Instagram User Analytics

Table of content

Sl.No	Content	Page No
1	Introduction	3
2	Tech-stack used	4
3	Approach	4
4	SQL Queries	5
5	Results	14
6	Conclusion	19

1. Introduction

User analytics we analyze how user interact with the software application. Instagram user analytics, in this project with help of given datasets and SQL queries analyze how users are engaged with instagram. The queries and its results give insights about how the users are interacting with the application. The insights help to analyze business more clearly.

Running SQL queries for certain tasks like marketing and investor metrics. In marketing analyze when the marketing team wants to launch advertisement campaign for that we have sub tasks which include rewarding most loyal user, remind interactive user to start posting, declaring contest winner, hash tag researching and launch AD campaign. In investor metrics, the investor wants to know how instagram is performing and it is not redundant like facebook, so we have like sub task to analyze user engagement and need to find bots and fake account.

2. Tech-stack used

In this project I have used:

- My SQL work bench 8.0 CE
- SQL

The purpose for I used My SQL woke bench software in this project to create and store database and also used for running SQL queries.

I used SQL for retrieve data from the database for a given tasks.

3. Approach

I used My SQL for implementing this project. Firstly I have imported the given datasets into my SQL work bench. Analyzed each table and its attribute also checked the connection with other tables. From each table found the primary key and foreign key for better understanding to use join in SQL. From the above information created entity diagram for the better visibility of the data base

4. SQL Queries

4.1 Database creation

```
CREATE DATABASE  ig_clone;
```

4.2 Table creation

user table:

```
  CREATE TABLE users(

            id INT AUTO_INCREMENT UNIQUE PRIMARY KEY,

            username VARCHAR(255) NOT NULL,

            created_at TIMESTAMP DEFAULT NOW()

);
```

photos table:

```
CREATE TABLE photos(

            id INT AUTO_INCREMENT PRIMARY KEY,

            image_url VARCHAR(355) NOT NULL,

            user_id INT NOT NULL,

            created_dat TIMESTAMP DEFAULT NOW(),

            FOREIGN KEY(user_id) REFERENCES users(id)

);
```

Comments table:

```
CREATE TABLE comments(

            id INT AUTO_INCREMENT PRIMARY KEY,

            comment_text VARCHAR(255) NOT NULL,

            user_id INT NOT NULL,

            photo_id INT NOT NULL,

            created_at TIMESTAMP DEFAULT NOW(),

            FOREIGN KEY(user_id) REFERENCES users(id),

            FOREIGN KEY(photo_id) REFERENCES photos(id)

);
```

Follows:

```
CREATE TABLE follows(

        follower_id INT NOT NULL,

        followee_id INT NOT NULL,

        created_at TIMESTAMP DEFAULT NOW(),

        FOREIGN KEY (follower_id) REFERENCES users(id),

        FOREIGN KEY (followee_id) REFERENCES users(id),

        PRIMARY KEY(follower_id,followee_id)

);
```

Tags:

```
CREATE TABLE tags(

        id INTEGER AUTO_INCREMENT PRIMARY KEY,

        tag_name VARCHAR(255) UNIQUE NOT NULL,

        created_at TIMESTAMP DEFAULT NOW()

);
```

Photos – Tags:

```
CREATE TABLE photo_tags(

        photo_id INT NOT NULL,

        tag_id INT NOT NULL,

        FOREIGN KEY(photo_id) REFERENCES photos(id),

        FOREIGN KEY(tag_id) REFERENCES tags(id),

        PRIMARY KEY(photo_id,tag_id)

);
```

A) Marketing:

1. Rewarding Most Loyal Users:

Your Task: Find the 5 oldest users of the Instagram from the database provided

1.1 SQL queries

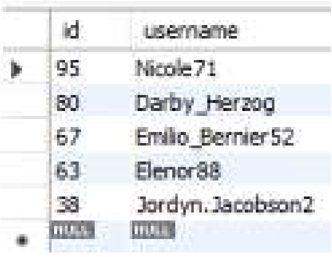
```
/*Find the 5 oldest users of the Instagram from the database provided*/

select users.id, users.username

from users

 order by created_at  asc limit 5;
```

Out put



	id	username
▶	95	Nicole71
	80	Darby_Herzog
	67	Emilio_Bernier52
	63	Elenor88
	38	Jordyn.Jacobson2
*	NULL	NULL

Insights:

- The above given queries gives the 5 oldest user from the datasets.
- The user Nicole71 who have been using the platform for the longest time.

