# Instagram User Analytics

By Stan Pereira

# Project Description

As a data analyst working with the product team at Instagram, I was tasked with analysing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

The user analysis involves tracking how users engage with a digital product, and the insights derived from this analysis will be used by various teams within the business.

The marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

# Approach

The initial step is the creation of the database & tables, along with importing the data into the tables. This is achieved by running the DDL & DML SQL queries provided in MySQL through the MySQL Workbench.

After the previous step, analysis on the data was made and insights were generated from the database by running SQL queries in MySQL Workbench.

# Tech Stack Used

The Software and Version Utilized

MySQL has a feature rich RDBMS that can be used to perform various data manipulation and analysis tasks such as data aggregation, table joining, data transformation, data visualization, etc.

It is also fast and scalable, allowing you can handle large amounts of data and perform complex queries efficiently.

MySQL Workbench 8.0 CE - Version 8.0.34 build 3263449 CE (64 bits) Community

MySQL is free and open source

MySQL supports multiple languages and platforms and different operating systems. You can also use various connectors and drivers to access MySQL.

# Management Queries

## Marketing Analysis

- Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
- Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
- Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.
- Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
- Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

## Investor Metrics

- User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
- Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

# Marketing Analysis #1

Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

**Task:** Identify the five oldest users on Instagram

**SQL Query:**

```
SELECT
    username as Username,
    created_at as Creation_Date
FROM
    users
ORDER BY Creation_Date
LIMIT 5;
```

| Username | Creation_Date |
|---|---|
| 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 |

**Insight:** The marketing team should reward those who have been using the platform for the longest time, including the user 'Darby_Herzog' who has not posted anything. It may entice him to post something.

# Marketing Analysis #2

Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

**Task:** Identify users who have never posted a single photo on Instagram.

**SQL Query:**

```
SELECT username as Inactive_Users
FROM
   users
      LEFT JOIN
   photos ON users.id = photos.user_id
WHERE photos.user_id IS NULL
ORDER BY Inactive_Users;
```

**Insight:** The marketing team should encourage the 26 users to start posting by sending them promotional emails. Follow-ups can also be made, so as to remove users who are still inactive.

| Inactive_Users |
|---|
| Aniya_Hackett |
| Bartholome.Bernhard |
| Bethany20 |
| Darby_Herzog |
| David.Osinski47 |
| Duane60 |
| Esmeralda.Mraz57 |
| Esther.Zulauf61 |
| Franco_Keebler64 |
| Hulda.Macejkovic |
| Jadyn81 |
| Janelle.Nikolaus81 |
| Jessyca_West |
| Julien_Schmidt |
| Kasandra_Homenick |
| Leslie67 |
| Linnea59 |
| Maxwell.Halvorson |
| Mckenna17 |
| Mike.Auer39 |
| Morgan.Kassulke |
| Nia_Haag |
| Ollie_Ledner37 |
| Pearl7 |
| Rocio33 |
| Tierra.Trantow |

# Marketing Analysis #3

Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

**Task:** Determine the winner of the contest and provide their details to the team.

**SQL Query:**

```
SELECT
    users.id AS ID, users.username AS
Username,
    photos.image_url AS Photo,
    COUNT(likes.photo_id) AS Total_Likes
FROM
    users
        INNER JOIN
    photos ON users.id = photos.user_id
        INNER JOIN
    likes ON photos.id = likes.photo_id
GROUP BY likes.photo_id
ORDER BY Total_Likes DESC
LIMIT 1;
```

| | ID | Username | Photo | Total_Likes |
|---|---|---|---|---|
| ▶ | 52 | Zack_Kemmer93 | https://jarret.name | 48 |

**Insight:** The user 'Zack_Kemmer93' has got the most likes (48) for one of his photos. All users should be informed of the contest winner and his photo. A small reward needs to be given to winner, which might entice other users to post more.

