Instagram User Analytics

This project focuses on leveraging SQL and MySQL Workbench to analyze user interactions and and engagement within the Instagram app.

As a data analyst, the goal is to extract valuable insights from the data that can inform decision-making decision-making processes and contribute to the growth and improvement of Instagram.

The project is divided into two main categories : **Marketing Analysis and Investor Metrics**.

The Marketing Analysis tasks involve identifying loyal users, encouraging inactive users, determining determining contest winners, researching popular hashtags, and finding the best day of the week to launch week to launch ads.

On the other hand, Investor Metrics include calculating user engagement metrics and identifying potential identifying potential fake accounts on the platform.

BY-: Sanjay Singh Kanwal

Business Problem

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.

Task: Identify the five oldest users on Instagram from the provided database.

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.

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.

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.

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.

Investor Metrics

User Engagement:

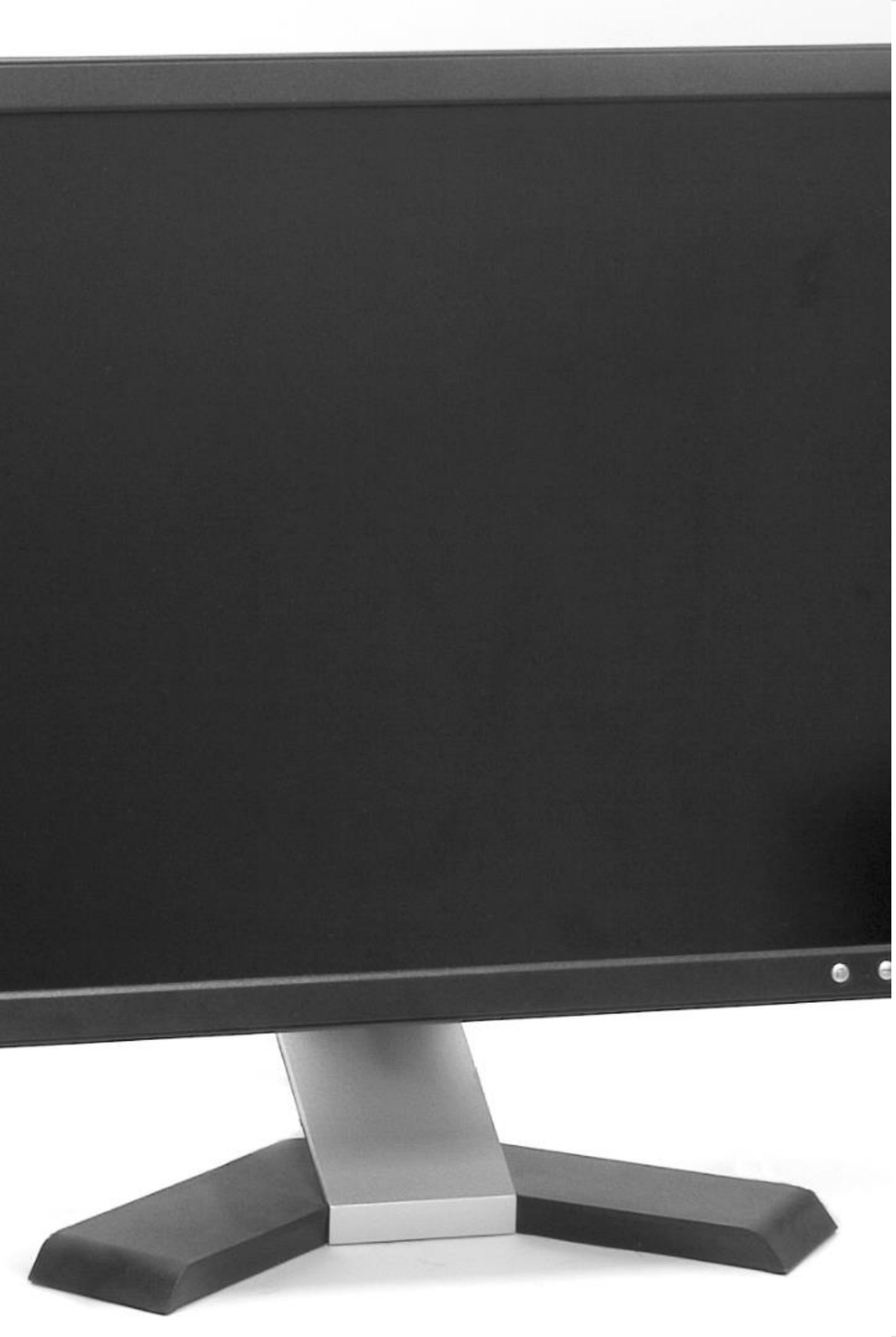
Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

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.