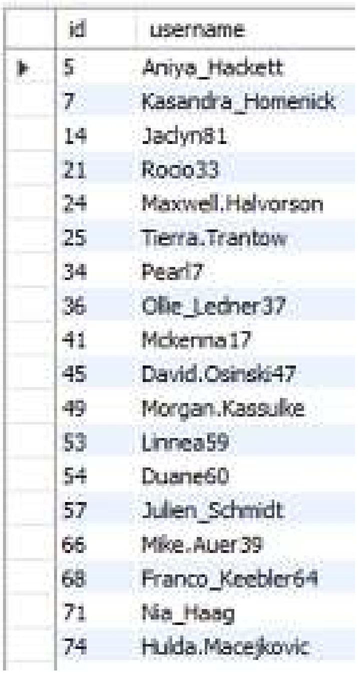
2. Remind Inactive Users to Start Posting:

Your Task: Find the users who have never posted a single photo on Instagram

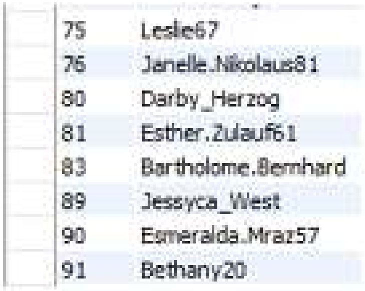
2.1 SQL queries

```
/*Find the users who have never posted a single photo on Instagram*/
 select  users.id,users.username
 from users
 left  join photos
 on users.id = photos.user_id
where photos.user_id is null
/* This are the user who never posted a single photo on instagram*/
```

Out put:



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

Insight:

- Above output shows that there are 26 user who have never post a single photo on instagram
- Sending them a promotional email may encourage them to do their first photo on instagram

3. Declaring Contest Winner:

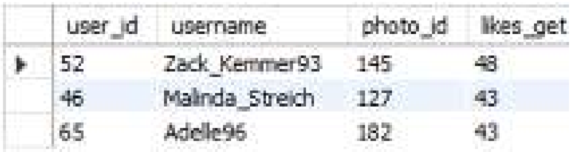
Your Task: Identify the winner of the contest and provide their details to the team

3.1 SQL queries

```
/*identify the winner of the contest and provide their details to the team.
Who gets the most likes on a single photo will win the contest*/

select b.user_id, users.username, b.photo_id, b.likes_get
from (select photos.user_id, maxlikes.photo_id, maxlikes.likes_get  /*sub queries is used*/
              from ( select  likes.photo_id ,count(likes.photo_id) as likes_get  /*sub queries is used*/
                                    from likes
                                    right join users
                                    on likes.user_id = users.id
                                    group by photo_id
                                    order by likes_get desc) as maxlikes
              inner join photos
              on photos.id = maxlikes.photo_id
              where maxlikes.photo_id = photos.id
              order by maxlikes.likes_get desc) as b
inner join users
on users.id = b.user_id
where b.user_id = users.id
order by b.likes_get desc limit 3
/*Zack_kemmer93 user_id 52 who got the most like on a single photo. He is the winner of the contest*/
```

Out put:



user_id	username	photo_id	likes_get
52	Zack_Kemmer93	145	48
46	Malinda_Streich	127	43
65	Adelle96	182	43

Insights:

- User who gets the most like on a single win the contest. Above out put show top three user.
- Zack_kemmer93 user_id 52 who get most like on a single photo. He is the winner of our contest.
- He get 48 likes on a single photo

4. Hash tag Researching:

Your Task: Identify and suggest the top 5 most commonly used hash tags on the platform

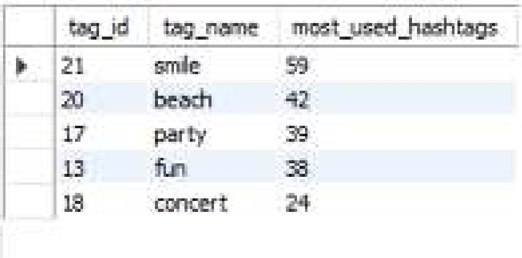
4.1 SQL queries

```
/*Identify and suggest the top 5 most commonly used hashtags on the platform*/

select * from photo_tags;

select photo_tags.tag_id,tags.tag_name, count(photo_tags.tag_id) as most_used_hashtags
from photo_tags
join tags
on photo_tags.tag_id = tags.id
group by photo_tags.tag_id
order by most_used_hashtags desc limit 5
```

Out put



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

Insights:

- Here output shows the top 5 hash tag that are commonly used
- Hash tag ‘smile’ is most used hash tag which is 59, and ‘beach’ is 42, party is 39 times etc.

