Project

Instagram User Analytics

NAME :- NIRAJ INGOLE

INTERNSHIP PROJECT

USING : SQL Fundamentals

Description:

This project of Instagram User Analytics comprises analysis of user data from Instagram. User analysis is a process carried out to derive business insights which in turn helps in development, these insights are used by several teams across the company.

In this project I am working with a product team of Instagram and answer the questions asked by the management team using SQL.

I will be using SQL to derive at solution for the problem statements.

You are required to provide a detailed report answering the questions below :

A) Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

1.**Rewarding Most Loyal Users:** People who have been using the platform for the longest time.

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

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

Your Task: Find the 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.

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

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

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

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

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

B) Investor Metrics: Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

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

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

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

How to do this Project?

Create a Database: You are supposed to run the following commands for creating the database to work on (the database has been provided in the attachments)

Perform Analysis: Use SQL to perform your entire analysis answering the questions asked above

Submit a Report: Make a report (PDF/PPT) to be presented to the leadership team. The report should/can contain the following details:

Approach:

about your project :

This project of Instagram User Analytics comprises analysis of user data from Instagram. User analysis is a process carried out to derive business insights which in turn helps in development, these insights are used by several teams across the company.

In this project I am working with a product team of Instagram and answer the questions asked by the management team using SQL.

I will be using SQL to derive at solution for the problem statements.

• The analysis is done by using basing SQL functions like SELECT, WHERE, COUNT, AS, FROM, GROUP BY, ORDER BY, etc.

A) Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

```
1  show databases;
2  use ig_clone;
3  select * from ig_clone.users;
4
```

1.**Rewarding Most Loyal Users:** People who have been using the platform for the longest time.

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

```
5  /*
6  1.Rewarding Most Loyal Users: People who have been using the platform for the longest time.
7            Your Task: Find the 5 oldest users of the Instagram from the database provided
8  */
9  select * from users,tags, photos,comments,follows,likes,photo_tags;
10 select * from users;
11 select distinct username,created_at  from users order  by  created_at limit 5;
```

Output

	username	created_at
1		
2	Darby_Herzog	2016-05-06 0:14:
3	Emilio_Bernier52	2016-05-06 13:04
4	Elenor88	2016-05-08 1:30:
5	Nicole71	2016-05-09 17:30
6	Jordyn.Jacobson	2016-05-14 7:56:
7		

Procedure:

To find the users using Instagram for the longest time, the created_at table from the dataset is used. The ORDER BY is used to find the oldest user, which sorts the table. This leaves us with the insights of oldest users.

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

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

```
/*
2.Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.
            Your Task: Find the users who have never posted a single photo on Instagram
*/
 select * from users;
select * from photos;
select  username from users
left join photos on
users.id=photos. user_id where  photos.user_id is null
order by users.username;
```

photo on Instagram

Output

	username
1	username
2	Adelle96
3	Aiyana_Hoeger
4	Alek_Watsica
5	Alexandro35
6	Alysa22
7	Andre_Purdy85
8	Aniya_Hackett
9	Aniya_Hackett
10	Annalise.McKenzie16
11	Aracely.Johnston98
12	Arely_Bogan63
13	Aurelie71
14	Bartholome.Bernhard
15	Bartholome.Bernhard
16	Bethany20
17	Bethany20
18	Billy52
19	Cesar93
20	Clint27
21	Colten.Harris76
22	Damon35
23	Darby_Herzog
24	Darby_Herzog
25	Dario77
26	Darwin29
27	David.Osinski47
28	David.Osinski47
29	Delfina_VonRueden68
30	Delpha.Kihn
31	Dereck65

	A
41	Esther.Zulauf61
42	Eveline95
43	Florence99
44	Franco_Keebler64
45	Franco_Keebler64
46	Frederik_Rice
47	Gerard79
48	Granville_Kutch
49	Gus93
50	Hailee26
51	Harley_Lind18
52	Harrison.Beatty50
53	Hulda.Macejkovic
54	Hulda.Macejkovic
55	Imani_Nicolas17
56	Irwin.Larson
57	Jaclyn81
58	Jaclyn81
59	Jaime53
60	Janelle.Nikolaus81
61	Janelle.Nikolaus81
62	Janet.Armstrong
63	Javonte83
64	Jaylan.Lakin
65	Jayson65
66	Jessyca_West
67	Jessyca_West
68	Jordyn.Jacobson2
69	Josianne.Friesen
70	Julien_Schmidt
71	Julien_Schmidt
72	Justina.Gaylord27
73	Kaley9
74	Karley_Bosco
75	Kasandra_Homenick
76	Kasandra_Homenick
77	Katarina.Dibbert
78	Kathryn80
79	Keenan.Schamberger60
80	Kelsi26
81	Kenneth64
82	Kenton_Kirlin
83	Lennie_Hartmann40
84	Leslie67
85	Leslie67
86	Linnea59
87	Linnea59
88	Malinda_Streich

Procedure:

To find inactive users we should find the users who haven't posted a single photo on Instagram, for that we consider users table which contains the user details and the photos table which contains the Instagram post details. By LEFT joining two tables and using WHERE clause the users who haven't posted can be identified.

IS null helps us identify the users in users table but not in photos table. This leaves us with the insights of people who haven't posted and been inactive 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.

