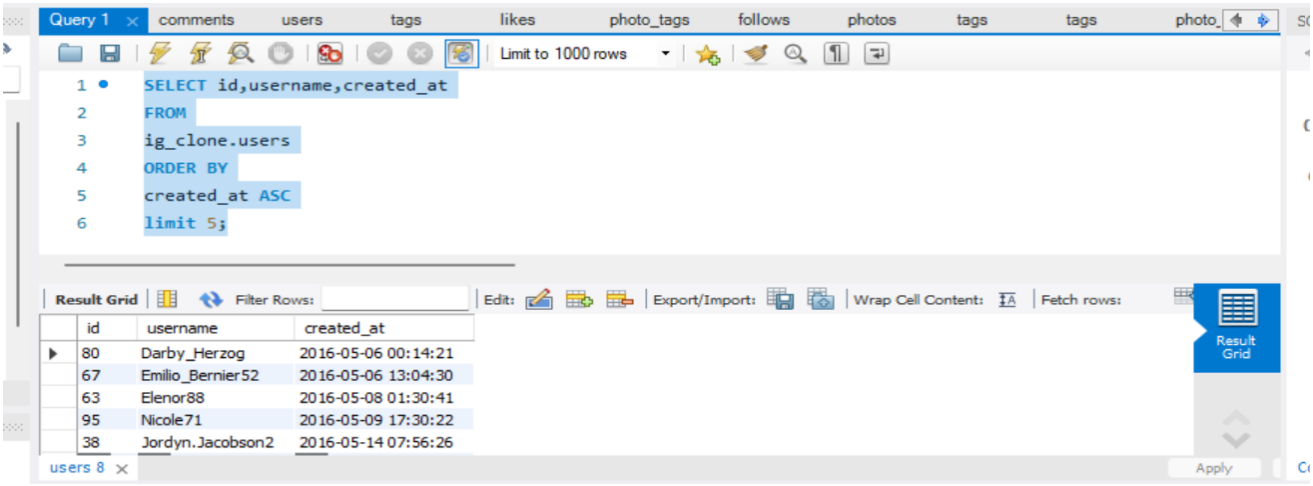
ig_clone.users

ORDER BY

created_at ASC

limit 5;

OUTPUT-



2nd Task- Find the users who have never posted a single photo on Instagram

Query- SELECT

U.username

FROM ig_clone.users U

left join

ig_clone.photos P

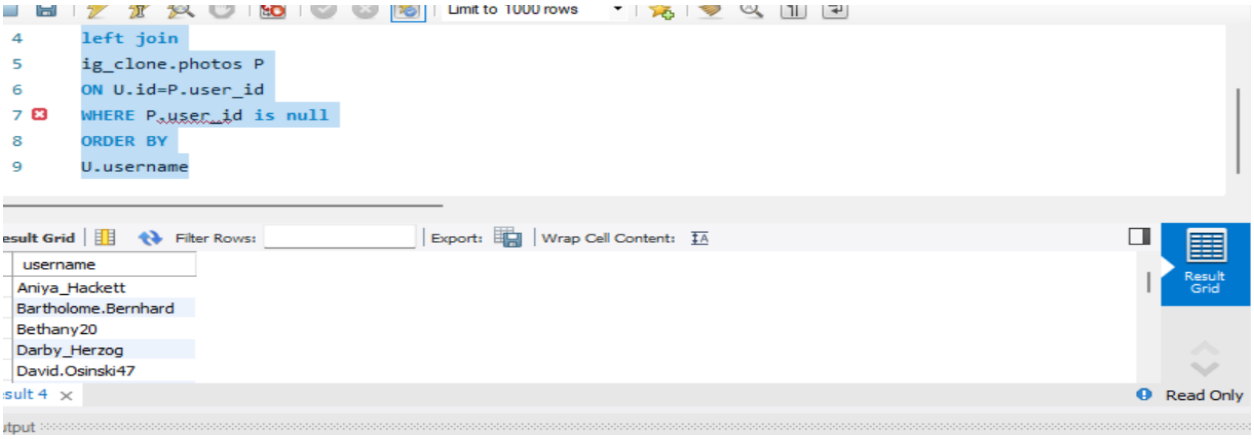
ON U.id=P.user_id

WHERE P.user_id is null

ORDER BY

U.username;

OUTPUT-



3rd Task- Identify the winner of the contest and provide their details to the team

Query- Select likes.photo_id , users.username , Count(likes.user_id) As like_user

