PROJECT REPORT

INSTAGRAM USER ANALYTICS

By

SHIVAM PAL

Contents

1 Project Description

1.1 About Project

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.

In this project, our task is to find the answer/insight of some questions asked by Instagram Product management team, in order to do better decision making.

1.2 How I handle the things?

First thing that we need to do is to ask appropriate question, finding bot user in Instagram database is difficult but finding person who likes all images of Instagram is easy. So, the question asked by product management teams are not actual question for us, we again need to ask an appropriate question.

Secondly, we use relevant tools to fetch required data from Instagram database.

1.3 What are the things that I am going to find out through the project?

1.3.1 Marketing

a) Rewarding Most Loyal Users: People who have been using the platform for the longest time.
b) Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.
c) Declaring Contest Winner: user who gets the most likes on a single photo will win the contest
d) Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
e) Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.

1.3.2 Investor Metrics

a) User Engagement: Are users still as active and post on Instagram or they are making fewer posts.
b) Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

2 Approach

To complete this project the first thing required is to ask appropriate question, then project is nothing more but just getting access to database and retrieving required data using SQL.

Marketing:

Asked Question	Appropriate Question
• **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.	• Find the 5 oldest users of the Instagram from the database provided
• **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.	• Finding the users who have never posted a single photo on Instagram
• **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.	• Finding user who gets most likes on a single photo.
• **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
• **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs.	• What day of the week do most users register on?

Investor Metrics:

Asked Question	Appropriate Question
• **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
• **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).

Marketing:

a) *Rewarding Most Loyal Users* - Find the 5 oldest users of the Instagram from the database provided

MySQL Query:

```
**************************
SELECT
        username,
        created_at
FROM
        users
ORDER BY
        created_at
LIMIT 5;
**************************
```

Output:

username	Cerated_at
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

b) Remind Inactive Users to Start Posting - Finding the users who have never posted a single photo on Instagram

MySQL Query:

```
***********************************
SELECT
        users.id,
        users.username
FROM
        users
LEFT JOIN
        photos
ON
        users.id = photos.user_id
WHERE
        image_url 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

c) Declaring Contest Winner - Finding user who gets most likes on a single photo.

MySQL Query:

```
**********************************
SELECT
        users.id user_id,
        username,
        photo_id,
        count(photo_id) as Total_likes
FROM
        likes
JOIN
        photos
ON
        likes.photo_id = photos.id
JOIN
        users
ON
        photos.user_id = users.id
GROUP BY
        photo_id
ORDER BY
        count(photo_id) DESC
LIMIT 1;
**********************************
```

Output:

user_id	username	Photo_id	Total_likes
52	Zack_Kemmer93	145	48

d) Hashtag Researching - Identify and suggest the top 5 most commonly used hashtags on the platform

MySQL Query:

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

Output:

tag_name	**no_of_times_used**
smile	59
beach	42
party	39
fun	38
concert	24

e) Launch AD Campaign - What day of the week do most users register on?

MySQL Query:

```
**********
SELECT
        dayname(date(created_at)) as Day_Name,
        count(dayname(date(created_at))) as user_registered
FROM
        users
GROUP BY
        Day_Name
ORDER BY
        user_registered DESC;

**********
```

Output:

Day_Name	User_registered
Thursday	16
Sunday	16

Note – Both Thursday and Sunday (Holiday) are days when most user create their Instagram account.

Investor Metrics:

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

MySQL Query:

****

```
SELECT
        FLOOR(count(username)/count(distinct(username))) avg_user_post
FROM
        users
INNER JOIN
        photos
ON
        users.id = photos.user_id;

SELECT
        count(distinct(photos.id))/count(distinct(username)) as total_photos_per_total_user
FROM
        users
LEFT JOIN
        photos
ON
        users.id = photos.user_id;
```

****

Output:

```
avg_user_post = 3
total_photos_per_total_user = 2.5700
```

b) Bots & Fake 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).

MySQL Query:

****

```
SELECT
        username,
        user_id,
        count(user_id) No_of_photos_liked
from
        users
Inner Join
        likes
On
        users.id = likes.user_id
group by
        user_id
having
        count(user_id) = 257; /*As total no of distinct photo on instagram is 257*/
```

****

Output:

username	user_id	No_of_photos_liked
Aniya_Hackett	5	257
Jaclyn81	14	257
Rocio33	21	257
Maxwell.Halvorson	24	257
Ollie_Ledner37	36	257
Mckenna17	41	257
Duane60	54	257
Julien_Schmidt	57	257
Mike.Auer39	66	257
Nia_Haag	71	257
Leslie67	75	257
Janelle.Nikolaus81	76	257
Bethany20	91	257

3 Tech-Stack

Software/Tool Used	Purpose
MS office Professional Plus 2019 (MS Word)	Documentation
MS office Professional Plus 2019 (MS Excel)	Rechecking Answers
MySQL Workbench 8.0 CE	Loading Database and Writing Query

4 Insights

- Instagram launched for people in May 2016.
- Around 26% of users are inactive in Instagram
- The most liked photo on Instagram is posted by Zack_Kemmer93, which is liked by 48% users of Instagram
- The most used hashtag is 'smile', 59% of user on Instagram use this.
- There are very high chances of success of any campaign, if it is launched on Sunday.
- Active user on an average posts 3 photos.
- 13% of Instagram IDs are fake

5 Result

- This project makes us understand how actually data related to any user are stored in databases (for optimization purpose data divided into two category one is static data here data not changes every time when user interact with application, like- username, mail_id, phone_no, address etc. Other type of data is dynamic data, which keep changes every time user interact, like- photos liked, comments, no of connections etc. For both type we need separate table)

- This project also makes us to learn how actually insight are drawn from data, to do that an appropriate question is needed to ask.

6 Drive Link

- https://drive.google.com/drive/folders/1XmYSjXu8YBEhD_e0eeuzeaz--ru30F_V?usp=sharing