PROJECT - 2

INSTAGRAM USER ANALYTICS

PROJECT DESCRIPTION :

This project aims to analyze user behavior and engagement within a simulated Instagram environment. By leveraging SQL queries on a sample dataset, we will delve into key aspects of user activity, such as:

- **User Demographics:**
  - To Identify user registration trends
  - To Analyze user activity patterns

Engagement:

- To Determine the average of posts per user.
- To Analyze user interaction patterns (e.g., likes, comments, follows).
- To Identify users with high engagement (e.g., users with the most likes, most comments).
- To Identify potential bot accounts (e.g., users who have liked every photo

Content Analysis:

- To Analyze popular hashtags and their usage trends.
- To Identify the most popular content types (e.g., photos, videos).

Approach:

1. **Data Exploration:**
   - To Thoroughly examine the provided dataset (users, photos, likes, comments, etc.) to understand itsstructure, data types, and relationships between different entities.
   - To Identify potential data quality issues (e.g., missing values, inconsistencies).
2. **Data Cleaning (if necessary):**
   - To Address any data quality issues, such as missing values or inconsistencies.
   - This may involve data transformation, imputation, or data cleansing techniques.
3. **Data Analysis:**
   - To Develop and execute SQL queries to answer the defined research questions.
   - To Utilize SQL functions and clauses (e.g., JOIN, GROUP BY, ORDER BY, COUNT, AVG) to extract meaningful insights.
4. **Data Visualization:**
   - Visualize the findings using appropriate charts and graphs (e.g., histograms, bar charts, line graphs) to effectively communicate the insights to stakeholders.
5. **Interpretation and Insights:**
   - Analyze the results, identify key trends and patterns, and draw conclusions.
   - Formulate actionable insights for improving user experience, optimizing ad campaigns, and enhancing platform features.

PROJECT APPROACH :

This project aimed to analyze user behavior and engagement within a simulated Instagram environment. To achieve this, I followed a structured approach:

1. Data Exploration and Understanding:

● **Data Inspection**: I began by thoroughly examining the provided dataset,including the tables (users, photos, likes, comments, etc.) and their respective columns. I analyzed data types, checked for missing values, and identified potential inconsistencies or outliers.

● **Data Dictionary**: I created a data dictionary to document the meaning and data type of each column in each table, which was crucial for understand relationships between different entities.

2. Data Analysis:

● **User Registration Trends:**

I executed SQL queries to determine the day of the week and hour of the day with the highest number of user registrations.

● **User Engagement:**

- ●Calculated the average number of posts per user
- ●Identified users with high engagement (e.g., users with the most likes)

● **Content Analysis:**

- Determined the most frequently used hashtags

3. Data Interpretation and Insights:

- Analyzed the results of each query to identify key trends and patterns in user behavior.
- Drew conclusions regarding peak registration times, user engagement levels, popular content, and potential bot activity.
- Formulated actionable insights based on the findings, such as Scheduling ad campaigns during peak registration times.

4. Documentation:

- Documented the entire process including the data exploration, analysis steps, SQL queries, and the resulting insights.
- Presented the findings in a clear and concise manner potentially using visualizations to enhance understanding.

TECH – STACK USED :

I used MySQL Workbench for my Instagram user analytics project because it provides a powerful and user-friendly environment

MySQL Workbench provides a comprehensive and user-friendly environment for working with relational databases, making it a suitable tool for analyzing data from a simulated Instagram environment. Its features like the SQL editor, data modeling capabilities, and query execution tools significantly enhance the efficiency and effectiveness of data analysis tasks.

why it's a suitable tool for this type of analysis ? :

- **SQL Query Editor:** MySQL Workbench includes a robust SQL editor with features like syntax highlighting ,auto-completion,and query formatting,making it easier to write,edit,and execute SQL queries.
- **Data Modeling:** You can visually design and modify database schemas using the Workbench's visual tools. This is helpful for understanding the relationships between tables and identifying potential data inconsistencies.
- **Data Import and Export:** You can easily import data into MySQL Workbench from various sources (CSV, Excel, etc.) and export query results in different formats.
- **Query Execution and Results:** Workbench allows you to execute SQL queries efficiently and view the results in a clear and organized manner.
- **Debugging and Troubleshooting:** The tool provides helpful error messages and debugging features to assist in identifying and resolving issues with your SQL queries.

