Instagram User Analytics

(SQL Fundamentals)

Project Description

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams.

I am working with the Product team of Instagram and the Product Manager has asked me to provide Insights on the questions asked by the Management Team.

I am 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.
   My 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.
   My 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.
   My 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.
   My 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.
   My 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
   My 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
   My 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).

Approach

First of all I have imported the dastabase on my MySQL Workbench. Then I analyzed the database carefully. Observing all the tables, columns, rows, and relationship among all the table, and created ER Diagram of complete database provided.

Before finding the answers of the questions I need to have the data understanding of the database provided as well as the business understanding. Then I have done Data Profiling and created a Data Model

like numbers of rows and columns we have in every Table, Datatypes,Keys,Relationships.

After doing all this , I started to find answers of the questions provided to me by the Product Team by Querying the database.

Tech-Stack Used

I have used MySQL Workbench v8.0.31 by Oracle for project execution in order to query the database. The ease of access and setup, troubleshooting support as well as the GUI made it a good tool for the project.

Insights

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.**

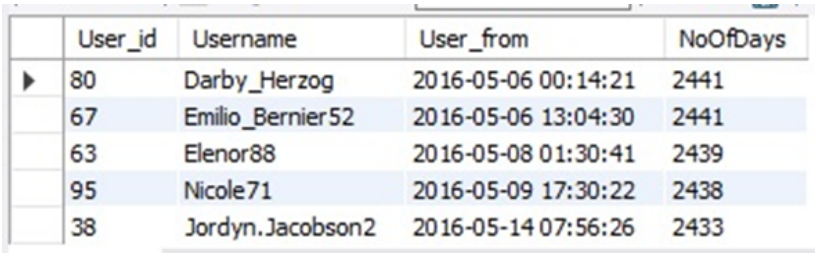
**/*
5 oldest users of the Instagram from the database provided.**

***/**

QUERY:-

```
SELECT id as User_id,
 username as Username,
 created_at as
User_from,(SELECT DATEDIFF(now(),created_at)) as NoOfDays
FROM users
ORDER BY created_at ASC
LIMIT 5;
```

OUTPUT:-



User_id	Username	User_from	NoOfDays
80	Darby_Herzog	2016-05-06 00:14:21	2441
67	Emilio_Bernier52	2016-05-06 13:04:30	2441
63	Elenor88	2016-05-08 01:30:41	2439
95	Nicole71	2016-05-09 17:30:22	2438
38	Jordyn.Jacobson2	2016-05-14 07:56:26	2433

CONCLUSION:- Here, we can see that these are the 5 oldest customers of Instagram. I have also fetched the number of days they have been user of instagram.

So, Marketing Team can Reward these users as these are the 5 most loyal customers.

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

/*Find the users who have never posted a single photo on Instagram*/

QUERY:-

```
SELECT
u.ID,u.username
FROM Users u
LEFT JOIN
photos p
ON
u.id=p.user_id
WHERE p.id IS
NULL;
```

OUTPUT:-

# 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

26 Rows returned.

CONCLUSION:-

There are 26 users who never posted a single photo on Instagram. The userId and username of those users are shown above. Marketing team can remind these 26 inactive users by sending them promotional emails to post their first photo.

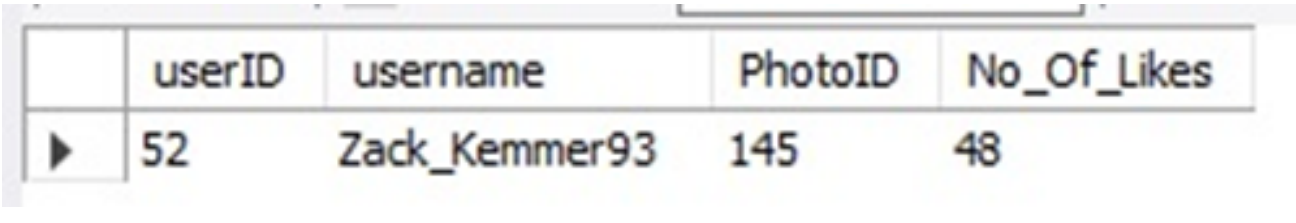
3.**Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.

/* Identify the winner of the contest and provide their details to the team */

QUERY:-

```
WITH
likescount as  /*I have created a Common Table expression consist of photoId and No. of likes
on that photo*/
(
SELECT
p.id,count(l.photo_id) as nooflikes
FROM photos p
INNER JOIN
likes l
ON
p.id=l.photo_id
GROUP BY p.id
)
/*Then I joined that likescount table with users and photos table to get the username of user
who posted that   photo*/
SELECT u.id as
userID,u.username,lc.id as PhotoID,lc.nooflikes as No_Of_Likes
FROM users u
INNER JOIN
photos p
ON
p.user_id=u.id
INNER JOIN
likescount lc
ON
lc.id=p.id
ORDER BY
nooflikes desc
LIMIT 1;
```

OUTPUT:-



	userID	username	PhotoID	No_Of_Likes
▶	52	Zack_Kemmer93	145	48

CONCLUSION:-

We can see here that username Zack_Kemmer93 got maximum likes on his photo with photoID 145.

So, he is the Contest Winner.

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

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

***/**

QUERY:-

```
SELECT
pt.tag_id,t.tag_name,count(pt.photo_id) nooftimesused
FROM
photo_tags pt
INNER JOIN
tags t
ON
pt.tag_id=t.id
GROUP BY
tag_id
ORDER BY
count(photo_id) desc
LIMIT 5;
```

OUTPUT:-

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

CONCLUSION:- We can see here the 5 top most popular used Hashtags on Instagram. I have also shown the number of times these hashtags used.

5.Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs. Provide insights on when to schedule an ad campaign

/*
What day of the week do most users register on?

*/

QUERY:-

```
SELECT (SELECT
dayname(created_at)) as day,count(username) as Users_Registered
FROM users
GROUP BY day
ORDER BY
count(username) desc;
```

OUTPUT:-

	day	Users_Registered
▶	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

CONCLUSION:-

We can see here that most users registered on Thursday and Sunday. So, Thursday or Sunday will be the best day to launch ADs.

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.

/*

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

*/

QUERY:-

```
SELECT
ROUND((SELECT COUNT(*) FROM PHOTOS)/(SELECT COUNT(*) FROM USERS),2) AS AVG;
```

OUTPUT:-

2.57

CONCLUSION:- Every user post average more than 2 posts on Instagram.

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

/*

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).

*/

QUERY:-

```
SELECT
users.id,username, COUNT(users.id) As totalLikesByUser
FROM users
JOIN likes ON
users.id = likes.user_id
GROUP BY
users.id
HAVING
totalLikesByUser = (SELECT COUNT(*) FROM photos);
```

OUTPUT:-

id	username	totalLikesByUser
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

13 Rows Returned.

CONCLUSION:-

These are the users who liked every photo posted on Instagram. So, we can say that these are bots.

Result

Working on this project I have got the real experience of working on corporate live projects. Thanks to the Trainity for giving me this opportunity.

I also able to use my concepts like Joins, Common Table Expression , Group By, Aggregate functions , Operators etc. which I learned and explore those concepts in different ways.

Prepared by: Raushan Kumar Chaurasiya

Data Analyst Trainee

POSITION

12/01/2023

DD/MM/YYYY