

Instagram User Analytics

Purpose:

The Instagram User Analytics project aims to leverage SQL and MySQL Workbench to analyze user interactions and engagement on the Instagram platform. The primary goal is to derive meaningful insights that can inform decision-making across various teams within the business, including marketing, product development, and user experience enhancement.

Approach:

1. Database Creation:

- A MySQL database, named **ig_clone**, was created using the provided commands.
- Tables were established to capture user data, photo details, comments, likes, follows, tags, and the relationship between photos and tags.

2. Analysis Tasks:

- **Marketing Analysis:**
 - *Loyal User Reward:* Identified the five oldest users on Instagram by querying the **users** table and ordering by the **created_at** timestamp.
 - *Inactive User Engagement:* Detected users who have never posted a photo by querying the **users** and **photos** tables.
 - *Contest Winner Declaration:* Determined the winner of a contest by finding the user with the most likes on a single photo using the **likes** table.
 - *Hashtag Research:* Identified the top five most commonly used hashtags by querying the **tags** and **photo_tags** tables.
 - *Ad Campaign Launch:* Analyzed user registration patterns to suggest the best day of the week for ad campaign launches.
- **Investor Metrics:**

- *User Engagement:* Calculated the average number of posts per user and provided the total number of photos divided by the total number of users.
- *Bots & Fake Accounts:* Identified potential bots by finding users who have liked every single photo, using the **users**, **photos**, and **likes** tables.

3. Reporting:

- The findings and insights derived from the analysis were compiled into a comprehensive report presented in PDF format.
- Each section of the report includes SQL queries, outputs, and relevant insights.

Tech-Stack Used:

MySQL Workbench: Chosen for its robust features, ease of use, and compatibility with MySQL databases.

Insights:

- The project uncovered valuable insights about user behaviour, engagement patterns, and potential areas for improvement.
- The identification of loyal users, inactive users, contest winners, and popular hashtags can inform marketing strategies.
- Investor metrics provide an understanding of overall user engagement and potential presence of fake accounts.

Results:

- The project resulted in a comprehensive report that provides actionable insights for various teams within Instagram.

- The findings contribute to informed decision-making processes, potentially influencing the platform's future development.





SQL Query Execution and Results

Figure 1: Identify the Five Oldest Users on Instagram

```

96
97  -- Identify the five oldest users on Instagram:
98 • SELECT * FROM users
99  ORDER BY created_at
100  LIMIT 5;
101

```

Result Grid			
Filter Rows: <input type="text"/>			
Edit:   			
Export/Import: 			
	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

Caption: SQL query to identify the five oldest users based on the 'created_at' timestamp. The results show the user details.

Insights:

- The identified users have been using the Instagram platform for the longest time.
- Recognizing these users is crucial for loyalty reward programs and targeted engagement.

Figure 2: Inactive User Engagement Analysis

```
106  -- Identify users who have never posted a single photo on Instagram:
107 •  SELECT u.id, u.username
108      FROM users u
109      LEFT JOIN photos p
110      ON u.id = p.user_id
111      WHERE p.id IS NULL;
```

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

	id	username
▶	5	Aniya_Hackett
	7	Kassandra_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

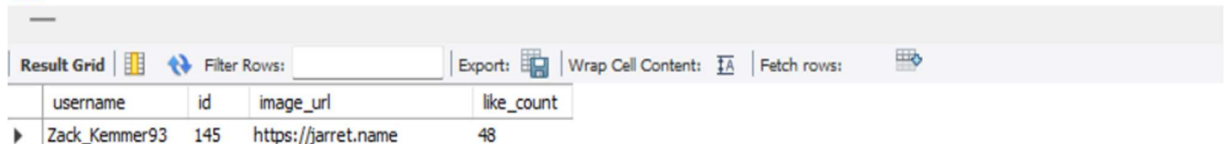
Caption: SQL query to identify users who have never posted a single photo. The results show user IDs and usernames.

Insights:

- These users have not posted any photos on Instagram, indicating inactivity.
- Targeting these users with promotional emails can encourage them to start posting and increase engagement.

Figure 3: Determine the Contest Winner with the Most Likes

```
116
117  -- Determine the winner of the contest with the most likes on a single photo:
118  SELECT username, photos.id, photos.image_url, COUNT(likes.user_id) AS like_count
119  FROM photos
120  INNER JOIN likes
121  ON likes.photo_id = photos.id
122  INNER JOIN users
123  ON photos.user_id = users.id
124  Group BY photos.id
125  ORDER BY like_count DESC
126  LIMIT 1;
127
```



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

Caption: SQL query to determine the contest winner with the most likes on a single photo. The results show the winner's details.

Insights:

- This user has the most likes on a single photo, making them the winner of the contest.
- Recognizing and rewarding such users can boost user engagement and participation in future contests.

Figure 4: Identify Top Five Most Commonly Used Hashtags

```
131
132 -- Identify and suggest the top five most commonly used hashtags:
133 • SELECT t.tag_name, COUNT(pt.photo_id) AS tag_count
134 FROM tags t
135 JOIN photo_tags pt ON t.id = pt.tag_id
136 GROUP BY t.tag_name
137 ORDER BY tag_count DESC
138 LIMIT 5;
139
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
tag_name	tag_count				
smile	59				
beach	42				
party	39				
fun	38				
concert	24				

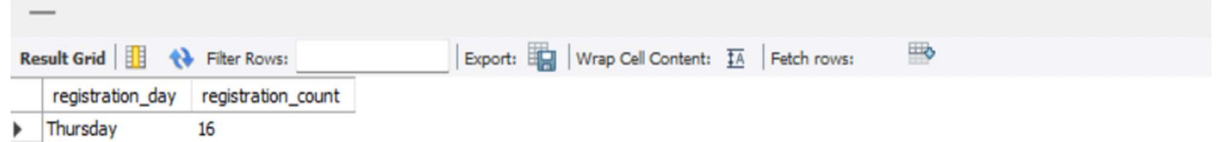
Caption: SQL query to identify the top five most commonly used hashtags. The results show tag names and usage counts.

Insights:

- These are the most popular hashtags on the platform, indicating trending topics.
- Brands can leverage these hashtags for increased visibility and audience reach.

Figure 5: Determine the Best Day for Ad Campaign Launch

