INSTAGRAM ANALYSIS PROJECT(Problem)

SQL Fundamentals

Difficulty Level:
Description:

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams.
These insights are then used by teams across the business to launch a new marketing campaign, 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.
You are working with the product team of Instagram and the product manager has asked you to provide insights on the questions asked by the management team.

You are required to provide a detailed report answering the questions below :

A) 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.
   Your 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.
   Your 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.
   Your 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.
   Your 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.
   Your 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

Your 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
   Your 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).

How to do this Project?

1. **Create a Database:** You are supposed to run the following commands for creating the database to work on (the database has been provided in the attachments)
2. **Perform Analysis:** Use SQL to perform your entire analysis answering the questions asked above
3. **Submit a Report:** Make a report (PDF/PPT) to be presented to the leadership team. The report should/can contain the following details:

Project Description
Give a brief about your project description i.e. what is this project about, how are you going to handle the things and what are the things that you are going to find out through the project.
Approach
Write a short paragraph about your approach towards the project and how you have executed it.
Tech-Stack Used
Do mention the software and the version used while making the project (For Eg. Jupyter Notebook, etc) and mention the purpose of using it.
Insights
Jot down the insights and the knowledge you gained while making the project. You need to write that what do you infer about the things. Make sure its brief and up to the point only. For Eg. If you got a graph then what do you understand by the graph, what changes can you make or what can you derive from the graph.
Result
Mention what have you achieved while making the project and how do you think it has helped you.
Drive Link
Save your file as a ".pdf" file and upload it to your Google Drive. Mention the sharable link **(link visibility should be set to public)** in your pdf file which you will be uploading. Do not directly upload your project.

Judgement Criteria:

SQL Understanding
The queries run should give correct output as well as should be easily understandable.

Case study completion
All the questions present must be answered completely having correct answers.

Insights

You need to use your own imagination to answer the case study while improvising it as well.

Plagiarism

Project submitted should not be copied from the internet or anywhere else, it should be your own work.

INSTAGRAM ANALYSIS PROJECT(Solution)

Description:

In this project we will look into how we can use the SQL commands to find the answers to the questions asked our client to provide them the right insight to take their business ahead.

Here we different tables available in a single database that signify the different parameters at the platform. With the help of these tables we can make the right connections within the tables to get the most accurate information out of the data.

Approach:

Here we will see if we can solve the problem in simplest manner in order to first find the right answer and then we can manually count/compare the results with the row specific commands and see it aligns with the with the whole set of end tables which is suppose to be our answer.

Tech-Stack Used:

Here we have the MySQL Workbench 8.0 CE as the software for development, the reason to select this software was the ease of use and its popularity among all the SQL development servers. This brings the benefit of getting the most help out of the internet communities in case of errors or issues.

Insights:

Top 5 loyal customers: The code written below was used for getting the answer to the top 5 loyal customers on the platform. For this the rank function was used over the created_at table to rank each member and the filtered out using the where function to get the answer.

```
show databases;

use ig_clone;

show tables;


--  ----------------------Top 5 loyal customers ------------


with ranked_users as (select *, rank() over (order by created_at) as rank1 from users)

select * from ranked_users

where rank1 <= 5;
```

OUTPUT:

id	username	created_at	rank1
80	Darby_Herzog	2016-05-06 00:14:21	1
67	Emilio_Bernier52	2016-05-06 13:04:30	2
63	Elenor88	2016-05-08 01:30:41	3
95	Nicole71	2016-05-09 17:30:22	4
38	Jordyn.Jacobson2	2016-05-14 07:56:26	5

list of inactive users:

For this task, the left join was used in order to join both the table in such as was that the list of all the users and the list users that have uploaded at least one picture cancel each other out and only the users that have never uploaded the picture stays in the list.

```
-- ---------------------- list of inactive users -----------

select * from users
left outer join photos on users.id = photos.user_id
where photos.user_id is NULL;
```

OUTPUT:

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
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:

In this part, the likes table was used to check how many likes each photo has, to do this the count function was used to check how many times each photo id was repeated in the table and rank function to give ranks to each photo based on the counts. Then the photo_id on the likes table and the ID on the photo table was joined to get the user name of each photo. Further, the users table was joined to this table with the ID row in users table to get the name of the users. Finally, the 'where' clause was used to filter out the details of the rank one which is the winner in this contest.

```
-- ----------------------Declaring Contest Winner----------------------

with winner_details as (

select * from (
select photo_id,count(photo_id) as No_of_likes, rank() over(order by count(photo_id) desc)
as ranks from likes
group by photo_id
order by No_of_likes desc
)
as winning_photo

join photos on
photos.id =winning_photo.photo_id
order by No_of_likes desc
)

select photo_id, No_of_likes, Image_url, user_id, users.username, ranks from winner_details
join users on
winner_details.user_id = users.id
where ranks = 1
order by No_of_likes desc;
```

OUTPUT:

photo_id	No_of_likes	image_url	user_id	username	ranks
145	48	https://jarret.name	52	Zack_Kemmer93	1