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.



Tech-Stack Used

For this project, MySQL Workbench was chosen as the primary tool for SQL analysis due to its robust capabilities in querying, exploring, and visualizing data, and its compatibility with MySQL databases, making it a suitable choice for handling the Instagram user data.

Approach

Data Loading and Exploration

The project begins with a thorough understanding of the database structure, including tables and columns. This exploration phase helps in preparing for data analysis.

SQL Queries

Formulate SQL queries to extract necessary information for each task.

Data Analysis

The SQL queries are executed to perform data analysis, revealing insights related to user behavior, engagement patterns, and other relevant metrics

Report Generation

The results of each SQL query are compiled into a structured report format. This step simplifies the communication of insights and facilitates decision-making

Marketing Analysis

LOYAL USER REWARD

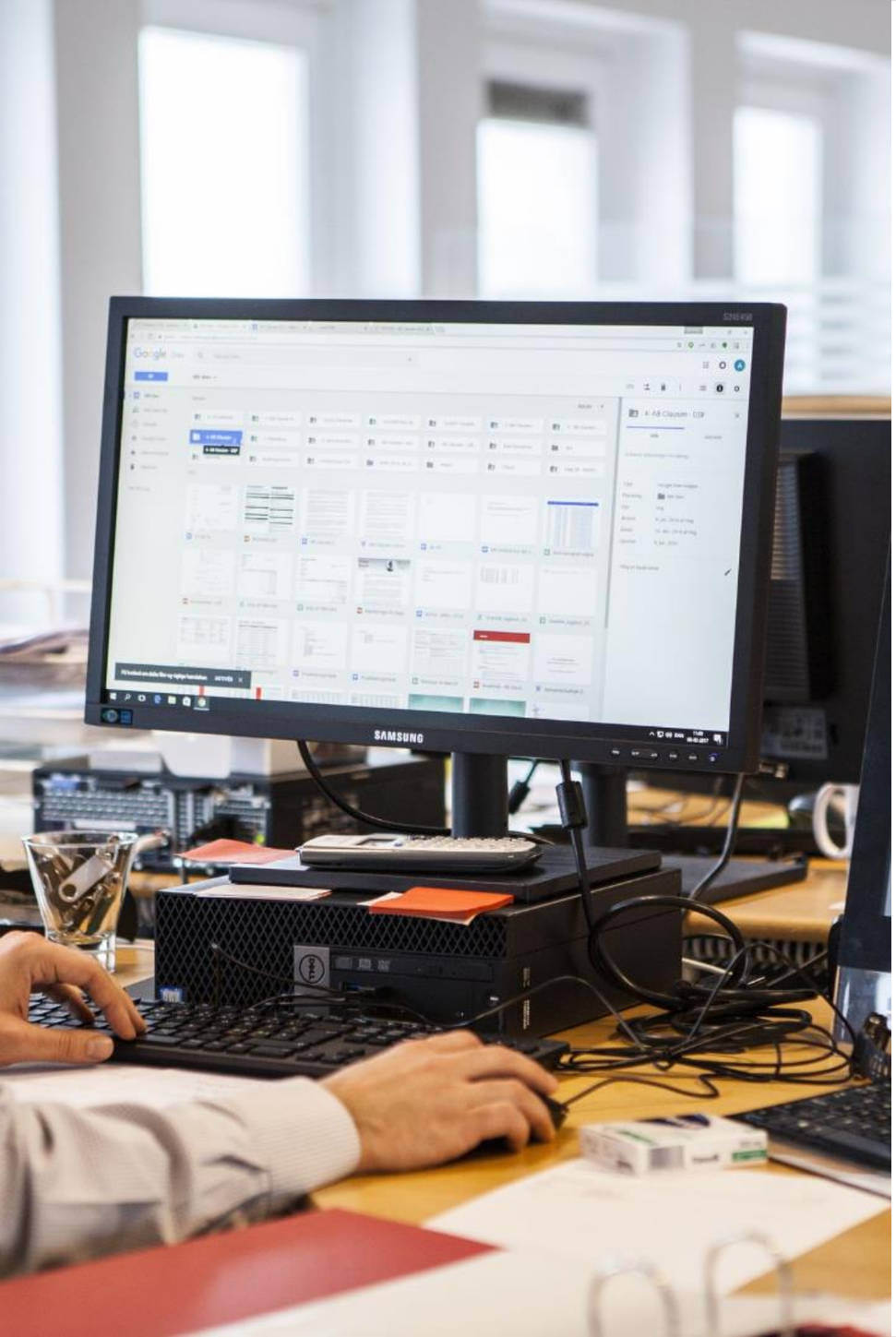
People who have been using the platform for the longest time.(Top 5 oldest Instagram users)