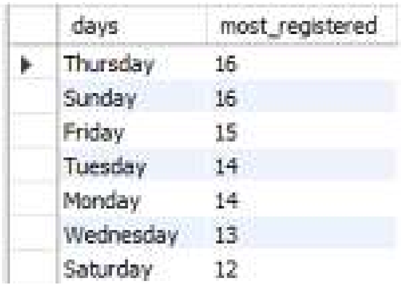
5. Launch AD Campaign:

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

5.1 SQL queries

```
select  dayname(users.created_at)as days,count(users.created_at) as most_registered
from users
group by  dayname(users.created_at)
order by  count(users.created_at) desc ;
```

Out put



days	most_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Insights:

- Thursday and Sunday when most of the user registered that is 16.and Friday 14 people registered, Tuesday and Monday 14 users registered.
- Schedule an ad campaign on Sundays which is a holiday and most of the users registered on this day. So the Sunday is the perfect day for ad campaign

B) Investor Metrics:

1) User Engagement:

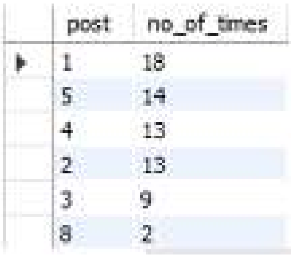
Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

1.1 SQL queries

```
/*Provide how many times does average user posts on Instagram.
 Also, provide the total number of photos on Instagram/total number of users*/

select photos_uploaded.post, count(photos_uploaded.post) as  no_of_times
                    from (     select user_id,users.username, count(image_url ) as post
                                  from photos
                                  join users
                                  on photos.user_id = users.id
                                  group by user_id
                                  order by post desc )as photos_uploaded
      group by post
  order by no_of_times desc ;
```

Out put



post	no_of_times
1	18
5	14
4	13
2	13
3	9
8	2

Insights:

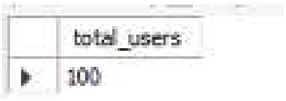
- From the output we can see that most of the users post a single post for 18 times
- They make 5 post 14 times, 4 and 2 posts are post 13 times

1.2 SQL queries

```
/*total number of photos on Instagram/total number of users*/

select  count(users.id) as total_users
from  users;
```

out put



total_users
100

Insights:

- Here we can see that total users are 100

2) Bots & Fake Accounts:

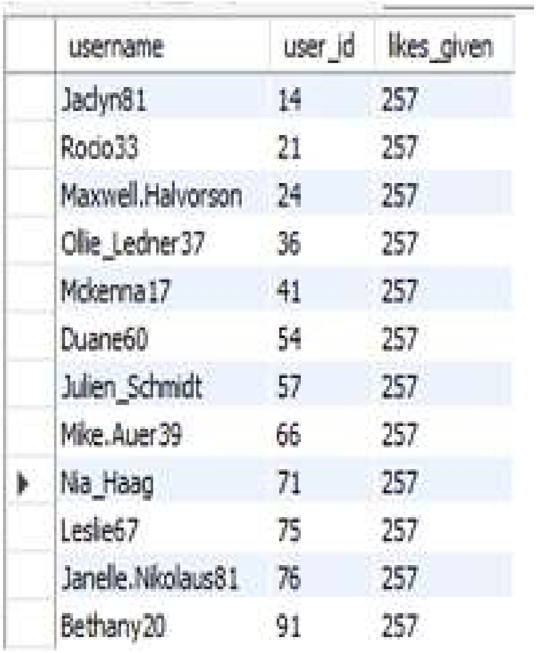
Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

2.1 SQL queries

```
/*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).*/

select username, user_id ,count(user_id) as likes_given
from likes
inner join users
on likes.user_id = users.id
group by user_id
having count(user_id) = 257
order by user_id
/* There are total of 257 photos uploaded by the users, and there are  13 users who likes every single
photos were uploaded.
there may be a chance of some of them were Bots*/
```

Out put



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

Insights:

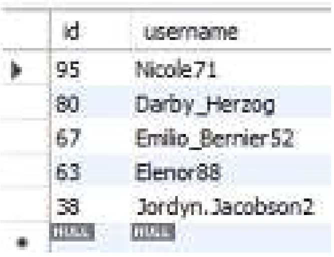
- From the output, that shows username and user id and the likes given by them
- There are total of 257 photos are uploaded by the users, these are the users who make likes every single photo were uploaded.
- There may be a chance of some of them are bots.

