SQL PROJECT ON INSTAGRAM USER ANALYTICS

This project aims to analyze Instagram user interactions and engagement with the Instagram app to provide valuable insights that can help the marketing team make better decisions for the growth of the business by helping them figure out the following mentioned problems:

MARKETING ANALYSIS:

1. Finding the most loyal users in order to reward them. i.e those who have been using the platform for the longest time?
   My approach for this is to Identify the five oldest users on Instagram from the provided database.
2. Finding inactive users who have never posted a single photo on Instagram and helping them engage by sending promotional emails.

   My approach: Identifying the users who have never posted a single photo on Instagram.

3. The marketing team had organized a contest where the user with the most likes on a single photo wins. The purpose of the contest was to encourage users to engage more and more.
   My approach: Determine the winner of the contest (i.e user with the most likes on a single photo) and provide their details to the team.
4. Finding the most popular hashtags as a partner brand wants to use those hashtags in their posts to reach the most people.
   My approach: Identifying and suggesting the top five most commonly used hashtags on the platform.
5. The team wants to launch an ad campaign, and for this, they need to know the best day of the week to launch ads.
   My approach: Determining the day of the week when most users register on Instagram and providing insights on when to schedule an ad campaign.

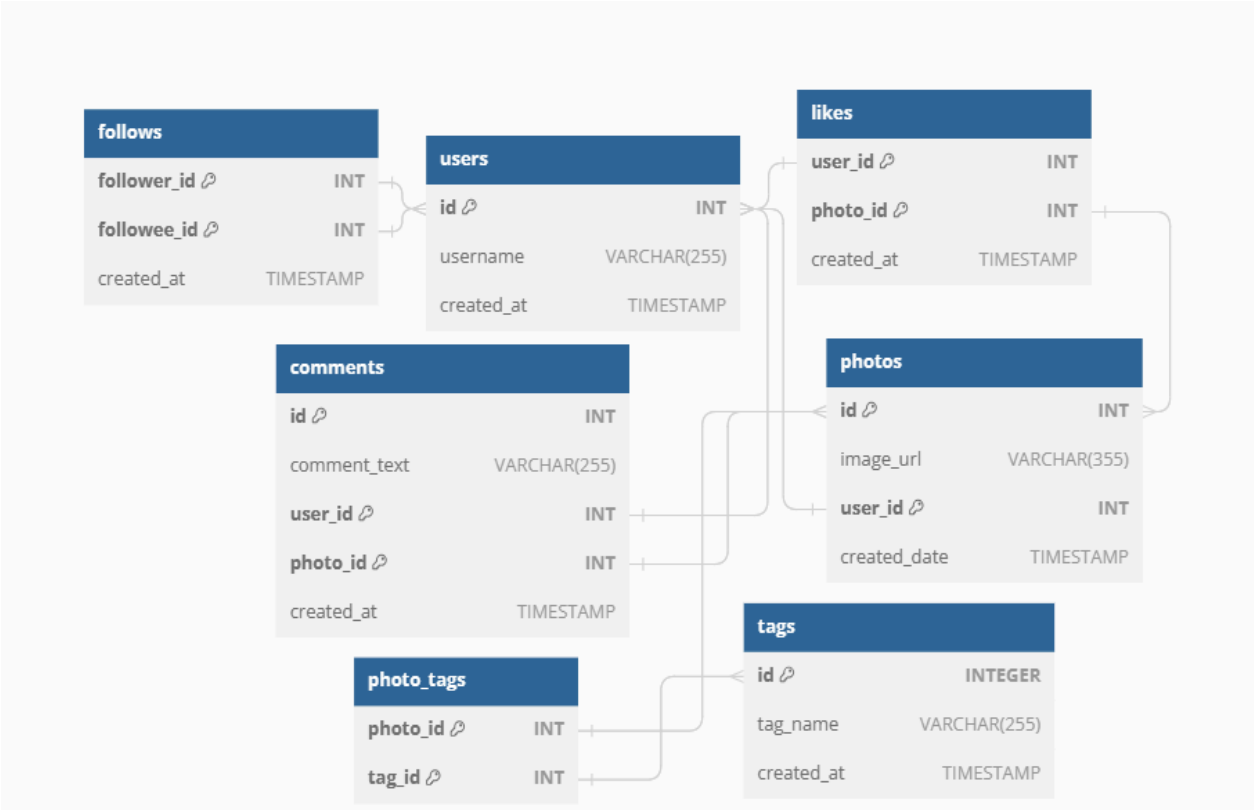
Also, we have extracted some insights about **Investor Metrics**.

