Instagram User Analytics

Project Description: This data analysis project is based on Instagram user data and I need to find out inferences from the data to cater to the business and marketing needs.

Approach: Tried to solve the project questions with the knowledge I got from the learning sessions of Trainity.

Tech-Stack Used: I used my SQL Workbench version 8.0.35 along with the SQL server that comes with it as I am doing MySQL training with Trainity.

Insights: I found out mainly the matter of concern from the data that many people are dormant users of Instagram and many are fake profiles. Finding out the fake profiles can help in improving the app reliability and the dormant users can be encouraged to use Instagram by giving them lucrative offers and contests. Even the trending hashtags photos can be found out just by analysing the data.

Result: I learned how to use SQL queries to derive conclusions from the data and Instagram user analytics project has helped me go through what kind of inferences can be derived from user data provided like active users, fake accounts etc.

Drive Link: given while submitting the report.

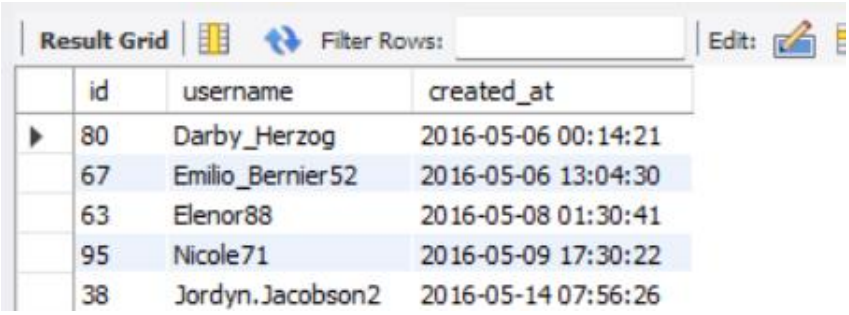
SQL tasks:

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:



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

   SQL Queries:

```
1  show databases;
2  use ig_clone;
3
4  #Loyal users
5  select * from users
6  order by created_at
7  LIMIT 5;
```

Insights:

ID numbers 80, 67, 63,95, and 38 have been on Instagram for the longest time.

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.

   Result:

Result Grid | Filter Rows:

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

SQL Queries:

```
#Users who never posted any photo
select users.id, username from users
left join photos
on users.id=photos.user_id
where photos.id is null;
```

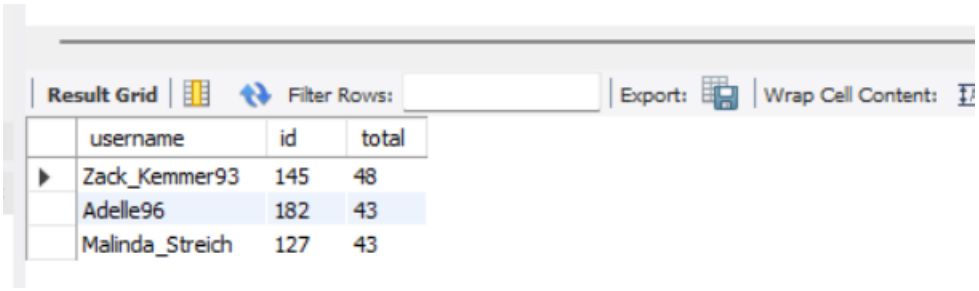
Insights:

User ID numbers and the names of the users who have never posted any photos on Instagram have been shown in the results.

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.

   Result:



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

username	id	total
Zack_Kemmer93	145	48
Adelle96	182	43
Malinda_Streich	127	43

   SQL Queries:

```
15    #contest winner
16 •  select username, photos.id, count(likes.user_id) as total
17    from photos
18    inner join likes
19    on likes.photo_id=photos.id
20    inner join users
21    on photos.user_id=users.id
22    group by photo_id
23    order by total desc
24    limit 3;
```

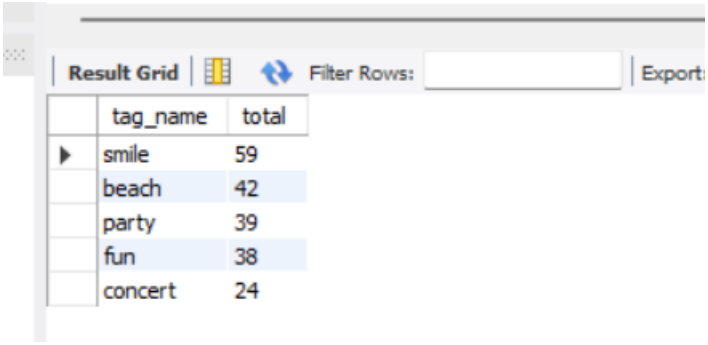
   Insights:

   As shown in the result, the winner of the contest should be Zack_Kemmer93 as he has the highest likes i.e. 48 on a single photo.