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

```
/*
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.
       Your Task: Identify the winner of the contest and provide their details to the team
*/
 select * from users;
select * from photos;
 select * from likes;

select u. username ,
     l.photo_id,
count(l.user_id)as like_user
from
ig_clone.likes l
inner join
ig_clone.photos p
 on
l.user_id= p.id
inner join
 ig_clone.users u
 on p.user_id=u.id group by l.user_id,u.username
 order by like_user desc limit 1
 ;
```

output

	username	photo_id	like_user	
1				
2	Kenton_Kirlin	1	257	
3				

Procedure:

To identify the winner, we should COUNT the likes of all the series. So, we join the INNER JOIN likes and photos table. As we use GROUP BY function on photo_id and username, the query groups all the essentials as per the arguments. We use ORDER BY to find the user who got highest likes and LIMIT function to just output one element from the table. By following all these procedures, we can conclude who is the winner of the contest.

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

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

```
50
51  /*
52  4.Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the
53   most people on the platform.
54         Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform
55   */
56  select  t.tag_name,
57   count(p.photo_id)as num_tags
58   from ig_clone.photo_tags p inner join ig_clone.tags t
59   on p.tag_id = t.id  group by tag_name order by num_tags desc limit 5;
60  /*
```

Output

	tag_name	num_tags		
1	tag_name	num_tags		
2	smile	59		
3	beach	42		
4	party	39		
5	fun	38		
6	concert	24		
7				

Procedure:

To identify the most commonly used hashtags we need to count number of times the hashtag has been used. For that we need data from both photo_tags and tags table. We INNER JOIN both the tables and use COUNT function to count number of times the hashtag has been used. GROUP BY is used to separate and group the data based on the argument tag_name, which leaves us with a list of hashtags and number of times they've been used. We ORDER BY descending order and LIMIT with 5 to output the top 5 commonly used Hashtags. By following this procedure, we can find the commonly used hashtags.

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

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

```
/*
5.Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.
Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign
 */

 select   weekday(created_at) as weekday,
 count(username) as num_users
 from ig_clone.users
 group by  weekday
 order by weekday desc;
```

Output

	weekday	num_users		
2	6	32		
3	5	24		
4	4	30		
5	3	32		
6	2	26		
7	1	28		
8	0	28		

Procedure:

We need to find in which day of the week the users register most on. To find that we need to get data from users table and I'll be using WEEKDAY in MySQL. By using WEEKDAY functions on created_at table which has the data we can get the day on which the user has registered and COUNT to find the number of users registered on that particular day. We use GROUP BY on weekday and get the grouped output and use ORDER BY to sort it in descending order.By following the procedure, we can find the perfect day to launch the AD campaign. But in my views instead of finding the users register date we can find on which does users post more photos, which in turn means more activity. This would give us great result

B) Investor Metrics: Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

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

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

```
74  1.User Engagement: Are users still as active and post on Instagram or they are making fewer posts
75                  Your Task: Provide how many times does average user posts on Instagram.
76                  Also, provide the total number of photos on Instagram/total number of users
77  */
78    select * from users;
79  select * from photos;
80  with created_at as(
81  select  users.id as user_id,
82         count(photos.id) as photo_id
83         from
84         ig_clone.users
85         left join
86         ig_clone.photos
87         on users.id=photos.user_id
88         group by users.id
89  )
90         select sum(photo_id) as total_photos,
91         count(user_id) as total_user,
92         ( photo_id/user_id) as photo_per_users
93         from  created_at;
```

Output

	total_photos	total_user	photo_per_users	
1				
2	514	200	10	
3				
4				

Procedure:

In this section we will be getting multiple insights, the first one is getting the post per user. We will be counting the photo_id to get number of times the user posts and group it to 12 get individual results. We use CTE as temporary result set and from there we find the total

number of photos and users using aggregate functions like SUM. The given problem statements are answered accordingly.

2.Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy 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.Bots & Fake Accounts: The investors want to know if the platform is crowded with
 fake and dummy 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).
*/
with photo_count as(
select
user_id,count(photo_id) as num_like
from
ig_clone.likes
group by
user_id

order by num_like  desc
)
select * from photo_count
where num_like=
(select count(*) from ig_clone.photos);
```

Output

	user_id	num_like
1	21	257
2	71	257
3	5	257
4	66	257
5	41	257
6	14	257
7	57	257
8	24	257
9	76	257
10	75	257
11	54	257
12	91	257
13	36	257

Procedure:

To find the fake account we can use the mentioned method to find the accounts which likes every photo, which is an un usual activity mostly carried out fake bots. For that we use the Likes table from the dataset. We'll be counting all the likes and group it based on the user_id to get output. We order by descending order to find the users who liked all the photos.

By following this procedure, we got a list of accounts which we consider as fake accounts. We can also find user accounts which comments on every post, because it is also a suspicious activity. This may provide us with more efficient results.

Result

By completing this project, I have come to know about the importance of Structured Query Language (SQL) in data analysis. We can use SQL in relational databases and provide with insights which makes our life easier. Instagram user analytics is like a real-life project we get in our work, this project has helped me to understand how the database creation and analysis works.

I have successfully answered and attached the outputs of each and every problem statement I have worked on MySQL Workbench.