QUERY/PROGRAM

```
SELECT username, created_at

FROM users

ORDER BY created_at ASC

LIMIT 5;
```

OUTPUT

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



Marketing Analysis

INACTIVE USER ENGAGEMENT

Remind Inactive users to Start Posting(Users who never posted a single photo on Instagram)

QUERY/PROGRAM

```
SELECT u.id , u.username

FROM users u

LEFT JOIN photos p ON u.id = p.user_id

WHERE p.id IS NULL

ORDER BY u.id;
```

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

Marketing Analysis

CONTEST WINNER DECLARATION

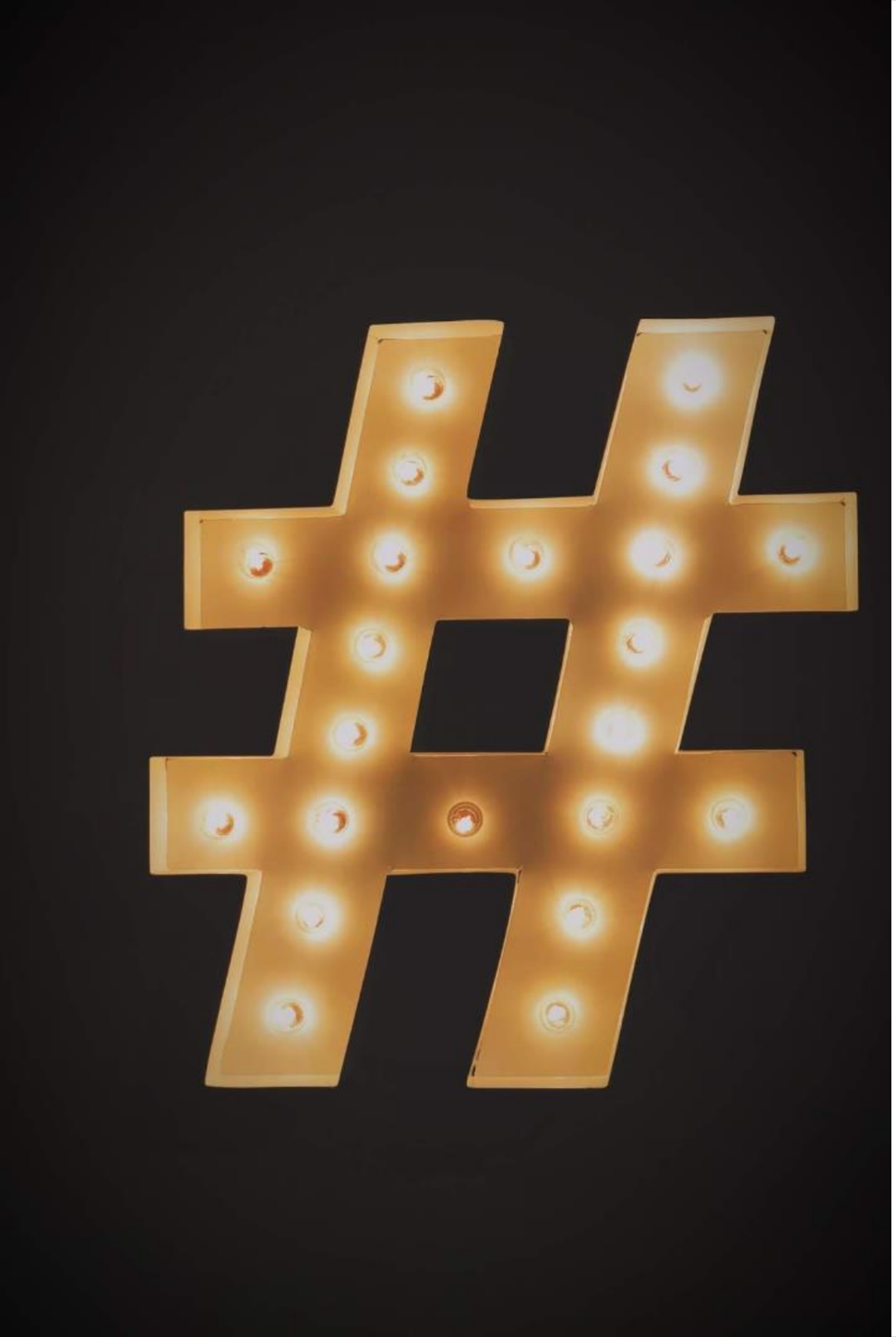
The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner. Identify the winner of the contest and provide their details to the team.