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.

   Result:



Result Grid | Filter Rows: | Export:

tag_name	total
smile	59
beach	42
party	39
fun	38
concert	24

SQL Queries:

```
25
26      #Most popular hashtag
27      select tags.tag_name, count(*) as total from photo_tags
28      left join tags
29      on photo_tags.tag_id=tags.id
30      group by tags.id
31      order by total desc
32      limit 5;
33
```

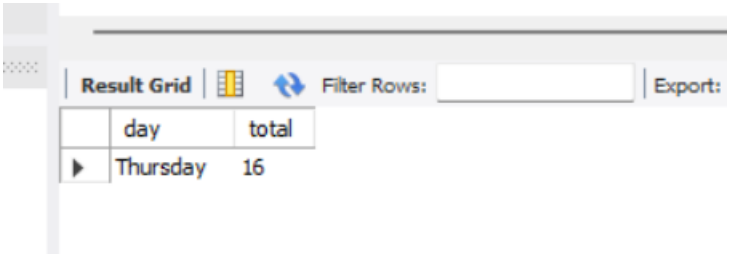
Insights:

As shown in the result, the highest number of tags is for smile, then beach, and so on.

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.

   Result:



Result Grid | Filter Rows: | Export:

day	total
Thursday	16

SQL Queries:

```
34    #Ad campaign launch
35    select dayname(created_at) as day, count(*) as total
36    from users
37    group by day
38    order by total desc
39    limit 1;
40
41
```

Insights:

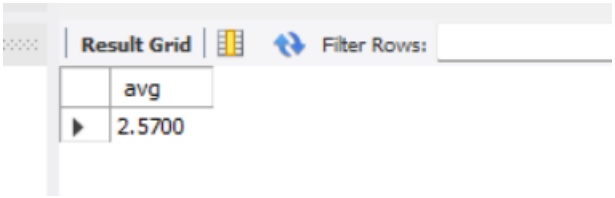
As shown in the result, the best day to launch an ad campaign is on Thursday as on that day most (16) users registered.

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.

Result:



SQL Queries:

```
48
49    #user engagement
50    select (select count(*) from photos)/(select count(*) from users) as avg;
51
```

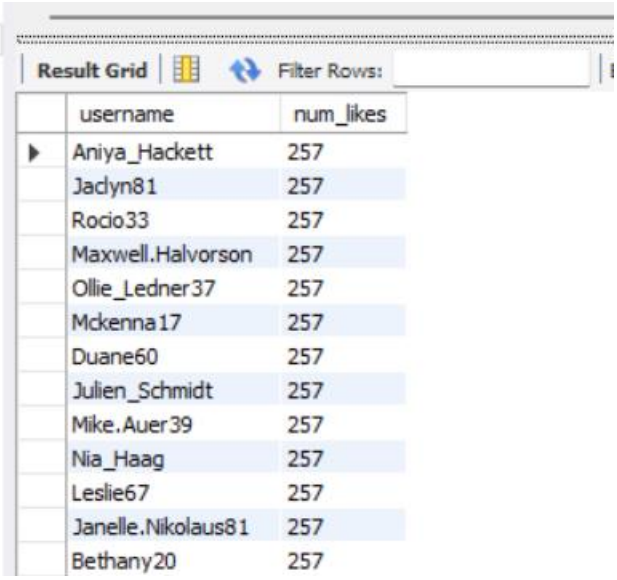
Insights:

As shown in the result, each user posts about 2.57 photos.

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.

Result:



Result Grid | Filter Rows:

username	num_likes
Aniya_Hackett	257
Jaclyn81	257
Rocio33	257
Maxwell.Halvorson	257
Ollie_Ledner37	257
Mckenna17	257
Duane60	257
Julien_Schmidt	257
Mike.Auer39	257
Nia_Haag	257
Leslie67	257
Janelle.Nikolaus81	257
Bethany20	257

SQL Queries:

```
55
56      #fake account
57 •    select username, count(*) as num_likes from users
58      join likes on users.id=likes.user_id
59      group by users.id
60      having num_likes = (select count(*) from photos);
61
```

Insights:

As shown in the result, all the users shown have liked all the photos posted on Instagram, which is not possible for a normal user hence the suspect is that they might be fake accounts.