INSTAGRAM USER ANALYTICS

Problem Description:

Our goal in this data analysis project is to examine user engagement and interactions with the Instagram app using SQL and MySQL Workbench. The analysis's conclusions will play a crucial role in shaping the choices made by the company's marketing, product development departments. Our goal is to offer practical suggestions that can will help in the expansion and improvement of the Instagram platform by examining user behaviour, registration trends, and engagement data.

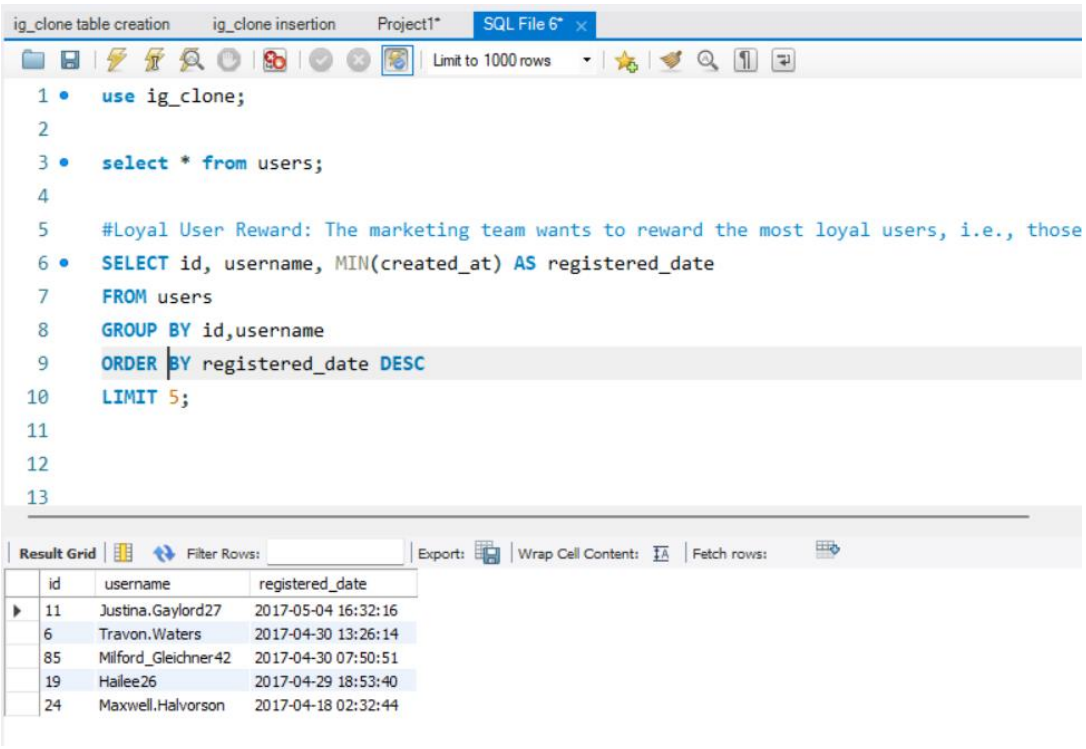
Approach:

A) Marketing Analysis:

1.Loyal User Reward:

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

SQL Query:



Explanation:

Query 1:

- The query “MIN(created_at)” is used to get the minimum date the users have created their accounts and an alias “registered_date” is provided to it using AS keyword and the SELECT statement is used to select the username and registered_date from users table.
- “GROUP BY username” query is used to group the result by username (which we got from the users table)
- “ORDER BY registered_date” query is to sort the result-set by oldest registered date
- “LIMIT 5” query makes sure that the result set contains only 5 records

2. Inactive User Engagement:

Problem statement: The team wants to encourage inactive users to start posting by sending them promotional emails.
Your Task: Identify users who have never posted a single photo on Instagram.