QUERY/PROGRAM

```
SELECT u.id AS user_id, u.username, p.id AS photo_id,
p.image_url, COUNT(*) AS total

FROM photos p

INNER JOIN likes l ON p.id = l.photo_id

INNER JOIN users u ON p.user_id = u.id

GROUP BY p.id

ORDER BY total DESC

LIMIT 1;
```

OUTPUT

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



Marketing Analysis

HASHTAG RESEARCH

A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.(Top 5 commonly used #Hashtags on Instagram)

QUERY/PROGRAM

```
SELECT t.tag_name, COUNT(*) AS Total

FROM tags t

JOIN photo_tags ON t.id = photo_tags.tag_id

GROUP BY t.tag_name

ORDER BY Total DESC

LIMIT 5;
```

OUTPUT

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

Marketing Analysis

AD CAMPAIGN LAUNCH

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?)

QUERY/PROGRAM

```
SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS registration_count

FROM users

GROUP BY registration_day

ORDER BY registration_count DESC

LIMIT 10;
```

OUTPUT

registration_day	registration_count
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Investor Metrics

User Engagement

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.

QUERY/PROGRAM

```
SELECT

(SELECT COUNT(DISTINCT id) FROM photos)/(SELECT COUNT(DISTINCT id) FROM users) AS
AVG_NO_POSTS_PER_USER,

(SELECT COUNT(id) FROM photos) AS TOTAL_PHOTOS,

(SELECT COUNT(id) FROM users) AS TOTAL_USERS;
```

