Instagram User Data Analysis Using SQL

Name : Aditya .C. Mohite
Date : 07/1/25

Objective

The objective of this project is to analyze Instagram user data to derive meaningful insights that can support the product, marketing, and development teams in making informed decisions. By understanding user interactions and engagement, the project aims to identify opportunities to improve user experience, increase engagement, and grow the platform effectively.

The project involved using SQL in MySQL Workbench to query and analyze a database containing Instagram user data. Tasks were categorized into marketing analysis and investor metrics. SQL queries were crafted to answer specific business questions, and the results were analyzed to generate actionable insights.

Approach

- **Step 1: Creating and Setting Up the Database**
  Imported the provided database file into MySQL Workbench.
  Verified the database structure, including tables, columns, and data types, to ensure accuracy.

- **Step 2: Writing SQL Queries for Each Task**
  Crafted efficient and accurate SQL queries to address specific business questions.
  Queries were designed to retrieve data for marketing analysis and investor metrics.

- **Step 3: Analyzing Query Outputs**
  Reviewed and validated the outputs of the SQL queries to ensure correctness.
  Interpreted the data to extract meaningful insights relevant to the posed questions.

- **Step 4: Summarizing Findings and Insights**
  Consolidated the results into clear and actionable insights.
  Focused on presenting findings that support decision-making for product, marketing, and development strategies.

Tech-stack used

Database:
MySQL Workbench: User-friendly and efficient for managing databases and running SQL queries.

Programming:
SQL: Essential for extracting, analyzing, and interpreting data.

Other Tools:
PowerPoint/Word: Used to create a clear and professional report/presentation.

Why These Tools?
Selected for their ease of use, compatibility, and ability to efficiently perform data analysis and reporting.

Marketing Analysis

Task : Identify the five oldest users on Instagram

SQL Query:

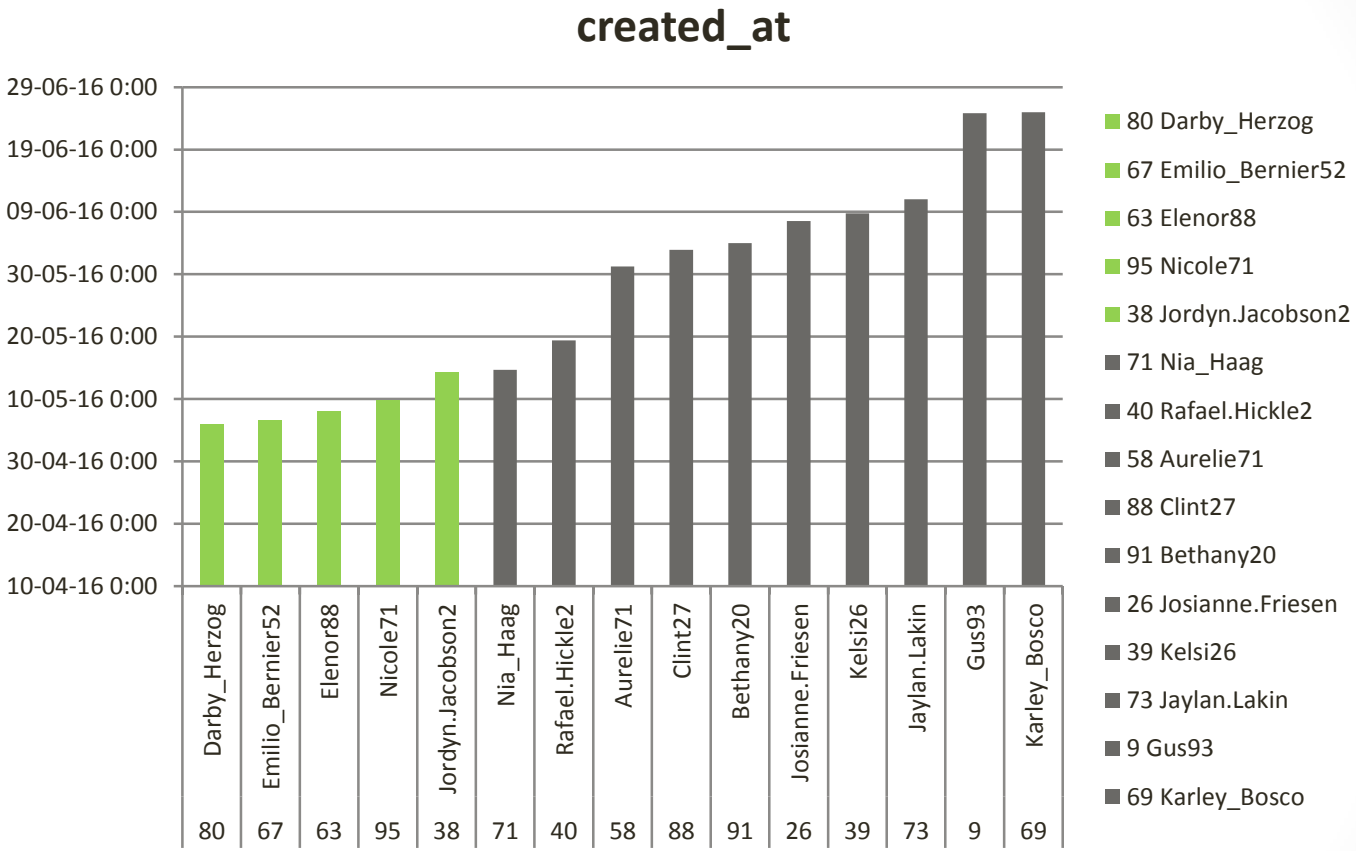
```
1  USE ig_clone;
2  SELECT * FROM users
3  order by created_at asc
4  LIMIT 5;
```

