INSTAGRAM USER ANALYTICS

PROJECT DESCRIPTION –

As we all know that Instagram is one of the most used social sites in today's world. It has attractive and user-friendly interface. You can share your favourite photo. It allows users to upload media that can be edited with various in built filters and hashtags. Instagram is one of the best as much as affordable social media platform for Digital market as well. All the Instagram data has been stored when we like or comment on someone's picture. To understand all these, I created a small project that show some of the features of Instagram.

As a start, I created a database by the name Instagram ig_clone. The database ha total 7 table as mentioned below.

- USER
- PHOTOS
- COMMENTS
- LIKES
- FOLLOWS
- TAGS
- PHOTO_TAGS

As per given data, we can find that user table consists of 3 attributes, id being the primary key and the photos table has its own primary key by the name ID.

PROJECT APPROCH –

The project was fairly simple, I cloned the project and executed the commands as given in the data set by the Guvi_tech team. Once the database was created, various sorting and data extracting queries were used to get the data/insights required.

Tech Stack Used –

I have downloaded MYSQL workbench 8.0 CE, it provides me easy user experience in SQL environments. It provides best visual console and better visibility in dataset.

Project Insights:

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.

   Task - Find the 5 oldest users of the Instagram from the provided database.

Query –

```
select * from users
order by created_at
limit 5;
```

Query explanation –
SELECT - it helps you to select the database.
FROM - tells the database to select data from the user table.
(*) tells to select all data in that data set.
ORDER BY- Used to sort the data according to our command.
LIMIT - To limit this much value only.

Output –

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

Conclusion - Users 80, 67, 63, 95, 38 are the 5 oldest users on the platform.

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.

   Task – To Identify users who have never posted a single photo on Instagram.

Query –

```
select username from users as u
left join photos as p
on u.id = p.user_id
where p.id is null;
```

Query explanation –
Used Join syntax to join table to print an exact detail of the users.

Output –

Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

Conclusion – There are 26 users mentioned above who have never posted a single photo on Instagram..

3. Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo win.

   Task -To find who have most likes on single photo.

Query –

```
Select username, p.id, p.image_url, count(l.user_id) as Total_likes
from photos as p
join likes as l
on l.photo_id=p.id
```

```
join users as u
on p.user_id=u.id
group by p.id
order by Total desc
limit 1;
```

Output –

Zack_Kemmer93	145	https://jarret.name	48

Conclusion - Result shows the user whose photo is most liked with 48 likes on the Instagram.

4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

   Task - To Identify top five most commonly used hashtags.

```
Query –
Select t.tag_name, count(*) as Total
from photo_tags as pt
join tags as t
on pt.tag_id=t.id
group by t.id
order by Total desc
Limit 5;
```

Output –

smile	59
beach	42
party	39
fun	38
concert	24

Conclusion – The above 5 hashtags has been used most frequently (The number indicate that how many time such hashtag has been used)

5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

   Task – To Determine the day of the week when most users register on Instagram.

   Query –

```
Select
dayname(created_at) as day, count(*) as Total
```

```
from users
group by day
order by total desc
limit 1;
```

Output –

Thursday	16

Conclusion – On Thursday, most of the user has been registered on Instagram.

A) 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.

Task – To 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.

Querry –

```
Select
(select count(*) from photos) / (select count(*) from users) as Phtos_per_users;
```

Output –

2.5700

Conclusion – The average of photos per user is 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.

Querry –

```
select u.id,u.username, count(*) as num_likes
from users as u
join likes as l
on u.id = l.user_id
group by u.id
having num_likes = (select count(*) from photos);
```

Output –

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

Conclusion – As we can see above, there is 13 fake accounts in Instagram as per given dataset.

Results – While doing this project I found out so many important terms of MYSQL that helps in solving complex problems irrespective of how large the database is. Data analysis using SQL queries to extract insights from database by which we track how users engage and interact with our digital platform in an attempt to derive business insights for marketing, product & development teams.

Thank you.

Shubham Shrikrishna Terekholkar.