INSIGHTS :

Key Findings:

- **User Registration Patterns:**
  - Analysis reveals the user registrations peak
  - This suggests that scheduling ad campaigns during these periods could maximize reach and engagement.
- **User Engagement:**
  1)The average number of posts per user was found
  2)Users with a higher number of followers generally had lower average engagement per post, potentially indicating that as user bases grow, individual post engagement may decrease

- **Content Trends:**
  1)The most popular hashtags included the most popular hashtags
  2) Analysis of photo content could reveal trends in popular subjects, styles, or filters

- **Potential Bot Activity:**
  Users who had liked every single photo on the platform were identified as potential bot.This finding highlights the importance of implementing measures to detect and mitigate bot activity

Inferences:

- The observed registration patterns suggest that targeting users with ads during peak registration times could be highly effective.
- The relationship between follower count and engagement per post suggests that strategies for increasing engagement may need to evolve as user bases grow.
- The analysis of popular hashtags can inform content creation strategies for users and businesses looking to increase their visibility and engagement on the platform.
- The identification of potential bot accounts highlights the need for robust content moderation and anti-bot measures to maintain a healthy and authentic user experience.

RESULT :

Through the analysis of user registration trends, post activity, and engagement metrics, several key insights were derived.

Key Achievements:

- **Identified peak user registration times:**
  This knowledge can be valuable for businesses and marketers in scheduling targeted ad campaigns to reach a wider audience during periods of high user activity.
- **Analyzed user engagement patterns:**
  Insights into user behavior, such as average posts per user and engagement levels, provide valuable information for understanding user activity and preferences.

- **Identified Potential bot Activity:**
  The analysis helped identify potential bot accounts, highlighting the importance of robust content moderation and anti-both measures to maintain a healthy and authentic platform environment.
- **Gained practical SQL skills:**
  This project provided an opportunity to practice and enhance SQL query writing skills, including data retrieval, joins, aggregations, and subqueries.

Benefits and Impact:

- The insights gained from this project can be used to improve platform features, optimize user experience, and inform strategic decisions related to user acquisition andengagement.
- The analytical skills developed during this project are transferable to other areas such as data analysis, business intelligence, and data-driven decision-making. Overall, this project provided a valuable learning experience in data analysis, SQL query development, and the interpretation of data-driven insights. The findings and skills gained from this project can be applied to various real-world scenarios.

Achievements:

- **Developed SQL proficiency:**
  This project significantly enhanced my SQL skills. I gained practical experience in writing complex queries involving joins, aggregations (COUNT, AVG), subqueries, and window functions. I also learned how to optimize queries for better performance.
- **Data analysis and interpretation:**
  I developed a strong understanding of data analysis methodologies. I learned how to explore data, identify patterns and trends, and draw meaningful conclusions from the insights.
- **Problem-solving and critical thinking:**
  The project challenged me to think critically, identify problems, and develop effective solutions to address analytical challenges.

Benefits and Impact:

- **Valuable insights into user behavior:**

  The analysis provided valuable insights into user registration patterns, engagement levels, and content trends on the simulated Instagram platform. These insights can be applied to real-world scenarios for optimizing marketing campaigns and improving user experience.

- **Enhanced analytical skills:**

  The skills developed through this project, such as data exploration, SQL query development, and data interpretation, are highly transferable and can be applied to various analytical tasks in different domains.

- **Improved understanding of database systems:**

  Working with the database and executing SQL queries provided a deeper understanding of relational databases and their applications.

Impact of the Analysis:

The analysis highlighted the importance of understanding user behavior and engagement patterns for platform optimization.