```
144
145  -- Determine the day of the week when most users register on Instagram:
146 • SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS registration_count
147 FROM users
148 GROUP BY registration_day
149 ORDER BY registration_count DESC
150 LIMIT 1;
151
```



registration_day	registration_count
Thursday	16

Caption: SQL query to determine the day of the week when most users register on Instagram. The results show the most active registration day.

Insights:

- Thursday is the day with the highest number of user registrations.
- Scheduling ad campaigns on Thursdays may maximize reach and engagement.

Figure 6: Top 3 Peak Registration Hours on Thursdays

```
151
152 -- top 3 peak registration hours on Thursdays
153 • SELECT HOUR(created_at) AS registration_hour, COUNT(*) AS registration_count
154 FROM users
155 WHERE DAYNAME(created_at) = 'Thursday'
156 GROUP BY registration_hour
157 ORDER BY registration_count DESC
158 LIMIT 3;
159
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
registration_hour	registration_count			
23	2			
9	2			
7	2			

Caption: SQL query to identify the top 3 peak registration hours on Thursdays. The results show peak hours and registration counts.

Insights:

- Thursday evenings, especially during the peak hours, see the highest user registrations.
- Optimizing ad campaigns during these hours may lead to increased user engagement.

Figure 7: Top 5 Peak Registration Hours Across All Days

```
159
160  -- top 5 peak registration hours across all days,
161  -- including the day associated with each peak hour
162 • SELECT
163     DAYNAME(created_at) AS registration_day,
164     HOUR(created_at) AS registration_hour,
165     COUNT(*) AS registration_count
166 FROM users
167 GROUP BY registration_day, registration_hour
168 ORDER BY registration_count DESC
169 LIMIT 5;
```

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

	registration_day	registration_hour	registration_count
▶	Tuesday	12	3
	Sunday	17	3
	Wednesday	17	3
	Friday	23	3
	Friday	19	3

Caption: SQL query to identify the top 5 peak registration hours across all days. The results show the most active hours and registration counts.

Insights:

- These are the top 5 peak registration hours across all days, providing a comprehensive view.
- Ad campaigns can be strategically scheduled during these peak hours for maximum impact.

Figure 8: User Engagement Metrics

```
178
179  -- Average number of posts per user
180 •  SELECT
181     AVG(post_count) AS avg_posts_per_user
182     FROM (SELECT user_id, COUNT(*) AS post_count
183           FROM photos
184           GROUP BY user_id) AS user_posts;
185
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	avg_posts_per_user			
▶	3.4730			

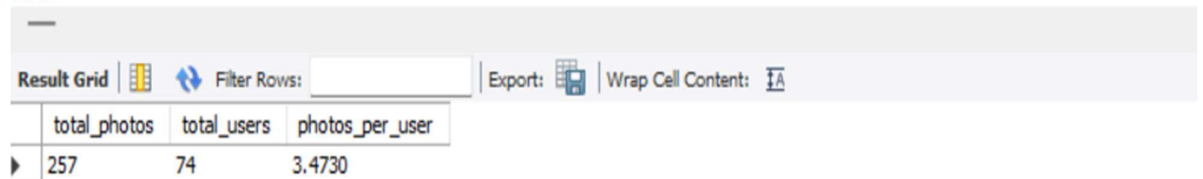
Caption: SQL query to calculate the average number of posts per user. The result shows the average posts per user.

Insights:

- On average, users contribute a certain number of posts, indicating overall engagement.
- Monitoring this metric helps understand user activity and content creation trends.

Figure 9: Total Number of Photos and Users

```
185
186  -- Total number of photos divided by the total number of users
187 • SELECT COUNT(*) AS total_photos, COUNT(DISTINCT user_id) AS total_users,
188        COUNT(*) / COUNT(DISTINCT user_id) AS photos_per_user
189 FROM photos;
190
```



The screenshot shows a SQL query result grid with the following data:

	total_photos	total_users	photos_per_user
▶	257	74	3.4730

Caption: SQL query to calculate the total number of photos and users, along with the photos per user ratio.

Insights:

- The platform has a certain number of photos contributed by a specific number of users.
- Calculating the photos per user ratio provides insights into user activity and content distribution.

- Figure 10: Identify Potential Bots

```

197  -- Identify potential bots who have liked every single photo on the site:
198 • SELECT id, username
199  FROM users u
200  WHERE NOT EXISTS (
201      SELECT id
202      FROM photos p
203      WHERE NOT EXISTS (
204          SELECT 1
205          FROM likes l
206          WHERE l.user_id = u.id AND l.photo_id = p.id
207      )
208  );

```

Result Grid

	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
*	NULL	NULL

- Caption:** SQL query to identify potential bots who have liked every single photo. The results show user IDs and usernames.
- Insights:**
 - Users listed here have liked every single photo on the site, a behaviour unusual for normal users.
 - Identifying such users is essential for investors to assess the presence of potential fake or bot accounts on the platform.

Conclusion:

The Instagram User Analytics project demonstrates the power of SQL in extracting meaningful insights from user data. The findings can guide strategic decisions, enhance user experience, and contribute to the platform's growth.