Task : Instagram User Analytics

Analysis done on the following points:-

Part (A). Marketing :-

1. Rewarding Most Loyal Users
2. Remind Inactive Users to Start Posting
3. Declaring the contest winners
4. Hashtag Researching
5. Launch AD Campaign

Part (B). Investor Metrics :-

1. User Engagement
2. Bots and Fake Accounts

Software used : MySQL Workbench 8.0 CE

Marketing

1.Rewarding the most Loyal users: People who have been using the platform for the longest time.
Task: Find the 5 oldest users of the Instagram from the database provided

To find the 5 oldest users of Instagram:

1. We will use the data from the **users** table by using (*).
2. Then using the **order by** function we will order the desired output by sorting with the **created_at** column in **ascending** order.
3. Then using the **limit** function, the output will be displayed for top 5 oldest Instagram users.

SQL QUERY:

```
Select  * FROM users ORDER BY created_at ASC LIMIT 5;
```

Output/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

These are the 5 oldest user.

2.Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.

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

To Find the inactive users i.e. the users who have never posted a single photo on Instagram:

1. We will first select **username** and **id** column from the users table.
2. Then we will use **NOT IN** operator to exclude any id values that are common in both tables.
3. Then we will use **DISTINCT** to fetch unique user from the photos table.

SQL QUERY:

```
SELECT username, id FROM users WHERE id NOT IN (SELECT DISTINCT user_id FROM photos);
```

Output/Result:

username	id
Aniya_Hackett	5
Kasandra_Homenick	7
Jaclyn81	14
Rocio33	21
Maxwell.Halvorson	24
Tierra.Trantow	25
Pearl7	34
Ollie_Ledner37	36
Mckenna17	41
David.Osinski47	45
Morgan.Kassulke	49
Linnea59	53
Duane60	54
Julien_Schmidt	57
Mike.Auer39	66
Franco_Keebler64	68
Nia_Haag	71
Hulda.Macejkovic	74
Leslie67	75
Janelle.Nikolaus81	76
Darby_Herzog	80
Esther.Zulauf61	81
Bartholome.Bernhard	83
Jessyca_West	89
Esmeralda.Mraz57	90
Bethany20	91

So, there are in total 26 users of the 100 users who have never posted a single photo on Instagram .

3.Declaring Contest Winner: The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
Task: Identify the winner of the contest and provide their details to the team

To find the user who gets the most likes on a single photo :
1. First we will select the **u.id,u.username, p.id AS photo_id, p.image_url**, and **count(*)** as num_like .
2. Then, we will inner join the three tables i.e : users ,photos and likes, **on user id to photo user_id** and then **p.id to l.photo_id**.
3. Then, by using **group by** function we will group the output on the basis of **p.id** and **u.id**.
4. Then, using **order by** function we will sorting the data on the basis of the **num_likes** in **descending** order
5. Then, to find the most liked photo we will using **limit** function to view only the top liked photo's information

SQL QUERY:

```
SELECT u.id, u.username, p.id AS photo_id, p.image_url, COUNT(*) AS num_likes
FROM users u
JOIN photos p ON u.id = p.user_id
JOIN likes l ON p.id = l.photo_id
GROUP BY u.id, p.id
ORDER BY num_likes DESC
LIMIT 1;
```

Output/Result:

id	username	photo_id	image_url	num_likes
52	Zack_Kemmer93	145	https://jarret.name	48

So, the id 52 with username Zack_Kemmer93 is the winner of the contest because his photo with photo_id 145 has the highest number of likes i.e. 48

4.Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform

Task: Identify and suggest the top 5 most commonly used hashtags on the platform.

To find the top 5 most commonly used hashtags on Instagram:

1. We need to select the **tag_name** column from the **tag** table and the **count(*) as count** to count the number of tags used individually.
2. Then, we need to **join tags** table and **photo_tags** table, on **t.id =pt.tag_id** cause they contain the same content in them i.e. **tag_i**d.
3. Then using the **group by** function we need to group the desired output on the basis of **t.tag_name**.
4. Then using the **order by** function we need to sort the output on the basis of **num_like** i.e count in **DESC** order
5. Then, to find the top 5 most used tag names we will use the **limit** 5 function.