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

Insights:

- The five oldest users have been active since Instagram's early days, showing long-term loyalty.
- These users could be prioritized for special rewards or recognition programs.


created_at
29-06-16 0:00
19-06-16 0:00
09-06-16 0:00
30-05-16 0:00
20-05-16 0:00
10-05-16 0:00
30-04-16 0:00
20-04-16 0:00
10-04-16 0:00
Darby_Herzog
Emilio_Bernier52
Elenor88
Nicole71
Jordyn.Jacobson2
Nia_Haag
Rafael.Hickle2
Aurelie71
Clint27
Bethany20
Josianne.Friesen
Kelsi26
Jaylan.Lakin
Gus93
Karley_Bosco
80
67
63
95
38
71
40
58
88
91
26
39
73
9
69
80 Darby_Herzog
67 Emilio_Bernier52
63 Elenor88
95 Nicole71
38 Jordyn.Jacobson2
71 Nia_Haag
40 Rafael.Hickle2
58 Aurelie71
88 Clint27
91 Bethany20
26 Josianne.Friesen
39 Kelsi26
73 Jaylan.Lakin
9 Gus93
69 Karley_Bosco

Task : Identify users who have never posted a single photo

SQL Query:

```
1
2  USE ig_clone;
3  SELECT users.id,users.username FROM users
4  LEFT JOIN photos ON users.id=photos.user_id
5  WHERE photos.user_id is NULL;
```

Output:

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jadyn81
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

Insights:

- A significant number of users have never posted..
- Marketing campaigns targeting these users could boost engagement.

Task : Determine the User with the Most Likes on a Single Photo.

SQL Query:

