Instagram User Analytics Project

Project Description

As a data analyst working with the product team at Instagram, I have analysed user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business. For example, the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

In this project, I have used SQL and MySQL Workbench as your tool to analyse Instagram user data and tried to find answer to questions posed by the management team. The insights will help the product manager and the rest of the team make informed decisions about the future direction of the Instagram app.

I hope the findings would potentially influence the future development of one of the world's most popular social media platforms.

Approach

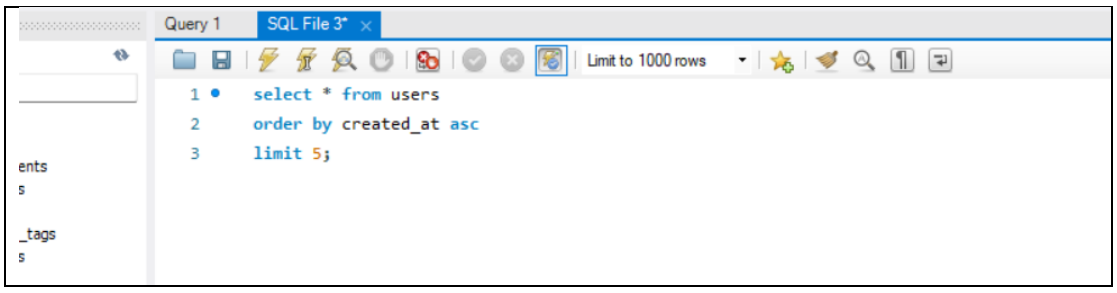
I have approached each question keeping myself in the foot of the team.

Tech-Stack Used

MySQL Workbench 8.3.0,
SQL Server Management Studio 8.0.37
It has very efficient UI and very fast.

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.



Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell

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
NULL	NULL	NULL

Analysis : - The five oldest loyal users are listed above. They could be given loyalty reward. This would increase the participation of more users in future launch of any feature or event.

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.

Limit to 1000 rows

```
1  select distinct id, username
2  from users
3  where id not in (select distinct user_id from photos)
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap C

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

These are the users who have never posted photos and could be send notification of promotional email.

3. **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.

```sql
select u.username, a.user_id,a.id as `photo id`, a.`no of likes` from users u
join
(select p.user_id, p.id, count(*) as `no of likes` from photos p join likes l on p.id =l.photo_i
group by p.user_id, p.id
order by `no of likes` desc) a
where u.id = a.user_id
```

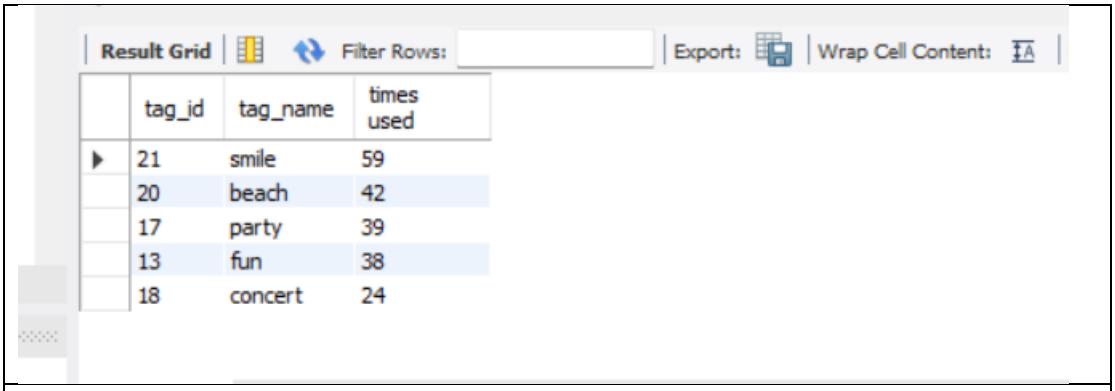
username	user_id	photo id	no of likes
Zack_Kemmer93	52	145	48
Malinda_Streich	46	127	43
Adelle96	65	182	43
Seth46	44	123	42
Presley_McClure	10	30	41
Annalise.McKenzie16	16	52	41
Delpha.Kihn	20	61	41
Meggie_Doyle	55	147	41
Elenor88	63	174	41
Kathryn80	72	192	41

Result 39

Analysis : Zack_Kemmer93 with photo id 145 has received max likes. We can award him and make people aware that these kinds of award exist as a result . People would create more impactful content.

4. **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.

