Instagram User Analytics

Project Description

This project aim to carry out the in depth analysis of user engagement process with the instagram platform which will help the product team to launch better features for the platform.

This project focuses mainly on two key aspects Marketing and Investors Metrics.
Based on the user engagement and the data collected the insights needs to be carried out and presented to the product team. The project will answer to the important questions like:

>Rewarding Loyal Users.
>Remind Inactive Users to Start Posting.
>Declaring Contest Winner.
>Hashtag Researching.
>Launch AD Campaign.
>User Engagement.
>Bots and Fake Accounts.

Approach

For this project, I have used MySQL Workbench to extract the required data from the given database using the Join function, subqueries, Aggregation, where condition, Group by, Distinct and other functions required. keeping the Primary key and foreign key in consideration provided all the reports asked by the marketing department and Investor metrics department.

Queries were utilized to create a database from the provided raw data. Sorting and data extracting queries were then implemented to obtain the required data/insights.

Tech Stack Used

MySQL Workbench :- The tech stack used included MySQL Workbench v8.0.39, which was an excellent tool for querying the database, thanks to its ease of access, simple setup, and GUI, as well as its troubleshooting support.

Insights

- This Instagram user analytics project helped me to dig deeper into the world of SQL and helped me understand how complex queries work and how to cultivate business insights from given data .It enabled me to ask the right necessary questions and narrow down solutions to the given problems.

- During the Instagram data analytics project, I focused on analyzing user engagement, content performance, hashtag usage, content type performance. These insights helped in refining content strategies, optimizing posting schedules, and improving overall engagement on the platform.

- The project answers the questions and thus provides the required insights relevant for the product team. This project has helped me to understand how data can be of relevance to a business and has helped to hone my SQL skills.

Results

Here are the query statements which I executed and the corresponding results.

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

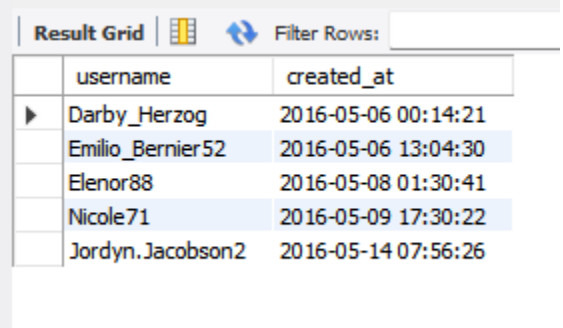
1. **Loyal User Reward:** The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
   Task: Identify the five oldest users on Instagram from the provided database.

   Code:-

```
#Find the 5 oldest users of the Instagram from the database provided.

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

Result:-



Result Grid | Filter Rows:

username	created_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

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.

Code:-

```
#Identify users who have never posted a single photo on Instagram.
SELECT username
FROM users
LEFT JOIN photos
ON users.id=photos.user_id
WHERE photos.id IS NULL;
```

Result:-

username
Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

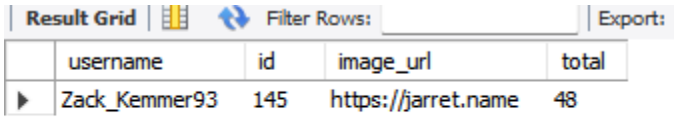
3. **Contest Winner Declaration:** The team has organized a contest where the user with the most likes on a single photo wins.
   Task: Determine the winner of the contest and provide their details to the team.

   Code:-

```sql
#Determine the winner of the contest and provide their details to the team.

SELECT
username, photos.id, photos.image_url, count(likes.user_id) AS total
FROM photos
INNER JOIN  likes
ON likes.photo_id=photos.id
INNER JOIN users
ON photos.user_id=users.id
GROUP BY photos.id
ORDER BY total DESC
LIMIT 1;
```

Result:-



Result Grid | Filter Rows: | Export:

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

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.

Code:-

```sql
#Identify and suggest the top five most commonly used hashtags on the platform.

SELECT
tags.tag_name,
COUNT(*) AS total
FROM photo_tags
JOIN tags
ON photo_tags.tag_id = tags.id
GROUP BY tags.id
ORDER BY total DESC
LIMIT 5;
```

Result:-

Result Grid | Filter Rows:

tag_name	total
smile	59
beach	42
party	39
fun	38
concert	24

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.

   Code:-

```sql
#Provide insights on when to schedule an ad campaign.
SELECT
    DAYNAME(created_at) AS day_of_week,
    COUNT(id) AS User_Registered
FROM
    users
GROUP BY day_of_week
ORDER BY User_Registered DESC
LIMIT 1;
```

   Result:-

Result Grid | Filter Rows:

day_of_week	User_Registered
Thursday	16

B) 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.

Code:-

```
#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.

WITH cte AS
(
SELECT
    u.id AS userid, COUNT(p.id) AS photoid
FROM
    users u
        LEFT JOIN
    photos p ON u.id = p.user_id
GROUP BY u.id
)
SELECT
    SUM(photoid) AS total_photo, COUNT(userid) AS total_users
FROM
    cte;
```

```
SELECT
    AVG(Number_of_Post) AS average_post_per_user
FROM
    (SELECT
        user_id, COUNT(*) AS Number_of_Post
    FROM
        photos
    GROUP BY user_id) AS average_post_count;
```

Result:-

Result Grid | Filter Rows:

	average_post_per_user
▶	3.4730

Based on the results, there are-

- 74 active users who have posted at least once.
- 100 total users(as per the data)

- 257 total posts made.
- Total photos / Total users = 257/100 = 2.57

So the average will be 257/74 = 3.47, based on the data we can say that an average user posts 3-4 times.

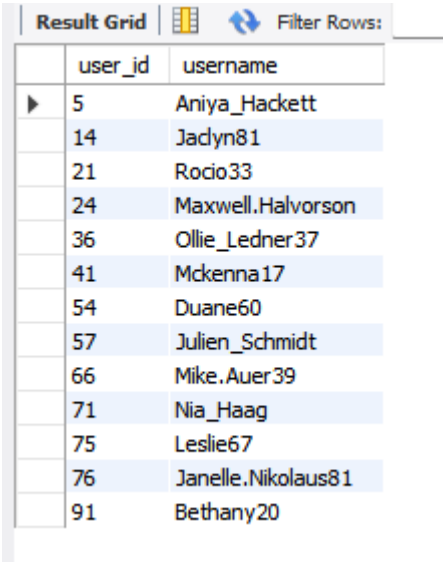
2. **Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.
   Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

   Code:-

```sql
#Identify users (potential bots) who have liked every single photo on the site.

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

   Result:-



Result Grid | Filter Rows:

user_id	username
5	Aniya_Hackett
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson
36	Ollie_Ledner37
41	Mckenna17
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
71	Nia_Haag
75	Leslie67
76	Janelle.Nikolaus81
91	Bethany20

We have 13 such users based on the data who have liked all 257 posts, user id for the same are specified above.

Conclusion

From this project, I got an idea about how as a business or data analyst we work on real time data to take any data driven decision.

By identifying optimal posting times, high-performing content types, and effective hashtag strategies, we were able to tailor our approach to maximize reach and interaction.

Additionally, understanding the demographics of the most engaged audience allowed for more targeted content creation. These findings will serve as a foundation for ongoing strategy refinement, ultimately enhancing overall user engagement and follower growth on the platform.

It helped me a lot to understand the analysis process well, and to provide insights for the best decision possible.