5. Result

- (A.1.1 SQL queries) my first task was find the 5 oldest users from the datasets, so I have used order by in created_at column. So it give me oldest user from the table.

```
select users.id, users.username
from users
 order by created_at  asc limit 5;
```

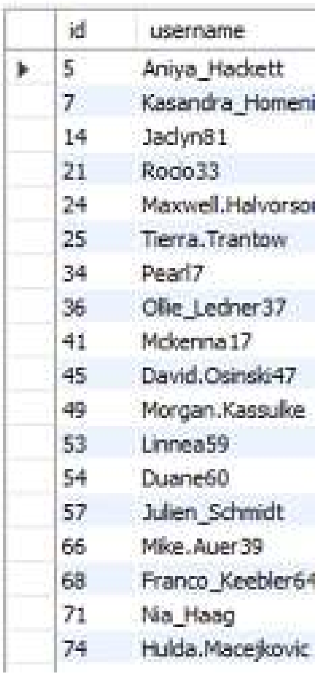
out put



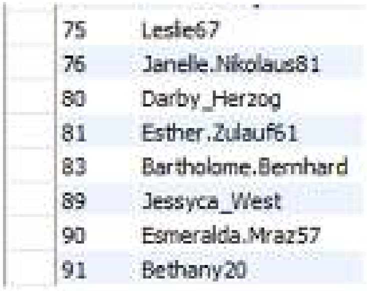
id	username
95	Nicole71
80	Darby_Herzog
67	Emilio_Bernier52
63	Elenor88
38	Jordyn.Jacobson2
NULL	NULL

- (A.2.1 SQL queries) second task was find the users who have never post a single photo. In order to find that I have to take two tables one is user table and another one is photo. For finding those users I need to take the non matching values from the tables. In order to do that I just use left join and where condition. It get me the results.

```
select  users.id,users.username
 from users
 left  join photos
 on users.id = photos.user_id
where photos.user_id is null
```



id	username
5	Aniya_Hackett
7	Kasandra_Homeni
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorso
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_Keebler6
71	Nia_Haag
74	Hulda.Macejkovic

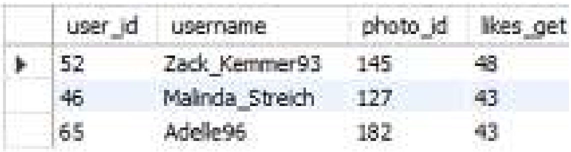


75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

- (A.3.1 SQL queries) next task is find the winner of the contest. Who get most likes for single photo is the winner of the contest. This queries is just a little difficult for me, identify the winner I need to compare three tables so I used sub queries. I used join and where condition group by and order by.

```
select b.user_id, users.username, b.photo_id, b.likes_get
from (select photos.user_id, maxlikes.photo_id, maxlikes.likes_get
                  from ( select  likes.photo_id ,count(likes.photo_id) as likes_get
                                          from likes
                                          right join users
                                          on likes.user_id = users.id
                                          group by photo_id
                                          order by likes_get desc) as maxlikes
                  inner join photos
                  on photos.id = maxlikes.photo_id
                  where maxlikes.photo_id = photos.id
                  order by maxlikes.likes_get desc) as b
inner join users
on users.id = b.user_id
where b.user_id = users.id
order by b.likes_get desc limit 3
```

OUT PUT:

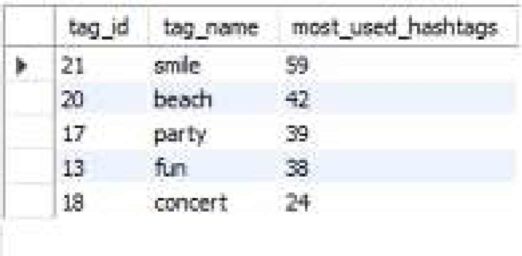


user_id	username	photo_id	likes_get
52	Zack_Kemmer93	145	48
46	Malinda_Streich	127	43
65	Adelle96	182	43

