Instagram User Analytics

A SQL PROJECT

By Debarati Chatterjee

Content

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.
- These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.
- I have been working with the product team of Instagram and In this project I have provided insights on the questions asked by the management team for the product manager of the company.

Approach

Before to creating the tables for performing different queries, I first went over the provided dataset. In order to get business insights, I also linked the data pieces and organized the tables. I then got the necessary findings and generated important information for the business to make well-considered decisions.

Tech-Stack Used

To complete the project, I have used the software given below

MySQL: Importing and analyzing the data and reviving the insight by performing SQL queries for the report.

MS Word: Presenting the report in detail

Insight

Based on some analytical question, I need to provide the insight of dataset.

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

Question-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

SQL Query:

```
select *
from users
order by created_at asc
limit 5;
```

Output:

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

Findings: The oldest 5 user of Instagram are

i. Darby_Herzog since 06.05.2016
ii. Emilio_Bernier52 since 06.05.2016
iii. Elenor88 since 08.05.2016
iv. Nicole71 since 09.05.2016
v. Jordyn.Jacobson2 since 14.05.2016

Question-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

SQL Query:

```
select users.id, users.username
from users left join photos on users.id=photos.user_id
where photos.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

Finding: The above list of 26 users with ID and Username had not post any single picture on Instagram. We need to send them promotional email individually to post their first photo on Instagram.

Question-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

SQL Query:

```
select users.id, users.username, photos.id, photos.image_url,
       count(likes.photo_id) as most_likes
from users inner join photos on users.id=user_id
           inner join likes on photos.id=likes.photo_id
group by users.id,photos.id
order by most_likes desc
limit 1;
```

Output:

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

Finding: Zack_kemmer93 is the winner of most liked picture with 48 total like on a single picture.

Question-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

SQL Query:

```
Select tags.id, tags.tag_name, count(tag_id) as tag_count
from tags inner join photo_tags on tags.id=photo_tags.tag_id
group by tags.id
order by count(tag_id) desc
limit 5;
```

Output:

id	tag_name	tag_count
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

Finding: Top 5 commonly used hastag on Instagram for high reach for a post is

i. #smile
ii. #beach
iii. #party
iv. #fun
v. #concert

Question-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

SQL Query:

```
select dayname(created_at) as Weekday,
       count(id) as total_user_registered
from users
group by Weekday
order by count(id) desc;
```

Output:

Weekday	total_user_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Finding: Thursday and Sunday are the best possible day to launch 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

Question-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

SQL Query:

Query 1:

```
with Avgpost_table as
(select users.id, users.username, count(photos.id) as total_post
from users inner join photos on users.id=photos.user_id
group by users.id)

select round(avg(total_post),2) as average_user_posts
from Avgpost_table;
```

Query 2:

```
select
round((select count(*) from photos)/(select count(*) from users),2)
as avgpost;
```

Output:

Output 1:

average_user_posts
3.47

Output 2:

avgpost
2.57

Finding: Average post by user is 3.47 which is greater than average post on Instagram which is 2.57. That indicate the active users post 3 to 4 pictures on average.

Question-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).

SQL Query:

```
select users.id, users.username as Botaccount, count(photo_id) as fakelikes
from users inner join likes on users.id=likes.user_id
group by users.id
having fakelikes= (select count(*) from photos);
```

Output:

id	Botaccount	fakelikes
5	Aniya_Hackett	257
14	Jaclyn81	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

Finding: The list of 13 accounts above indicates the possible bot account of user that we should aware of.

Result:

Key result from the project to summarize the detailed analysis:

- There are 5 users who are using the app since May 2016. Darby_Herzog is the oldest one.
- There are 26 users who have not posted even single picture. Promotional email should be sent to them.
- The winner of the most likes contest is Zack_Kemmer93.
- Top 5 hashtags which partner brands can consider are #smile, #beach, #party, #fun and #concert.
- Users registers mostly on Thursdays and Saturdays; hence company can launch Advertisement Campaign on these week days.
- Active users have posted 3 to 4 photos on average on Instagram.
- There are 13 user accounts that can be possibly bot account and action should be taken.

Conclusion

Instagram is wildly used social networking site with growing number of users. To report for user analytics of Instagram I have learned a lot about Business orientation that are applied in critical decision making. My biggest takeaway from the project is that I gained knowledge in data management, table creation, query execution, syntax use, and various error types and fixes while performing SQL queries. In addition, working on this project gave me insight into a variety of variables that are essential for a company to function well and expand.


Thank You