From ig_clone.likes likes

Inner join ig_clone.photos photos

On likes.photo_id = photos.id

Inner join ig_clone.users users

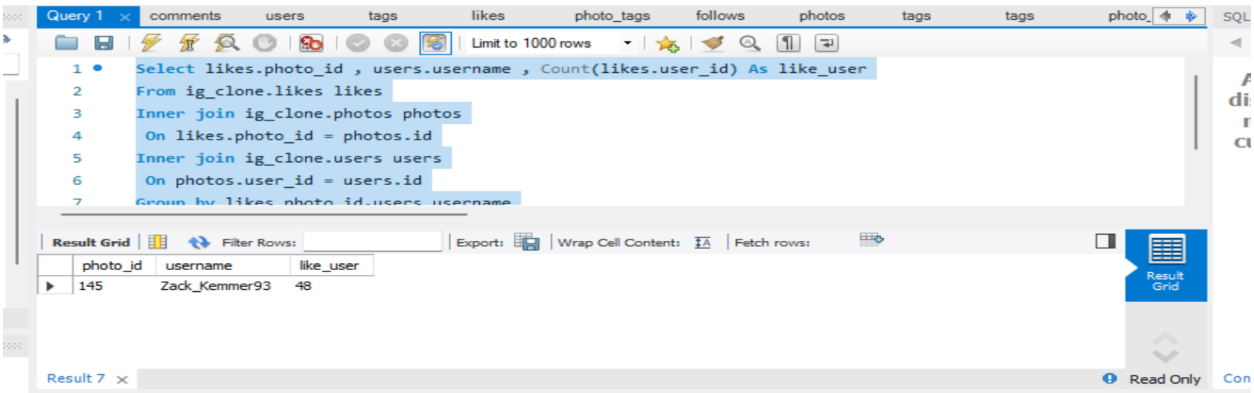
On photos.user_id = users.id

Group by likes.photo_id,users.username

Order by like_user desc

Limit 1

OUTPUT-



4th Task- Identify and suggest the top 5 most commonly used hashtags on the platform

