Instagram User Analytics

(Project-1)

By Shahequa Modabbera

LinkedIn

GitHub

Project Description:

This project is about how the users engage and interact with Instagram. We will analyse these users 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.

The things that we are going to find out through this project are as follows:

- **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.
- **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.
- **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.
- **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
- **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs.
- **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts.
- **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts.

Approach:

We are working with the product team of Instagram and the product manager has asked us to provide insights on the questions asked by the management team. We use SQL to derive different insights from the dataset provided by the management team. First, we run the necessary commands for creating the database to work on. Then, we performed analysis to generate valuable insights for the company.

Execution:

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

1. **Rewarding Most Loyal Users:** Find the 5 oldest users of the Instagram from the database provided

```
with oldest_users as
(
select username, created_at
from ig_clone.users
order by created_at
limit 5
)
select * from oldest_users;
```

2. **Remind Inactive Users to Start Posting:** Find the users who have never posted a single photo on Instagram

```
select u.username
from ig_clone.users u
left join ig_clone.photos p
on u.id = p.user_id
where p.user_id is null
order by u.username;
```

3. **Declaring Contest Winner:** Identify the winner of the contest and provide their details to the team

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

4. **Hashtag Researching:** Identify and suggest the top 5 most commonly used hashtags on the platform

```
select t.tag_name, count(p.photo_id) as num_tags
from ig_clone.photo_tags p
inner join ig_clone.tags t
on p.tag_id = t.id
```

```
group by tag_name
order by num_tags desc
limit 5;
```

5. **Launch AD Campaign:** What day of the week do most users register on? Provide insights on when to schedule an ad campaign

```
select
dayname(date(created_at)) as Day_Name,
count(dayname(date(created_at))) as user_registered
from ig_clone.users
group by Day_Name
order by user_registered desc;
```

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

6. **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

Average user posts:

```
with CTE as
(
select u.id as userid, count(p.id) as photoid
from ig_clone.users u
left join ig_clone.photos p
on u.id = p.user_id
group by u.id
)
select sum(photoid)/count(userid) as avg_user_post
from CTE
where photoid > 0;
```

Photos per user:

```
with CTE as
(
select u.id as userid, count(p.id) as photoid
from ig_clone.users u
left join ig_clone.photos p
on u.id = p.user_id
group by u.id
)
select
    sum(photoid) as total_photos,
    count(userid) as total_users,
    sum(photoid)/count(userid) as photos_per_user
from CTE;
```

7. **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).

```
select user_id, count(photo_id) as num_likes
from ig_clone.likes
group by user_id
having count(photo_id) = (select count(*) from ig_clone.photos);
```

Tech-Stack Used:

- MySQL Workbench (Version 8.0 CE): MySQL Workbench provides data modelling, SQL development, and various administration tools for configuration. It also offers a graphical interface to work with the databases in a structured way. It is easy to use MySQL to create a database and perform analysis answering the questions given in the description.
- Microsoft Word 2021: It is used to make a report (PDF) to be presented to the leadership team.

Insights:

A) Marketing

- There are total of 100 users using Instagram clone.
- Around 26% of the users are inactive in Instagram. We can remind the inactive users by sending them promotional emails to post their 1st photo.
- The most liked photo in Instagram is posted by Zack_Kemmer93, which is liked by 48% of the users. The team can start the contest for the most liked photos. This will make the users to post more such good posts.
- The most used hashtag is “smile”. Around 59% of the users use the “smile” hashtags. If a partner brand uses the “smile” hashtag, it will be able to reach the most users in the platform.
- The best days to launch ads are Sunday and Thursday. As the most users register on Instagram on Sunday and Thursday.

B) Investor Metrics

- Average users post on Instagram = 3.5 (approx.)
- Photos per user = Total photos / Total users = 2.57
- This will help the investors in knowing whether the users are active and post on Instagram or not.
- 13% of Instagram IDs are fake and dummy accounts.

Results:

From this project, I get to know about the business perspectives of a company. I learned how the marketing team with the help of data analyst team can grow the business altogether. This

project helps in understanding how actually the data gained from the users are stored in databases and we can generate the beautiful insights from it. I learned how to ask the right questions given the circumstances. I also get to know how to use the hashtags to reach the most audience while posting and what time to post (after doing research about the Instagram Analytics). It helped me in building my SQL concepts.

SQL Query Link:

Instagram User Analytics SQL Code