OUTPUT

AVG_NO_POSTS_PER_USER	TOTAL_PHOTOS	TOTAL_USERS
2.5700	257	100

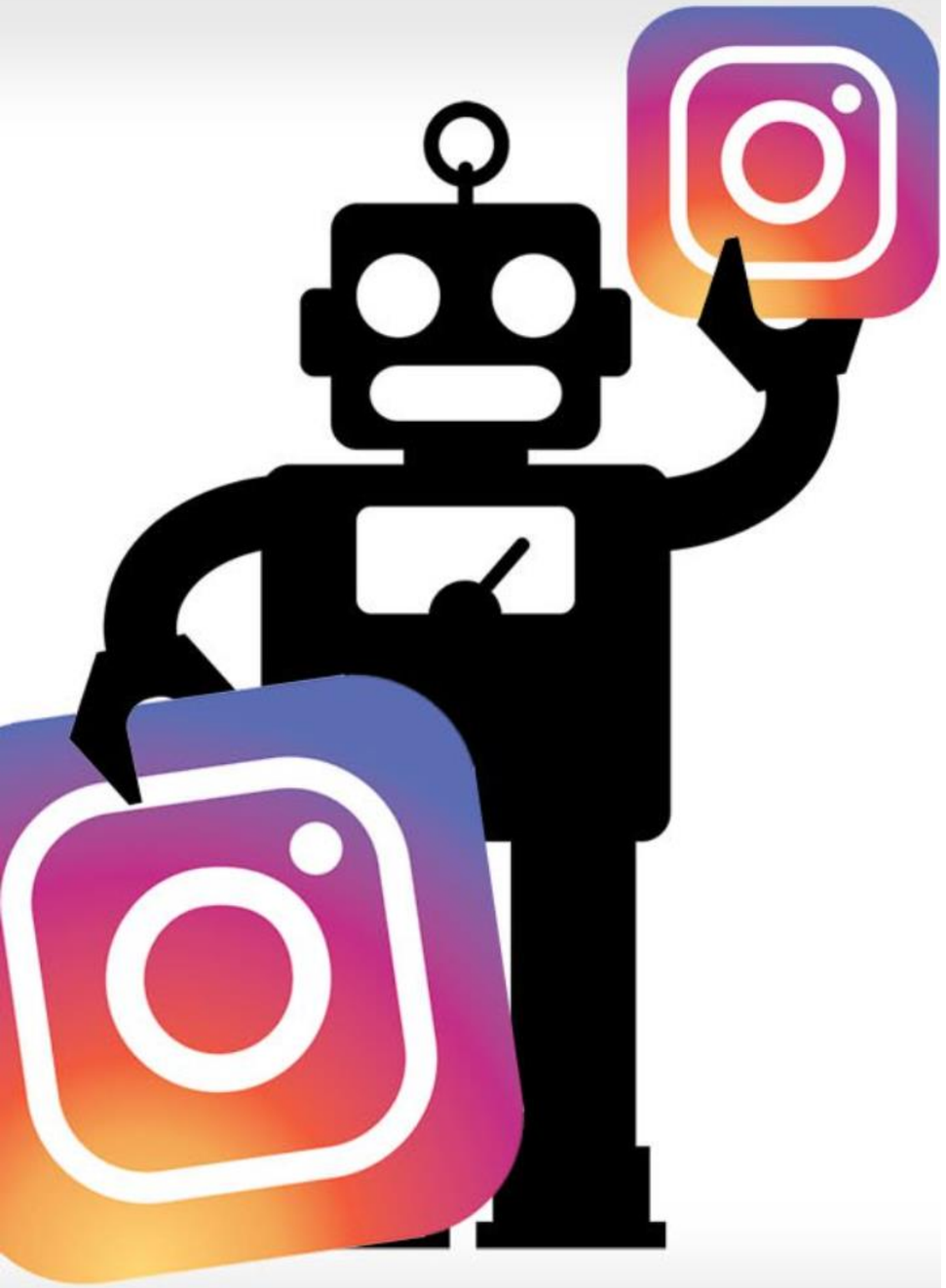
Investor Metrics

Query to find how many times each user posts on Instagram :

```
SELECT

u.username AS username, p.user_id,

HOUR(p.created_dat) AS post_hour,

COUNT(*) AS post_count

FROM

users u

INNER JOIN

photos p ON u.id = p.user_id

GROUP BY

u.username, post_hour, p.user_id

ORDER BY

p.user_id;
```

OUTPUT

username	user_id	post_hour	post_count
Kenton_Kirlin	1	12	5
Andre_Purdy85	2	12	4
Harley_Lind18	3	12	4
Arely_Bogan63	4	12	3
Travon.Waters	6	12	5
Tabitha_Schamberger11	8	12	4
Gus93	9	12	4
Presley_McClure	10	12	3
Justina.Gaylord27	11	12	5
Dereck65	12	12	4
Alexandro35	13	12	5
Billy52	15	12	4
Annalise.McKenzie16	16	12	4
Norbert_Carroll35	17	12	3
Odessa2	18	12	1
Hailee26	19	12	2
Delpha.Kihn	20	12	1
Kenneth64	22	12	1
Eveline95	23	12	12
Josianne.Friesen	26	12	5
Darwin29	27	12	1
Dario77	28	12	4
Jaime53	29	12	8
Kaley9	30	12	2
Aiyana_Hoeger	31	12	1
Irwin.Larson	32	12	4
Yvette.Gottlieb91	33	12	5
Lennie_Hartmann40	35	12	2
Yazmin_Mills95	37	12	1
Jordyn.Jacobson2	38	12	2
Kelsi26	39	12	1
Rafael.Hickle2	40	12	1
Maya.Farrell	42	12	3
Janet.Armstrong	43	12	5
Seth46	44	12	4
Malinda_Streich	46	12	4
Harrison.Beatty50	47	12	5
Granville_Kutch	48	12	1
Gerard79	50	12	3
Mariano_Koch3	51	12	5
Zack_Kemmer93	52	12	5
Meggie_Doyle	55	12	1
Peter.Stehr0	56	12	1
Aurelie71	58	12	8
Cesar93	59	12	10
Sam52	60	12	2
Jayson65	61	12	1
Ressie_Stanton46	62	12	2
Elenor88	63	12	4
Florence99	64	12	5
Adelle96	65	12	5
Emilio_Bernier52	67	12	3
Karley_Bosco	69	12	1
Erick5	70	12	1
Kathryn80	72	12	5
Jaylan.Lakin	73	12	1
Donald.Fritsch	77	12	6
Colten.Harris76	78	12	5
Katarina.Dibbert	79	12	1
Aracely.Johnston98	82	12	2
Alysa22	84	12	2
Milford_Gleichner42	85	12	2
Delfina_VonRueden68	86	12	9
Rick29	87	12	4
Clint27	88	12	11
Frederik_Rice	92	12	3
Willie_Leuschke	93	12	2
Damon35	94	12	1
Nicole71	95	12	2
Keenan.Schamberger60	96	12	3
Tomas.Beatty93	97	12	2
Imani_Nicolas17	98	12	1
Alek_Watsica	99	12	3
Javonte83	100	12	2