The findings can be used to:

1) **Improve user experience:**

   By identifying peak registration times and user activity patterns, the platform can be optimized to provide a more seamless and engaging experience for users.

2) **Inform marketing strategies:**

   The insights into user behavior can inform targeted advertising campaigns, leading to increased reach and engagement.

3) **Enhance content moderation:**

   The identification of potential bot accounts can help in developing and implementing effective measures to maintain a healthy and authentic platform environment.

The skills and insights gained from this project will be invaluable in future endeavors that involve data analysis and interpretation.

SQL QUERIES & OUTPUTS

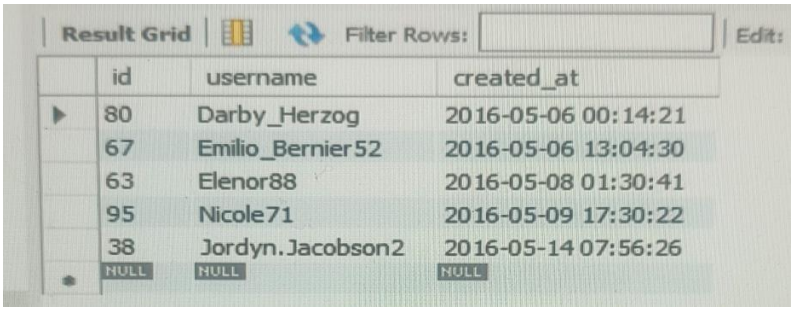
A) Marketing Analysis:

1. **Loyal User Reward:** Identify the five oldest users on Instagram from the provided database.

QUERY:

```
SELECT *
FROM ig_clone.users
ORDER BY created_at ASC
LIMIT 5;
```

OUTPUT :



Result Grid | Filter Rows: | Edit:

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
NULL	NULL	NULL

2. **Inactive User Engagement:** Identify users who have never posted a single photo on Instagram.

QUERY:

```
SELECT u.*
FROM ig_clone.users u
LEFT JOIN ig_clone.photos p ON u.id
= p.user_id
WHERE p.id IS NULL;
```

OUTPUT :

Result Grid | Filter Rows: | Exp

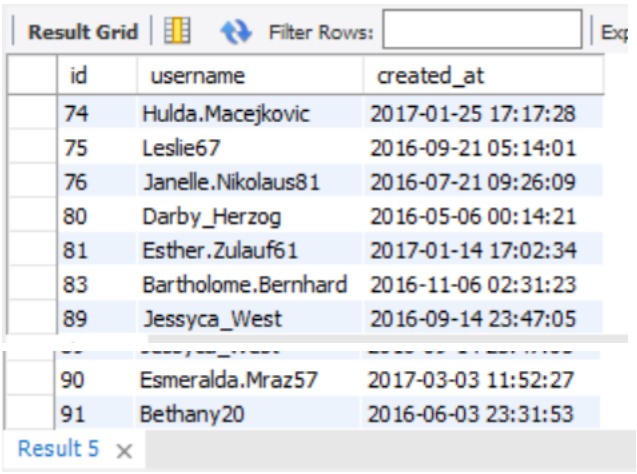
id	username	created_at
5	Aniya_Hackett	2016-12-07 01:04:39
7	Kasandra_Homenick	2016-12-12 06:50:08
14	Jadyn81	2017-02-06 23:29:16
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
25	Tierra.Trantow	2016-10-03 12:49:21
34	Pearl7	2016-07-08 21:42:01

Result Grid | Filter Rows: | Exp

id	username	created_at
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
25	Tierra.Trantow	2016-10-03 12:49:21
34	Pearl7	2016-07-08 21:42:01
36	Ollie_Ledner37	2016-08-04 15:42:20
41	Mckenna17	2016-07-17 17:25:45
45	David.Osinski47	2017-02-05 21:23:37