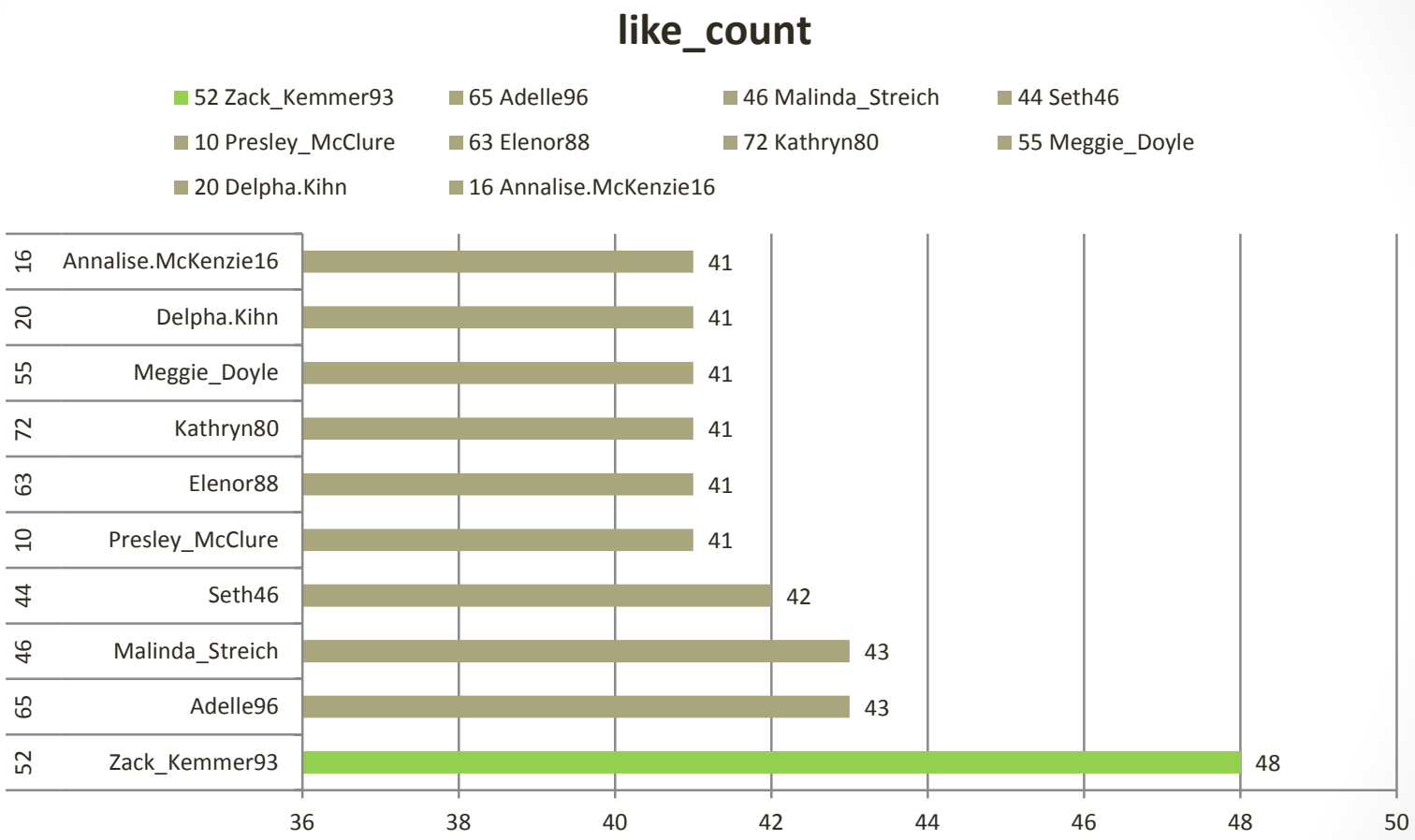
```sql
USE ig_clone;
SELECT photos.user_id,users.username,likes.photo_id,photos.image_url,COUNT(likes.user_id) AS like_count FROM likes
JOIN photos ON likes.photo_id = photos.id
JOIN users ON photos.user_id = users.id
GROUP BY likes.photo_id
ORDER BY like_count DESC
LIMIT 1;
```

Output:

user_id	username	photo_id	image_url	like_count
52	Zack_Kemmer93	145	https://jarret.name	48

Insights:

- The contest winner is identified by the photo with the highest likes.
- Highlighting this user can boost community participation in future contests.


like_count
52 Zack_Kemmer93
65 Adelle96
46 Malinda_Streich
44 Seth46
10 Presley_McClure
63 Elenor88
72 Kathryn80
55 Meggie_Doyle
20 Delpha.Kihn
16 Annalise.McKenzie16
16 Annalise.McKenzie16
41
20 Delpha.Kihn
41
55 Meggie_Doyle
41
72 Kathryn80
41
63 Elenor88
41
10 Presley_McClure
41
44 Seth46
42
46 Malinda_Streich
43
65 Adelle96
43
52 Zack_Kemmer93
48
36
38
40
42
44
46
48
50

Task : Identify the Top Five Most Commonly Used Hashtags.

SQL Query:

```sql
use ig_clone;
SELECT photo_tags.tag_id,tags.tag_name,COUNT(photo_tags.photo_id) as tag_use_count
FROM photo_tags
JOIN tags on photo_tags.tag_id = tags.id
GROUP BY photo_tags.tag_id
ORDER BY tag_use_count DESC
LIMIT 5;
```