- (A.4.1 SQL queries) 4th task is to find the 5 most commonly used hash tags. I used count() function and order by to find most used hash tag, also I need to join two tables that is photo_tag and tags tables so I used join.

```
select photo_tags.tag_id,tags.tag_name, count(photo_tags.tag_id) as
most_used_hashtags
from photo_tags
join tags
on photo_tags.tag_id = tags.id
group by photo_tags.tag_id
order by most_used_hashtags desc limit 5
```

OUT PUT

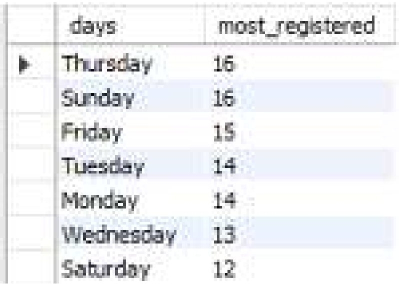


tag_id	tag_name	most_used_hashtags
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

- (A.5.1 SQL queries) 5th task is that what day of the week do most users register on? Provide insights on when to schedule an ad campaign. To do that I used dayname() function to find the day, and I used count() function to find the most registered day

```
select  dayname(users.created_at)as days,count(users.created_at) as
most_registered
from users
group by  dayname(users.created_at)
order by  count(users.created_at) desc ;
```

OUT PUT

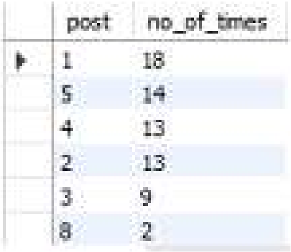


days	most_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

- (B.1.1 SQL queries) 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. To find how many time does average user post on instagram 1st I find posts make by users and then I count how many times. For that purpose I used sub query, order by ,group by, count() function etc.

```
select photos_uploaded.post, count(photos_uploaded.post) as  no_of_times
 from (            select user_id,users.username, count(image_url ) as post
                           from photos
                           join users
                           on photos.user_id = users.id
                           group by user_id
                           order by post desc )as photos_uploaded
              group by post
  order by no_of_times desc ;
```

OUT PUT

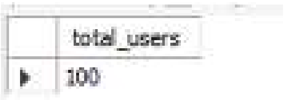


post	no_of_times
1	18
5	14
4	13
2	13
3	9
8	2

- (B.1.2 SQL queries) next tasks is find total number of users. To find that I used count() function.

```
select  count(users.id) as total_users
from  users;
```

OUT PUT

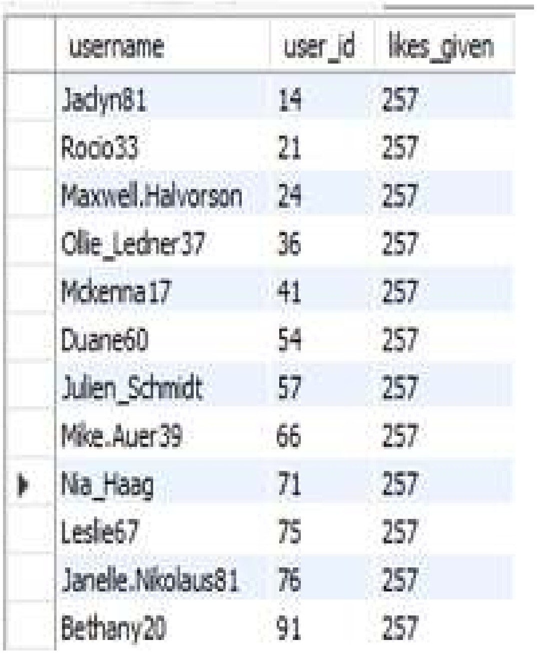


total_users
100

- (B.2.1 SQL queries) my next task is find bots. To do that I have to find who have liked every single photo uploaded. I used count() function to get likes given by the user and I used join to join two tables, also I used group by, order by having condition

```
select username, user_id ,count(user_id) as likes_given
from likes
inner join users
on likes.user_id = users.id
group by user_id
having count(user_id) = 257
order by user_id
```

OUT PUT



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

6. CONCLUSION

From this project instagram user analytics, analyze the user activity on instagram how user engage and interact with the application. And provide the insights to marketing and business development team for improve the user experience and helps the business to grow. The marketing team want to launch a new marketing campaign. There are several tasks given in order to complete I used My SQL work bench 8.0 CE and used several SQL queries to get the answer.

From this project I have learned and improved my existing skill and knowledge. I gained better understanding of My SQL work bench and SQL queries.