1. **User Engagement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
   My approach: Evaluate 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.
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.

Database and tools:

- MySQL
- MySQL Workbench 8.0 CE

Schema- ig_clone



SQL Query and output:

QUERY-1

```
138
139     # REWARDING MOST LOYAL USERS (5 OLDEST USERS)
140
141 •   Select *
142     from users
143     order by created_at
144     limit 5;
145
146
147
```

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

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

Identifying loyal users can help to increase user engagement by giving users a reason to be more actively using Instagram on a regular basis. Further, it can help in user retention. The 5 most loyal users (i.e those using it for a longer time) are of id 80,67,63,95 and 38.

QUERY-2

```
145
146     # IDENTIFYING USERS WHO HAVE NOT POSTED AN IMAGE
147
148 •   select id, username
149     from users where id not in
150     (select user_id
151     from photos
152     );
153
154
```

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

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

users 87

Here, we have identified those users who have not posted a single image since their signing up and the aim is to convert these inactive users to active users who are engaging in the platform by sending them promotional emails and encouraging them to follow up the trends.

Query-3

```
101
102    # contest winner declaration
103 •  with cte as
104    (
105    select l.photo_id, count(l.user_id), p.user_id
106    from likes as l
107    join photos as p
108    on l.photo_id= p.id
109    group by 1
110    order by count(l.user_id) desc
111    limit 1
112    )
113    select u.id, u.username, c.photo_id from users as u
114    join cte as c
115    where c.user_id=u.id;
```

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

id	username	photo_id
52	Zack_Kemmer93	145

It was observed that the user with id 52 had the most likes on his photos, and thus, he was the winner of the contest.

The motive behind this contest was to encourage users to engage more and more.

Query-4

```
117     # TOP 5 MOST COMMONLY USED HASHTAGS
118
119 •   select t.tag_name, count(pt.photo_id)
120     from tags as t
121     join photo_tags as pt
122     on pt.tag_id = t.id
123     group by 1
124     order by count(pt.photo_id) desc
125     limit 5
126     ;
127
128
```

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

tag_name	count(pt.photo_id)
smile	59
beach	42
party	39
fun	38
concert	24

Using the right set of hashtags can enable the content to be found by right people. So, hashtag research can help boost engagement, expand audience and build brand awareness of the partner brand.

The hashtag used most is #smile.

QUERY-5

```
129
130     # best day of the week for ad campaign
131
132 •   select dayname(created_at) ,count(dayname(created_at))
133     from users
134     group by dayname(created_at)
135     order by count(dayname(created_at))desc ;
136
137
138
```

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

dayname(created_at)	count(dayname(created_at))
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Knowing when most of new users are registering can help us the ad campaign to reach them .

The best days of the week for ad campaign are Thursday and Sunday, as both these days have the highest number of user registrations.

INVESTOR METRIC:

A. Finding user Engagement:

```
153  #USER ENGAGEMENT
154
155  select
156  count(distinct(p.id)) as `photos posted`,
157  count(distinct(u.id)) as `no of users`,
158  round( cast(count(distinct(p.id)) as float) / count(distinct(u.id)) ,2.0) as `avg posts per user`
159  from photos as p
160  right outer join users as u
161  on p.user_id=u.id
162  ;
163
164
```

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

photos posted	no of users	avg posts per user
257	100	2.57

So, there are in total 257 rows i.e. 257 photos in the photos table and 100 rows i.e. 100 ids in the users table which makes the desired output to be 257/100 = 2.57 (avg. users posts on Instagram).

B.Fake and Bots account detection:

```sql
# BOT DETECTION

with cte4 as(
select user_id from likes
group by user_id
having count(photo_id)=(select count(id) from photos)
)
select k.user_id, u.username
from cte4 as k
join users as u
on k.user_id=u.id;
```

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

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

Out of the total user id there are 13 such user id who have liked each and every post on Instagram (which is not practically possible). So, such users are considered as BOTS and Fake Accounts.