Query- SELECT

t.tag_name,

COUNT(p.photo_id) AS num_tags

FROM

ig_clone.photo_tags p

INNER JOIN

ig_clone.tags t

ON p.tag_id = t.id

GROUP BY

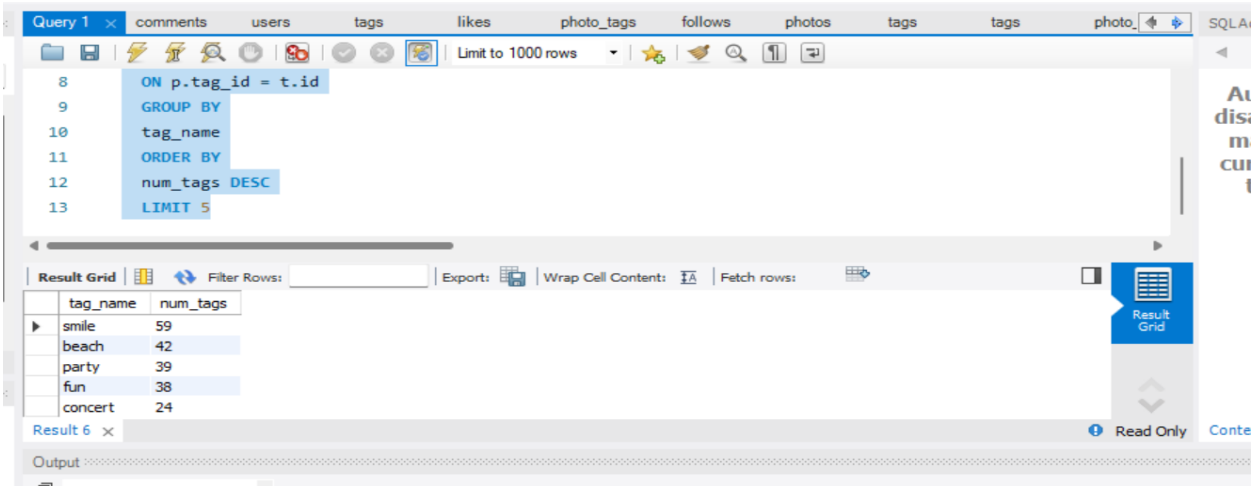
tag_name

ORDER BY

num_tags DESC

LIMIT 5

OUTPUT-

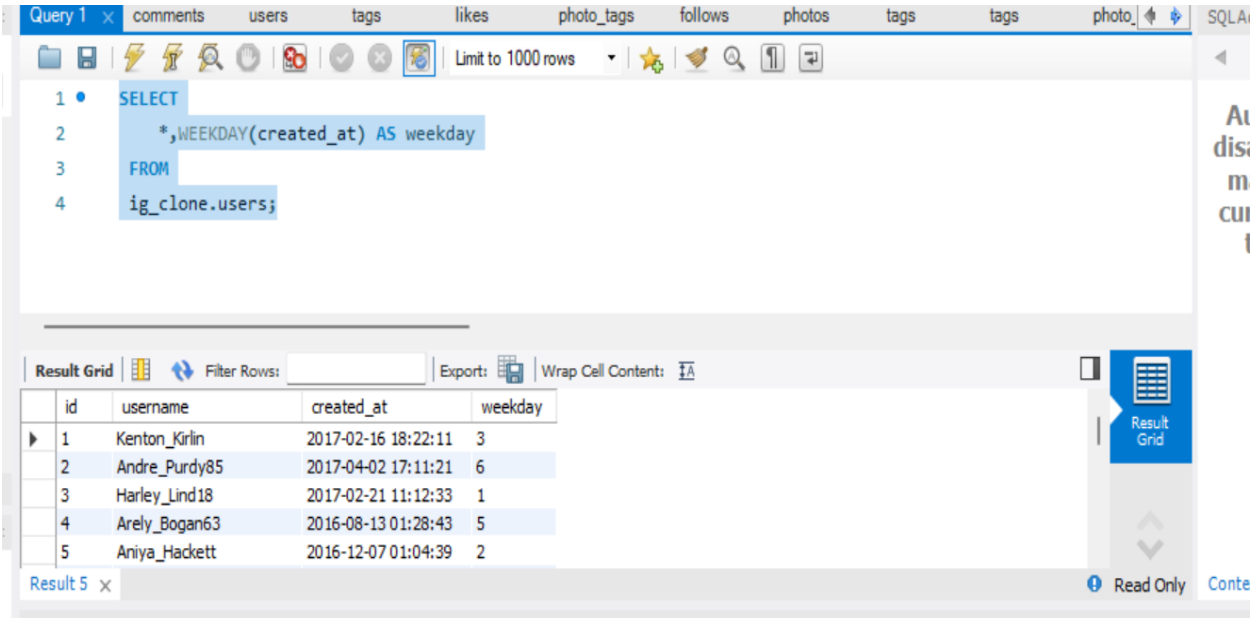


5th Task- What day of the week do most users register on? Provide insights on when to schedule an ad campaign

1st Part of the Query-

```
SELECT
    *,WEEKDAY(created_at) AS weekday
FROM
    ig_clone.users;
```