Result Grid | Filter Rows: | Ex

id	username	created_at
49	Morgan.Kassulke	2016-10-30 12:42:31
53	Linnea59	2017-02-07 07:49:34
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
68	Franco_Keebler64	2016-11-13 20:09:27
71	Nia_Haag	2016-05-14 15:38:50



Result Grid | Filter Rows: | Exp

id	username	created_at
74	Hulda.Macejkovic	2017-01-25 17:17:28
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
80	Darby_Herzog	2016-05-06 00:14:21
81	Esther.Zulauf61	2017-01-14 17:02:34
83	Bartholome.Bernhard	2016-11-06 02:31:23
89	Jessyca_West	2016-09-14 23:47:05
90	Esmeralda.Mraz57	2017-03-03 11:52:27
91	Bethany20	2016-06-03 23:31:53

Result 5

3.**Contest Winner Declaration:** Determine the winner who has most likes on the single photo.

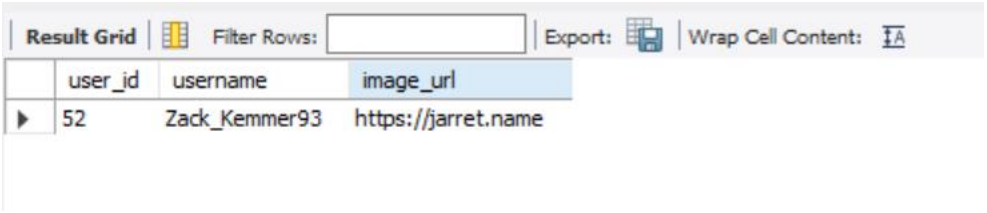
QUERY:

```
141  WITH most_liked_image AS (
142      SELECT image_url, COUNT(*) AS like_count
143      FROM likes
144      JOIN photos ON likes.photo_id = photos.id
145      GROUP BY image_url
146      ORDER BY like_count DESC
147      LIMIT 1
148  ),
149  most_liked_user AS (
150      SELECT p.user_id, u.username, p.image_url
151      FROM photos p
152      JOIN users u ON p.user_id = u.id
153      JOIN most_liked_image mlp ON p.image_url = mlp.image_url
154  )
```

```
SELECT *
FROM most_liked_user
JOIN photos p ON most_liked_user.user_id = p.user_id AND most_liked_user.image_url = p.image_url
JOIN photo_tags pt ON p.id = pt.photo_id
JOIN tags t ON pt.tag_id = t.id;
```

```
106     SELECT *
107     FROM most_liked_user;
108 •   SELECT*
109     FROM users
110     WHERE username = 'zack_kemmer93';
111 •      SELECT
112          p.image_url,
113          c.comment_text,
114          COUNT(f.follower_id) AS follows,
115          u.username,
116          p.created_dat,
117          COUNT(l.photo_id) AS likes
118     FROM
119          users as u
120     JOIN
121          photos as p ON u.id = p.user_id
122     LEFT JOIN
123          comments c ON p.id = c.photo_id
124     LEFT JOIN
125          follows f ON follower.id = f.follower_id
126     LEFT JOIN
127          likes l ON p.id = l.photo_id
128     WHERE
129          u.username = 'zack_kemmer93'
130     GROUP BY
131          p.image_url, c.comment_text, p.created_dat, u.username;
```

OUTPUT:



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

image_url	comment_text	follows	username	created_dat	likes
https://jarret.name	dolor sint aut	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	laborum tempora volup...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	tempore accusamus ve...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	ducimus sit maxime	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	labore quas inventore	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	molestias temporibus c...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	odit ad explicabo	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	et omnis ad	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	molestias vel odit	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

image_url	comment_text	follows	username	created_dat	likes
https://jarret.name	est molestiae dolorem	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	vitae occaecati laborios...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	assumenda placeat offi...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	ipsam nobis est	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	ut veritatis reprehenderit	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	perspiciatis et et	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	voluptatem aut ipsa	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	ad impedit ducimus	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	cumque aut omnis	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