```sql
select p.tag_id, t.tag_name ,count(*) as `times used`
from photo_tags p
join tags t
where p.tag_id = t.id
group by p.tag_id
order by `times used` desc
limit 5;
```



tag_id	tag_name	times used
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

Analysis : These are the most commonly used hastags. We can send advertises the list and can charge more per advertisement with these hash tags as a result revenue will increase for Instagram as these hashtags are mostly viewed and hence CTR (Click Through rate) would rise as more people would watch the ad.

5. **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.

```
select DAYNAME(created_at)  as day_of_week, count(*) as `total user`
from users
group by day_of_week
order by `total user` desc
```

day_of_week	total user
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Result 60

Analysis : - Most of the users register themselves on Thursday and Sunday. So, any ad campaign can be scheduled on or around these days.

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.

Assumption : Total post includes photos plus comments

```sql
with cte1 as (select count(*) as photo_count
            from photos ) ,
cte2 as (select count(*) as user_count
            from users ),
cte3 as ( select count(*) as comment_count
        from comments
)

select ((cte1.photo_count+ cte3.comment_count)/ cte2.user_count) as `average photo per user`
from cte1, cte2, cte3
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

average photo per user
77.4500

Total Photos divided by total users

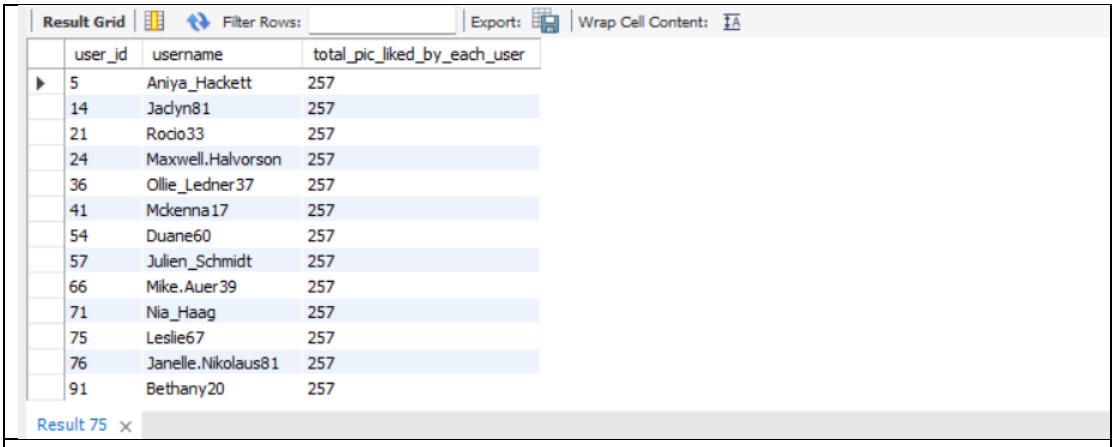
```sql
with cte1 as (select count(*) as photo_count
            from photos ) ,
cte2 as (select count(*) as user_count
            from users )

select cte1.photo_count/ cte2.user_count as `average photo per user`
from cte1, cte2
```

Result Grid | Filter Rows: | Export:

average photo per user
2.5700

2. **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.

```sql
select l.user_id, u.username ,count(*) as `total_pic_liked_by_each_user`
from likes l join users u
where l.user_id = u.id
group by user_id
having `total_pic_liked_by_each_user` = (select count(*) from photos)
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

user_id	username	total_pic_liked_by_each_user
5	Aniya_Hackett	257
14	Jaclyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

Result 75

These account seems like bot account, however to reach any conclusion we need more data like are they even posting, if they are posting is that impactful which means are people even liking the post, do they have followers or are they just following.

Insights

There are lot of active users, we can find lots of Bots account as well. The bot account findings need to be suffice with a lot more data and observation over time. The loyal customers are the once who have created the account long ago and loyalty rewards to them will ensure more such loyal accounts vis a vis new service launch. The major source of revenue are advertises. The advertisement model could be designed in a manner that max revenue would come if they wants to use most viewed hashtags. In this way revenue could be maximized. More over customer engagement is the top most priority and hence any we can send notifications to customer who have not posted any photos or comments till now. Posting of impactful content ensures sustaining of any social media website, this could be achieved by ensuring people with most photo likes are rewarded.

Result

The findings above can increase the revenue from advertisers and increase posting of more impactful content.