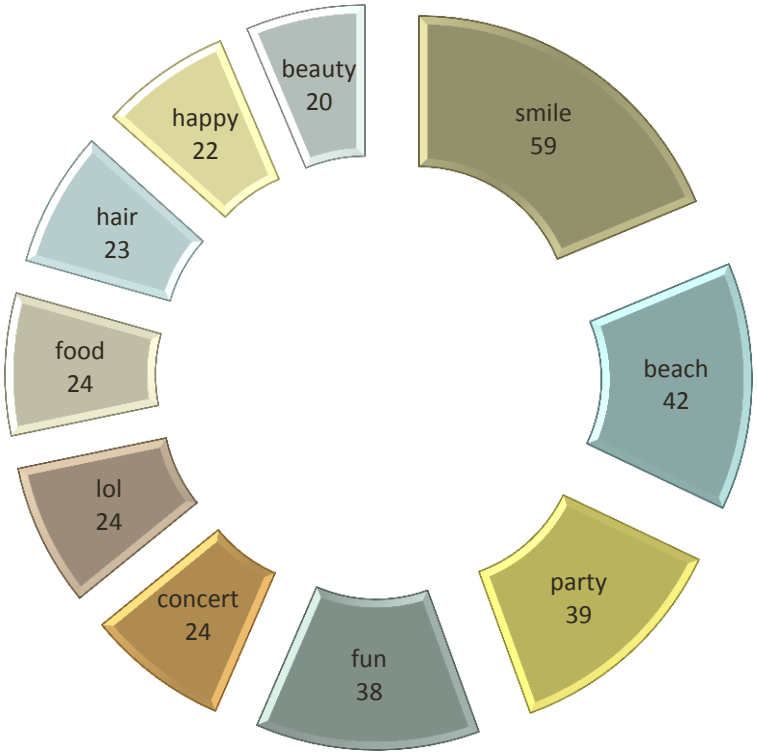
Output:

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

Insights:

- The most popular hashtags can help partner brands increase visibility.
- These hashtags indicate trending topics among users.

tags_use_count



Task : Identify the Top Five Most Commonly Used Hashtags.

SQL Query:

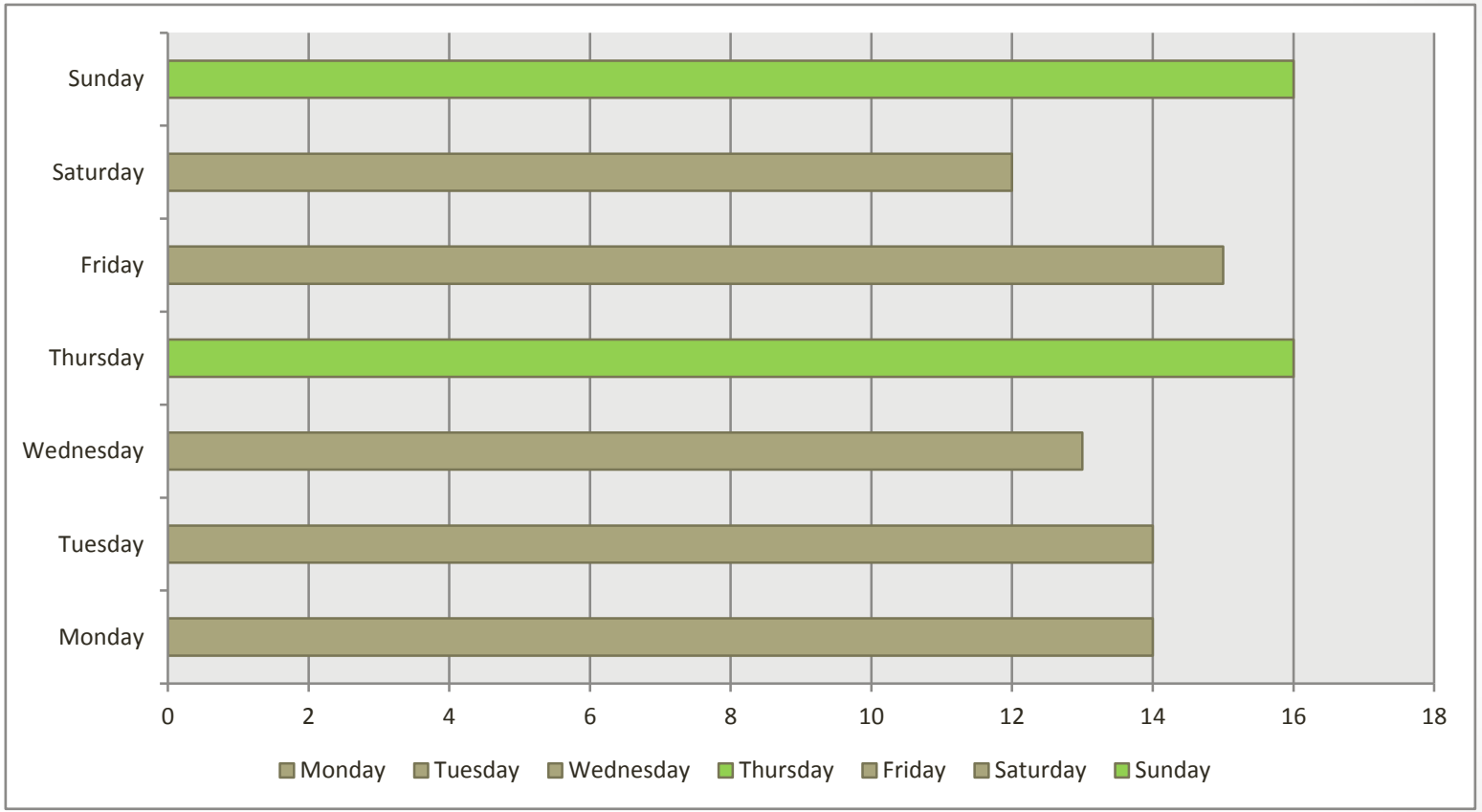
```sql
USE ig_clone;
SELECT dayname(created_at) as day_of_week,COUNT(id) as users_registered
FROM users
GROUP by day_of_week
ORDER BY users_registered DESC;
```

