Instagram User Analytics

Project Description: In this Project, I will be doing analysis of Instagram data to answer the questions raised by Product team related to Marketing and Investor Metrics. I will be providing insights to the product manager so that they can take crucial decisions.

Approach: For doing analysis, I have been given Instagram database (ig_clone) having 7 SQL tables – users, photos, comments, likes, follows, tags and photo_tags. These tables have data related to different metrics of Instagram. I will be running SQL queries on these tables to find the required information.

Tech-Stack Used: For executing SQL queries, I have used MySQL 8.0 Workbench because I have been using it for my learning since beginning and I find it easy to use and understand.

Data Insights

A. **Marketing:** Some campaigns are launched by Marketing team. Information required by them is as follows: -
   1. **Rewarding 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.
      Explanation: User data is found in 'users' table having 'created_at' column which is used to find oldest users.
      SQL Query: select * from users order by created_at limit 5;

      Results: Oldest users of Instagram -

id	username	created_at
80	Darby_Herzog	06-05-2016 00:14
67	Emilio_Bernier52	06-05-2016 13:04
63	Elenor88	08-05-2016 01:30
95	Nicole71	09-05-2016 17:30
38	Jordyn.Jacobson2	14-05-2016 07:56

   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.

Explanation: Users that post pictures on Instagram are listed in 'photos' table. Therefore, the users which are not there in 'photos' table, have never posted any photo. There are 26 such users as listed in below results.

SQL Query: select * from users where id not in (select distinct user_id from photos);

Results: List of inactive users -

id	username	created_at
5	Aniya_Hackett	07-12-2016 01:04
7	Kasandra_Homenick	12-12-2016 06:50
14	Jaclyn81	06-02-2017 23:29
21	Rocio33	23-01-2017 11:51
24	Maxwell.Halvorson	18-04-2017 02:32
25	Tierra.Trantow	03-10-2016 12:49
34	Pearl7	08-07-2016 21:42
36	Ollie_Ledner37	04-08-2016 15:42
41	Mckenna17	17-07-2016 17:25
45	David.Osinski47	05-02-2017 21:23
49	Morgan.Kassulke	30-10-2016 12:42
53	Linnea59	07-02-2017 07:49
54	Duane60	21-12-2016 04:43
57	Julien_Schmidt	02-02-2017 23:12
66	Mike.Auer39	01-07-2016 17:36
68	Franco_Keebler64	13-11-2016 20:09
71	Nia_Haag	14-05-2016 15:38
74	Hulda.Macejkovic	25-01-2017 17:17
75	Leslie67	21-09-2016 05:14
76	Janelle.Nikolaus81	21-07-2016 09:26
80	Darby_Herzog	06-05-2016 00:14
81	Esther.Zulauf61	14-01-2017 17:02
83	Bartholome.Bernhard	06-11-2016 02:31

89	Jessyca_West	14-09-2016 23:47
90	Esmeralda.Mraz57	03-03-2017 11:52
91	Bethany20	03-06-201631

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.
   Explanation: First, we will find out the most liked photo that can be found in 'likes' table.
   SQL Query: select photo_id, count(photo_id) from likes group by photo_id order by count(photo_id) desc limit 1;

   Results: Most liked Photo -

Photo_id	count(photo_id)
145	38

   Photo with id=145 got most likes. Now, we will find the user who posted this photo from 'photos' table.

   SQL Query: select user_id from photos where id = 145;

   Results: User who posed the most liked photo –

user_id
52

   Therefore, user with id – 52 is the winner of this contest. Let's find his details from 'users' table.

   SQL Query: select * from users where id = 52;

   Results: Contest Winner Details -

id	username	created_at
52	Zack_Kemmer93	01-01-2017 05:58

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 used hashtags on the platform.
   Explanation: Tags are used with pictures. To find most used tags, we need to find the count of pictures tagged by each tag. This information is available in 'photo_tags' table.
   SQL Query: select tag_id, count(distinct photo_id) from photo_tags group by tag_id order by count(distinct photo_id) desc limit 5;

Results:

tag_id	count(distinct photo _id)
21	59
20	42
17	39
13	38
18	24

Now, lets find these tags details in 'tags' table.

SQL Query: select * from tags where id in (21,20,17,13,18);

Results: Top 5 most used hashtags -

id	tag_name	created_at
13	fun	30-01-2023 16:49
17	party	30-01-2023 16:49
18	concert	30-01-2023 16:49
20	beach	30-01-2023 16:49
21	smile	30-01-2023 16:49

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.
Explanation: For this, we need to find the count of users grouped by days of the week based on their registration date. We have 'created_at' column to find registration date in 'users' table. We will use 'dayofweek()' function of MySQL to find day of the week from registration date and group users on its basis to find their count.
SQL Query: select dayofweek(created_at),count(id) from users group by dayofweek(created_at) order by count(id) desc;

Results: Count of users registered on different days of the week (**1** for Sunday, **2** for Monday, **3** for Tuesday, **4** for Wends day, **5** for Thursday, **6** for Friday, **7** for Saturday) –

dayofweek(created_at)	count(id)
5	16
1	16
6	15
3	14
2	14

4	13
7	12

According to these results, Maximum users registered on Thursday and Sunday. Therefore, Ad Campaign can be scheduled on either Thursday or Sunday.

B. **Investor Metrics:** Investors want to know if Instagram is performing well. Information required by them is as follows –

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.
   Explanation: For this we will find count of pictures posted by each user from 'photos' table and then find its average.
   SQL Query: select avg(c) from (select user_id, count(id) c from photos group by user_id)x;

   Results:

avg(c)
3.473

   This shows that on an average, a user posts 3.473 pictures on Instagram. This includes only active users. Inactive users are not counted for this. To include inactive users also, we need to find the total number of photos on Instagram/total number of users from 'users' and 'photos' tables.
   SQL Query: select p/u from (select count(id) u from users)x,(select count(id) p from photos)y;

   Results:

p/u
2.57

   Therefore, If we consider only active users than average posts per user is 3.473 and if we consider all users than average post per user is 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).
   Explanation: We have 257 photos in 'photos' table. So, we can say that bots users have liked 257 pictures. Therefore, we need to find count of pictures liked by each user from 'likes' table and the users who have liked 257 pictures will be bots. Then we can find their details from 'users' table.
   SQL Query: select * from users where id in (select user_id from likes group by user_id having count(photo_id) = 257);
   Results: List of Bots users –

id	username	created_at
14	Jaclyn81	06-02-2017 23:29
21	Rocio33	23-01-2017 11:51
24	Maxwell.Halvorson	18-04-2017 02:32
36	Ollie_Ledner37	04-08-2016 15:42
41	Mckenna17	17-07-2016 17:25
54	Duane60	21-12-2016 04:43
57	Julien_Schmidt	02-02-2017 23:12
66	Mike.Auer39	01-07-2016 17:36
71	Nia_Haag	14-05-2016 15:38
75	Leslie67	21-09-2016 05:14
76	Janelle.Nikolaus81	21-07-2016 09:26
91	Bethany20	03-06-2016 23:31

Result: In this analysis, I got to know how data can be helpful in taking business decisions. Data insights can help in making business strategies. It helps to know user engagement, active/inactive/bot users, app performance etc. By analysing data, dates for new launches can be decided to get maximum benefits from them. Also, this can be helpful in planning ways to enhance activities of existing users. This is the power of data Analysis. This analysis helped me to understand the advantages of data analysis and how it is done to find information.