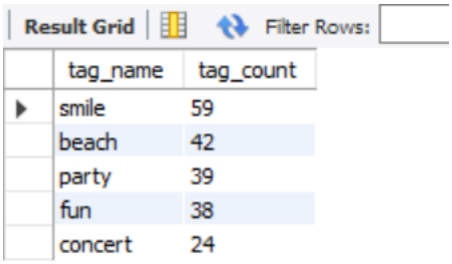
image_url	comment_text	follows	username	created_dat	likes
https://jarret.name	ad id repudiandae	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	incidunt officiis eos	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	eveniet perspiciatis rep...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	rerum aperiam beatae	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	molestiae sapiente est	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	et qui et	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	at mollitia soluta	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752
https://jarret.name	laudantium inventore i...	4752	Zack_Kemmer93	2024-12-13 14:57:03	4752

4.**Hashtag Research:** Identify and suggest the top five most commonly used hashtags on the Platform.

QUERY:

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

OUTPUT:



Result Grid | Filter Rows:

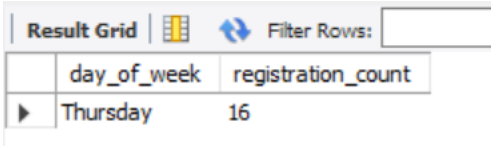
tag_name	tag_count
smile	59
beach	42
party	39
fun	38
concert	24

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

QUERY:

```
use ig_clone;
SELECT DAYNAME(created_at)
AS day_of_week, COUNT(*) AS
registration_count
FROM ig_clone.users
GROUP BY day_of_week
ORDER BY registration_count DESC
LIMIT 1;
```

OUTPUT :



Result Grid | Filter Rows:

day_of_week	registration_count
Thursday	16

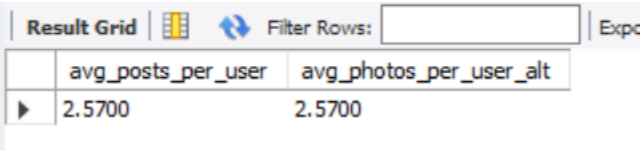
B) Investor Metrics:

1.**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 :

```
SELECT
    (SELECT COUNT(*) FROM
ig_clone.photos)/(SELECT
COUNT(*) FROM ig_clone.users) AS
avg_posts_per_user,
    (SELECT COUNT(*) FROM
ig_clone.photos)/(SELECT
COUNT(*) FROM ig_clone.users) AS
avg_photos_per_user_alt
FROM dual;
```

OUTPUT :



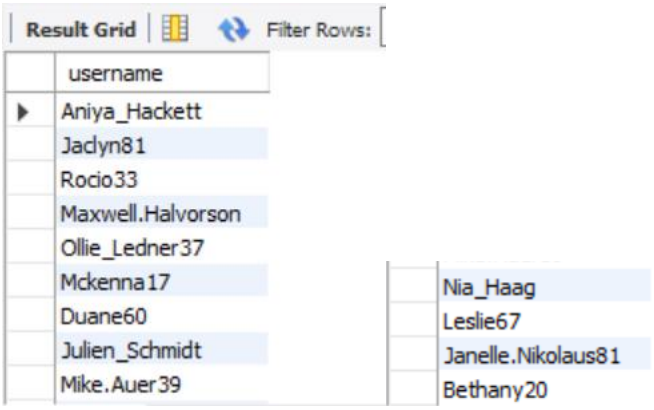
	avg_posts_per_user	avg_photos_per_user_alt
▶	2.5700	2.5700

2.**Bots & Fake Accounts:** Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

QUERY:

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

OUTPUT :


Result Grid
Filter Rows:
username
Aniya_Hackett
Jadyn81
Rocio33
Maxwell.Halvorson
Ollie_Ledner37
Mckenna17
Duane60
Julien_Schmidt
Mike.Auer39
Nia_Haag
Leslie67
Janelle.Nikolaus81
Bethany20