Hashtag Researching:

In this part of the question, the tags and the photo_tags tables where joined to get the number of times each hashtag was used and then the count of the use of the hashtag was used as measure to give ranks to these tags and only the top 5 were filtered out of it.

```
select tag_name, trending_tags, ranks from (
select tags.tag_name ,count(photo_tags.tag_id) as trending_tags,
rank() over(order by count(photo_tags.tag_id)desc) as ranks from photo_tags

join tags on
photo_tags.tag_id = tags.id

group by tags.tag_name
order by trending_tags desc) top5

where ranks <= 5;
```

OUTPUT:

tag_name	trending_tags	ranks
smile	59	1
beach	42	2
party	39	3
fun	38	4
food	24	5
lol	24	5
concert	24	5

Launch AD Campaign:

In this part of the the task, the best day for launching the AD campaign was found by converting the given date in the the users table to the week day's name and then checking how many times the particular day was repeated on the list and then ranking the list with most counts as rank 1.

```
-- ------------------Launch AD Campaign----------------------

select count(dayname(created_at)) as used_for, dayname(created_at) as which_day,
rank() over(order by count(dayname(created_at)) desc) as best_day from users
group by which_day;
```

OUTPUT:

used_for	which_day	best_day
16	Thursday	1
16	Sunday	1
15	Friday	3
14	Tuesday	4
14	Monday	4
13	Wednesday	6
12	Saturday	7

User Engagement:

In order to find the average number of number of times a user upload photo on instagram, the photos table was used to find the count of how many times a user ID was repeated to determine how many photos where uploaded by each user and then the sum of all the total photos and the average of the same was then used as a measure the average of the same.

```
select
avg(No_of_uploads) as avg_uploads, sum(No_of_uploads) as total_Photos from
(select count(user_id) as No_of_uploads from photos
group by user_id) as count_of_uploads;
```

OUTPUT:

avg_uploads	total_Photos
3.473	257

Bots & Fake Accounts:

In this part the list of accounts were to obtained which liked all the photos on the platform to determine the fake accounts. This was achieved by checking which user id has liked how many photos from the platform. Once that was achieved, the results were compared with the number of photos on the platform and then the ones that have likes all the photos were considered to be the fake accounts.

```
-- ----------------------Bots & Fake Accounts---------------------------------

select user_id, No_of_likes from (

select count(photo_id) as No_of_likes, user_id from likes
group by user_id
order by No_of_likes desc

) as bot_users

where No_of_likes = 257  ;
```

OUTPUT:

user_id	No_of_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

Results:

This project has been helpful for deep understanding of joints and basic functions of the SQL and how it can be used to find the most accurate data out of a large database. I think it is necessary to understand what we actually need out of data and where everything is stored in order to play with the whole database with ease.

This project has given me the confidence to use the SQL at a professional level and with some more fine tuning the code format can be improved.

FULL CODE:

```
show databases;

use ig_clone;

show tables;


--  ----------------------Top 5 loyal customers ------------

with ranked_users as (select *, rank() over (order by created_at) as rank1 from users)

select * from ranked_users

where rank1 <= 5;


-- ---------------------- list of inactive users -----------


select users.id, users.username from users

left outer join photos on users.id = photos.user_id

where photos.user_id is NULL;


-- ----------------------Declaring Contest Winner----------------------
```

```
with winner_details as (

select * from (

select photo_id,count(photo_id) as No_of_likes, rank() over(order by count(photo_id) desc)
as ranks from likes

group by photo_id

order by No_of_likes desc

) as winning_photo

join photos on

photos.id =winning_photo.photo_id

order by No_of_likes desc

)

select photo_id, No_of_likes, Image_url, user_id, users.username, ranks from winner_details

join users on

winner_details.user_id = users.id

where ranks = 1

order by No_of_likes desc;

-- ------------------------Hashtag Researching----------------------
```

```
select tag_name, trending_tags, ranks from (
select tags.tag_name ,count(photo_tags.tag_id) as trending_tags,
rank() over(order by count(photo_tags.tag_id)desc) as ranks from photo_tags

join tags on
photo_tags.tag_id = tags.id

group by tags.tag_name
order by trending_tags desc) top5

where ranks <= 5;

-- ------------------Launch AD Campaign-----------------------

select count(dayname(created_at)) as used_for, dayname(created_at) as which_day,
rank() over(order by count(dayname(created_at)) desc) as best_day from users
group by which_day;

--  --------------------Investor Metrics----------------------------------

-- ---------------------User Engagement-------------------------------------
select
avg(No_of_uploads) as avg_uploads, sum(No_of_uploads) as total_Photos from
(select count(user_id) as No_of_uploads from
photos group by user_id) as count_of_uploads;
```

```
-- ----------------------Bots & Fake Accounts---------------------------------

select user_id, No_of_likes from (

select count(photo_id) as No_of_likes, user_id from likes
group by user_id
order by No_of_likes desc

) as bot_users

where No_of_likes = 257  ;
```