Investor Metrics

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).

QUERY/PROGRAM

```
SELECT user_id,
username, count(*) AS
Total_Likes

FROM users

INNER JOIN likes

ON users.id =
likes.user_id

GROUP BY likes.user_id

HAVING TOTAL_LIKES =
(SELECT count(*) FROM
photos);
```

OUTPUT

user_id	username	Total_Likes
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

> The top 5 oldest and most loyal users are: **Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71, Jordyn.Jacobson2**

> The winner of the contest is Zack_Kemmer93 with user_id - 52. His photo with photo_id - 145 has the highest number of likes, at 48.

> The top 5 most commonly used #hashtags, along with their total counts are: **smile(59), beach(42), party(39), fun(38), concert(24)**

> The majority of users registered on Thursdays and Sundays, with 16 users registering on each day. It would prove beneficial to start AD Campaigns on these days.

> There are a total of 257 rows, or 257 photos, in the photos table. There are also 100 rows, or 100 user IDs, in the users table. This makes the desired output to be 257/100 = 2.57 (the average number of posts per user on Instagram).

> Out of all the users, 13 user IDs have liked every post on Instagram. These users are considered to be bots or fake accounts.

5 Why Approach

Implement the 5 Why Approach to identify the root cause of the issues and help resolve them.

- The "5 Whys" approach is a problem-solving technique used in various fields, including data analysis, to identify the root causes of an issue or a series of questions.
- In data analysis, the "5 Whys" can be applied to investigate and understand the underlying reasons behind data trends, anomalies, or problems.
- The "5 Whys" approach helps in systematically identifying and addressing the underlying causes of data-related issues, ensuring that solutions are targeted at the root of the problem rather than just addressing symptoms.

Question	Reason
Why did the Marketing team want to identify the most inactive users?	• The Marketing team aimed to proactively engage with inactive users through targeted promotional emails or campaigns. By re-engaging these users, the team aimed to boost user activity on the platform. Increased user activity benefits both users and advertisers, as active users are more likely to interact with new content.
Why did the Marketing team want to identify the top 5 hashtags used?	• Identifying the top five hashtags used on the platform is crucial for strategic content planning. Hashtags play a fundamental role in Instagram's content discovery and reach mechanism. By understanding the most popular hashtags, the Marketing team can align their content with trending topics, reach a wider audience, and increase engagement.
Why did the Marketing team want to determine the day of the week with the highest number of new user registrations?	• Determining the day of the week with the most new user registrations on Instagram is essential for optimizing ad campaign scheduling.Targeting a fresh audience that hasn't seen previous campaigns can be highly effective.
Why did the Investors want to know the average number of posts per user on Instagram?	• Investors were interested in the average number of posts per user on Instagram as a measure of user engagement. A higher average posts-per-user value indicates active and engaged users, which is a positive sign for the platform's health.
Why did the Investors want to identify the count of bots and fake accounts?	• Investors wanted to assess the authenticity and quality of the platform's user base by detecting potential bots and fake accounts. An excessive presence of fake accounts can lead to misleading engagement metrics and undermine the platform's trustworthiness.

Conclusion

Our project has achieved several significant milestones:

- Providing valuable insights to the marketing team and investors, enabling them to make data-driven decisions regarding user engagement, marketing strategies, and the platform's authenticity.
- Demonstrating the importance of SQL analysis in extracting actionable insights from vast datasets, highlighting the role of data-driven decision-making in shaping the platform's future.

- The insights derived from this analysis can potentially influence the development and strategic decision-making processes at Instagram, contributing to its continued success and growth as a leading social media platform.