OUTPUT-



WHERE

MONDAY-0

TUESDAY-1

WEDNESDAY-2

THURSDAY-3

FRIDAY-4

SATURDAY-5

SUNDAY-6

2nd part of the Query

SELECT

WEEKDAY(created_at) AS weekday,

COUNT(username) AS num_users

FROM

ig_clone.users

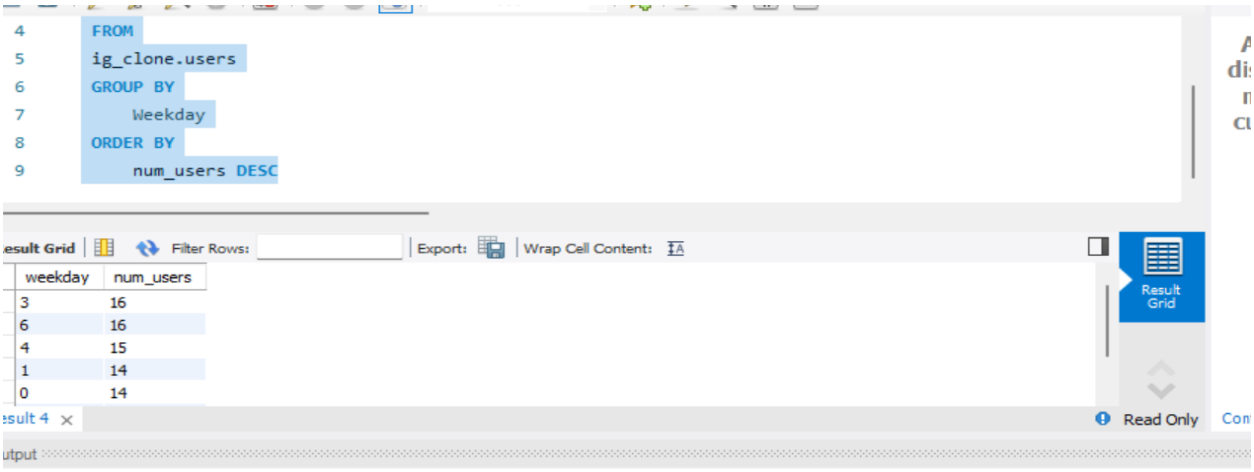
GROUP BY

Weekday

ORDER BY

num_users DESC

OUTPUT-



6th Task- Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

1st part of the Query

SELECT

u.id AS userid,

COUNT(p.id) AS photoid

FROM

ig_clone.users u

LEFT JOIN

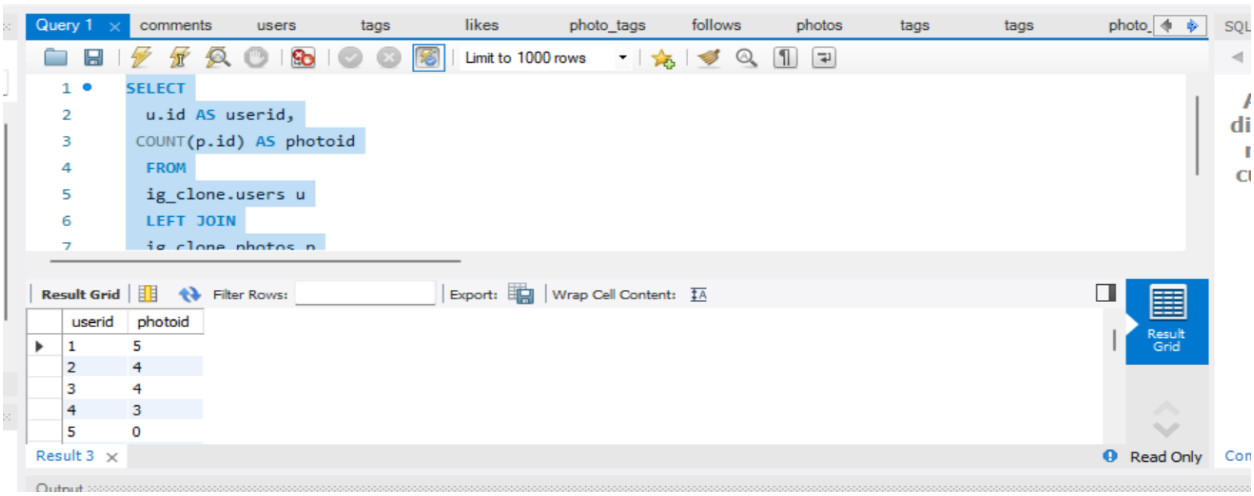
ig_clone.photos p

ON u.id = p.user_id

GROUP BY

u.id

OUTPUT-



2nd part of the Query

with CTE AS (

SELECT

u.id AS userid,

COUNT(p.id) AS photoid

FROM

ig_clone.users u

LEFT JOIN

ig_clone.photos p

ON u.id = p.user_id

GROUP BY

u.id

)