# Marketing Analysis #4

Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

**Task:** Identify and suggest the top five most commonly used hashtags on the platform.

**SQL Query:**

```
WITH HTR as (
   SELECT
      tag_id, tag_name,
      COUNT(tag_id) as Total_Count,
      RANK() OVER (ORDER BY COUNT(tag_id)
DESC) as Tag_Rank
   FROM
      photo_tags
         INNER JOIN
      tags ON photo_tags.tag_id=tags.id
   GROUP BY tag_id
)
SELECT *
FROM
   HTR
WHERE
   Tag_Rank <= 5;
```

| tag_id | tag_name | Total_Count | Tag_Rank |
|---|---|---|---|
| 21 | smile | 59 | 1 |
| 20 | beach | 42 | 2 |
| 17 | party | 39 | 3 |
| 13 | fun | 38 | 4 |
| 5 | food | 24 | 5 |
| 11 | lol | 24 | 5 |
| 18 | concert | 24 | 5 |

**Insight:** Even though the top five hashtags were asked for, the hashtag that was ranked 5th was tied with two others, hence a total of 7 hashtags are mentioned.

# Marketing Analysis #5

Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

**Task:** Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

**SQL Query:**

```
WITH DayTable as (
   SELECT
      DAYNAME(created_at) as Day_Of_Creation,
      COUNT(DAYNAME(created_at))
as Total_Reg_Users,
      RANK() OVER (order
by count(dayname(created_at)) DESC)
as DayRank
   FROM
      users
   GROUP BY Day_Of_Creation
)
SELECT *
FROM
   DayTable
WHERE
   DayRank=1;
```

| | Day_Of_Creation | Total_Reg_Users | DayRank |
|---|---|---|---|
| ▶ | Thursday | 16 | 1 |
| | Sunday | 16 | 1 |

**Insight:** There are two days of the week that have the most registrations. As the goal is to increase registrations, the marketing team may want to launch ads on both the days of most registrations, as this indicates that there is a high demand and interest.

# Investor Metrics #1

User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

**Task:** Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

**SQL Query:**

```
SELECT
   (SELECT
      COUNT(DISTINCT id) /
COUNT(DISTINCT user_id)
   FROM
      photos) AS Total_Posts_Per_Active_User,
   (SELECT
      COUNT(DISTINCT id)
   FROM
      photos) /
   (SELECT
      COUNT(DISTINCT id)
   FROM
      users)
AS Ratio_Total_Photos_to_Total_Users;
```

| Total_Posts_Per_Active_User | Ratio_Total_Photos_to_Total_Users |
|---|---|
| 3.4730 | 2.5700 |

**Insight:** The Average number of posts is taken from Active Users, but the ratio is taken from All Users. The information shows the gap between the users that are registered and actual users making posts. This gap needs to decrease.

# Investor Metrics #2

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

**Task:** Identify users (potential bots) who have liked every single photo on the site.

**SQL Query:**

```
SELECT
    user_id AS ID, username AS Username,
    COUNT(photo_id) AS TotalCount
FROM
    users
        JOIN
    likes ON users.id = likes.user_id
GROUP BY user_id
HAVING TotalCount = (SELECT
        COUNT(id)
    FROM
        photos)
ORDER BY username;
```

| ID | Username | TotalCount |
|---|---|---|
| 5 | Aniya_Hackett | 257 |
| 91 | Bethany20 | 257 |
| 54 | Duane60 | 257 |
| 14 | Jadyn81 | 257 |
| 76 | Janelle.Nikolaus81 | 257 |
| 57 | Julien_Schmidt | 257 |
| 75 | Leslie67 | 257 |
| 24 | Maxwell.Halvorson | 257 |
| 41 | Mckenna17 | 257 |
| 66 | Mike.Auer39 | 257 |
| 71 | Nia_Haag | 257 |
| 36 | Ollie_Ledner37 | 257 |
| 21 | Rocio33 | 257 |

**Insight:** As genuine users do not normally like every single photo, the 13 users in the list are Bots or Fake Accounts. This list distorts the data and can provide incorrect analysis. The users need to be removed from the database.

# Project Impact

| | |
|---|---|
| Impact for the Team | • The analysis from the data will help the product manager and the team make informed decisions about the future direction of the Instagram App. |
| Impact for Me | • The project allowed me to learn the fundamentals of SQL and its working. It has also shown me how to use it to extract various insights for data analysis. |

# Thank You