SQL QUERY:

```
SELECT t.tag_name, COUNT(*) AS count
FROM tags t
JOIN photo_tags pt ON t.id = pt.tag_id
GROUP BY t.tag_name
ORDER BY count DESC
LIMIT 5;
```

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

These are the 5 most used hashtag ,which will help brand to reach to most of the people.

5.Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.

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

To find the day of week on which most users register on Instagram:

1. First we define the columns of the desired output table using select **dayname(created_at)** as day_of_week and **count(*) as count** i.e total number of users from the users table
2. Then using the **group by** function we group the output table on the basis of day_of_week
3. Then using the **order by** function we order/sort the output table on the basis of **count** in **descending order**

SQL QUERY:

```
SELECT DAYNAME(created_at) AS day_of_week, COUNT(*) AS count
FROM users
GROUP BY day_of_week
ORDER BY count DESC;
```

Output/Result:

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

Most of the users registered on Thursday and Sunday i.e. 16 and hence it would be best to start AD Campaign on these two days

Investor Metrics

1.User Engagement: Are users still as active and post on Instagram or they are making fewer posts

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 the how many times does average posts on Instagram:

1. First, we will use **count** get total number of photos and user from photos and users table respectively .
2. Then we will divide total number photo and user respectively to get number of times does average user posts on Instagram

SQL QUERY:

```
select (select count(*) from photos) as total_photos, (select count(*) from users) as total_users,
(select count(*) from photos)/(select count(*) from users) as avg_posts_per_user;
```

Output/Result:

total_photos	total_users	avg_posts_per_user
257	100	2.5700

So, there are in total 257 photos in the photos table and 100 ids in the users table which makes the desired output to be 257/100 = 2.57

2.Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

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).

To find the bots and fake accounts :

1. First, we select the **u.id** and **u.username** column from the **users** table.
2. Then, we select the **count(*)** function as **count** to count total number of likes from the likes table
3. Then we will join **users** and **likes** table on the basis of **u.id** and **l.user_id**.
4. Then by using the **group by** function we group the desired output table on the basis of **l.user_id**
5. Then, we search for the values from the **cout(*)** from **photos** having equal values with **count** i.e the total number of like from **like** table.

SQL QUERY:

```
SELECT u.id, u.username,count(*) count
FROM users u
JOIN likes l ON u.id = l.user_id
GROUP BY l.user_id
having count =(select count(*) from photos);
```

2. Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

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).

Output/Result:

id	username	count
5	Aniya_Hackett	257
14	Jadyn81	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

So ,these are the id and username fond to be bots as the liked every single photo on the site.

INSIGHTS

- SQL queries can be used to extract specific information from a database, such as finding users who have not posted any photos or calculating the average number of photos per user or on which day most user registered on the site and many more.
- The structure of the database and the relationships between its tables are critical factors in developing effective SQL queries.
- SQL syntax and best practices, such as using subqueries, joins, and aggregate functions, are essential to creating efficient and accurate queries.
- Subqueries can be used in SQL to create more complex queries, such as finding users who have never posted a single photo or who have liked every single photo.
- To optimize performance and accuracy, it is important to understand the structure of the database and use best practices when writing SQL queries.
- The results of SQL queries can provide valuable insights into the characteristics and behavior of users and their interactions with data in the database.
- Careful analysis of the results of SQL queries is necessary to ensure that the insights gained are accurate and meaningful.

RESULT

While making the project , I have achieved several things:

- I have gained a deeper understanding of SQL syntax, best practices, and techniques for working with databases.
- For example, I have developed solutions to common tasks in a database, such as finding find and retrieve data from multiple databases.
- Project help me to demonstrate the power and versatility of SQL queries for analyzing and understanding complex data.

Overall, project have helped me to gain insights into the behavior of users in the database, which can be used to inform decisions and strategies related to the use and management of the database.
In addition , the skills and knowledge gained from this project can be applied to other databases and data analysis tasks, making us more effective and efficient in our work.