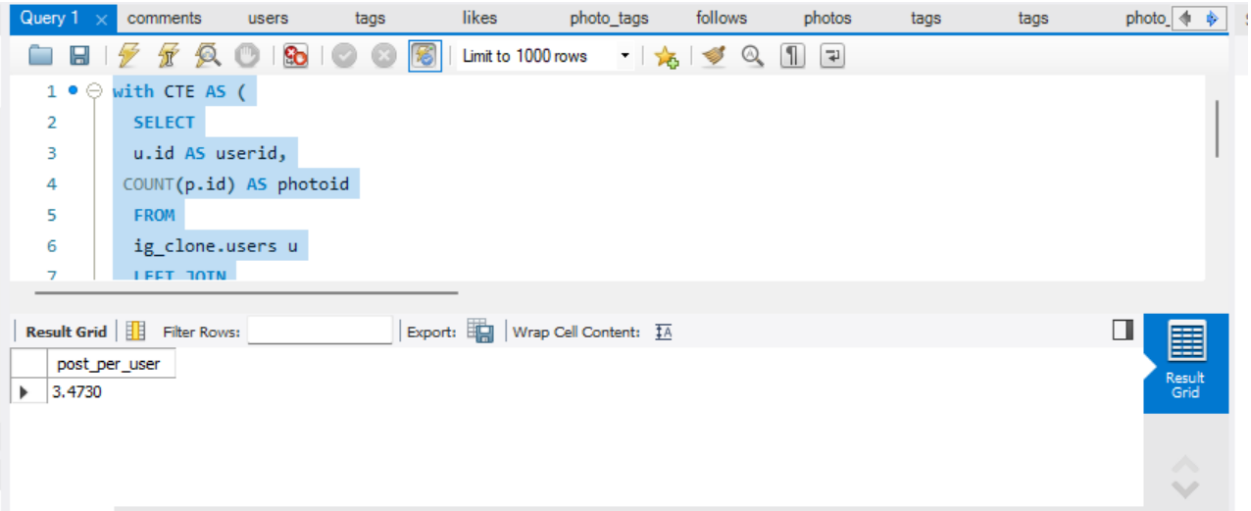
SELECT SUM(photoid)/COUNT(userid) AS post_per_user

FROM CTE

WHERE

photoid > 0

OUTPUT



7th 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)

Query

with photo_count As (

SELECT

user_id,

COUNT(photo_id) As num_like

FROM

ig_clone.likes

Group By

user_id

ORDER BY

num_like DESC

)

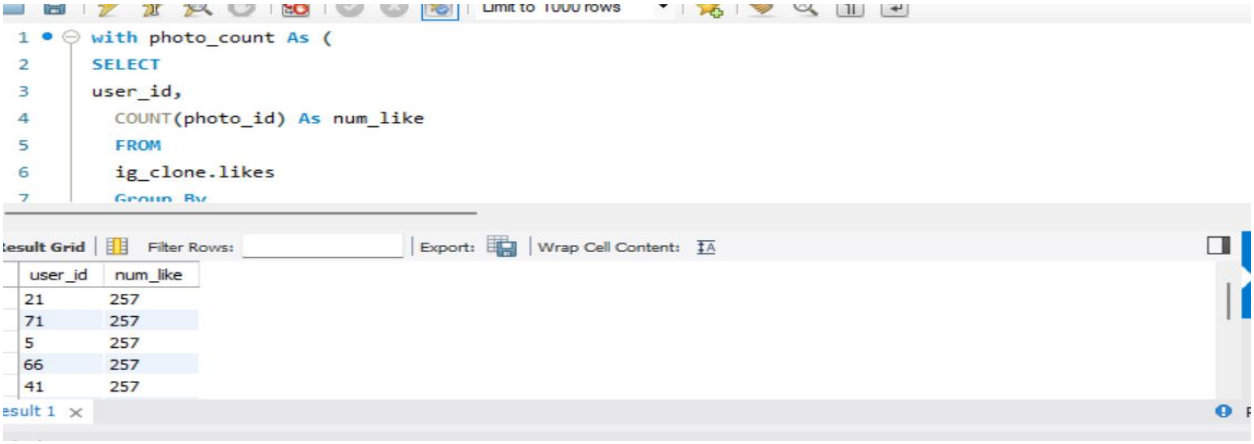
SELECT *

FROM photo_count

WHERE

num_like = (SELECT count(*) FROM ig_clone.photos)

OUTPUT-



Tech-Stack Used-

I have installed MY SQL V.8.0.31 for 64 bit system configuration. So I am using here My SQL workbench V.8.0.31 for written the queries. The purpose of using it-

- ,My SQL is an open source and free to use database system that helps to facilitate the proper and efficient management of databases.
- It is very powerful and simple to set up and easy to use. Once we have done the setup and are ready to use.
- We can perform many different operations(logical, Arithmetic e.t.c) using my SQL like create, delete a database, insert a record by using simple commands.

Insights-

In this the project I have learned lot of things about Data warehouse,data lake and database management system and also gained a clear vision about Job role of analytics is to help companies gain insight into their customers. Then, the companies can optimize their marketing and deliver a better product. Analytics gives businesses the quantitative data they need to make better, more informed decisions and improve their services. Then as a

example I got a brief idea about Netflix how it works by using analytics and how it utilizes the data to improve their business.

Result-

I have no idea about SQL. But during completing this project I have learned My SQL which is requires to be a data analytics. So this is one of my biggest achievement during this course and I think it would be beneficial for my career growth.