Output:

day_of_week	users_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Insights:

- The highest user registrations occur on Thursdays & Sundays.
- Scheduling ads on this day can maximize reach and engagement.


Sunday
Saturday
Friday
Thursday
Wednesday
Tuesday
Monday
0
2
4
6
8
10
12
14
16
18
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday

Investor Metrics

Task : Calculate the Average Posts per User and Total Photos Divided by Total Users.

SQL Query:

```sql
USE ig_clone;
SELECT CEIL(AVG(posts_per_user))
FROM (SELECT COUNT(id) as posts_per_user FROM photos GROUP BY user_id) as avg_posts;
```

```sql
USE ig_clone;
SELECT total_posts/total_users
FROM (SELECT COUNT(id) as total_posts from photos) as sq1,
(SELECT COUNT(id) as total_users from users) as sq2;
```

Output:

CEIL(AVG(posts_per_user))
4

total_posts/total_users
2.5700

Insights:

- The average number of posts per user provides insights into overall engagement.
- Low averages may suggest a need to encourage users to post more frequently..

Task : Identify Users Who Have Liked Every Single Photo on the Platform.

SQL Query:

```sql
USE ig_clone;

SELECT user_id,usr,photos_liked
FROM
    (SELECT likes.user_id,users.username as usr,COUNT(likes.photo_id) as photos_liked
    FROM likes
    JOIN users ON likes.user_id = users.id
    GROUP BY likes.user_id) as sq1
WHERE photos_liked = 257;
```

Output:

user_id	usr	photos_liked
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

Insights:

- Users who have liked every single photo are highly likely to be bots or fake accounts.
- This insight can help in identifying and managing non-genuine accounts, improving platform authenticity.
- Further investigation into these accounts could reveal patterns to strengthen anti-bot measures..

Results

- Conducted a comprehensive analysis of user engagement, identifying opportunities for growth and improvement.
- Delivered actionable insights for the marketing team, including strategies to engage inactive users and optimize ad campaign timing.
- Recommended measures for the product team, such as rewarding loyal users and leveraging popular hashtags.
- Identified potential bot accounts, supporting efforts to enhance platform authenticity.
- Provided data-driven insights to inform strategic decisions for improving user satisfaction and platform growth.

Insights

Loyal Users: The oldest users have been active for over a decade, ideal candidates for rewards.

Inactive Users: Many users have never posted, highlighting an opportunity for engagement campaigns.

Contest Winner: Identified the user with the most-liked photo, boosting contest participation potential.

Hashtag Trends: Top hashtags reveal trending topics and aid brand visibility.

Ad Strategy: Thursdays & Sundays see the most user registrations, ideal for ad campaigns.

Engagement Metrics: Moderate average posts per user suggest a need to encourage more activity.

Fake Accounts: Potential bots flagged for liking every photo, indicating a need for stricter verification.

Reflections

- **SQL Skills**: Gained experience in writing efficient queries and analyzing complex data.
- **Analytical Growth**: Improved ability to interpret data and provide actionable insights.
- **Challenges Overcome**:
  Navigated complex table relationships.
  Optimized query performance through iteration.
- **Key Takeaways**: Strengthened problem-solving and reinforced the importance of data-driven decisions.

THANK YOU