SQL Query:

```
23
24    #Inactive User Engagement: The team wants to encourage inactive users to start p
25
26 •  SELECT * FROM users
27    WHERE id NOT IN(
28    SELECT user_id FROM photos);
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

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
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
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

Result Grid | Filter Rows: | Edit:

id	username	created_at
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
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
NULL	NULL	NULL

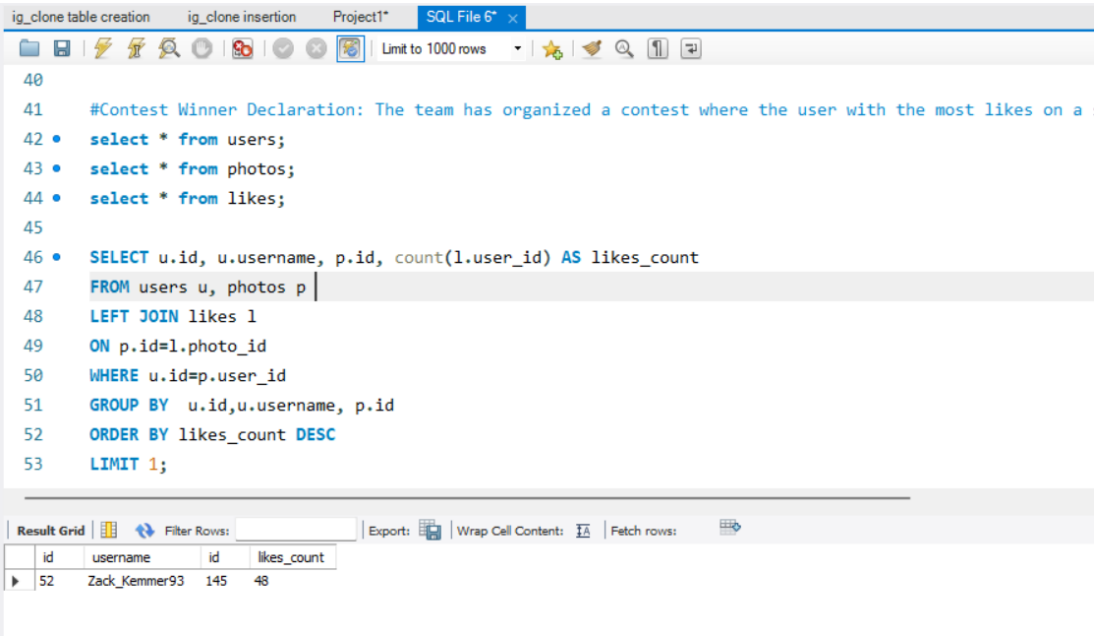
Explanation:

- "SELECT" query is used to select id, username from users table
- Here, "WHERE" statement is used on id from users table which can be accessed through users.id

- In the next query “SELECT” statement is used to select user_id from photos table
- Then, condition is checked where “users.id” is NOT IN “user_id from photos”
- This will give the id, username of users who never posted, as the photos table only consists of user_id who posted atleast one photo by checking NOT IN condition we are checking for id who are not in photos table

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

SQL Query:



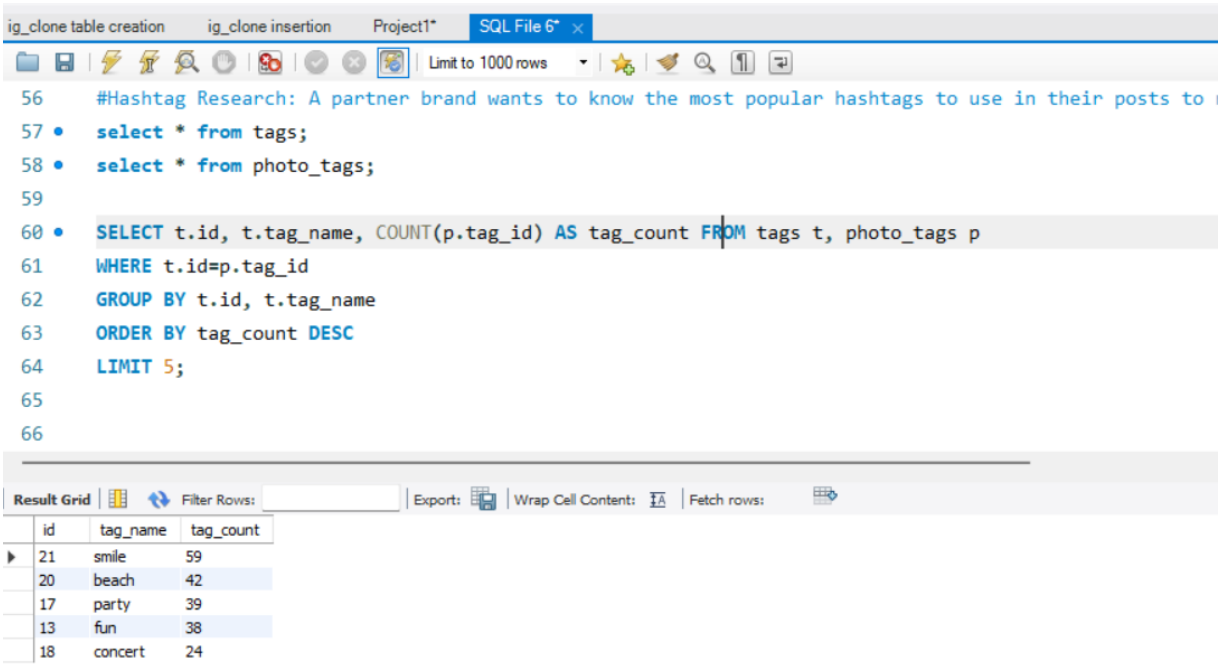
Explanation:

- To get the information of users who has most likes on a picture
- Firstly, we are selecting columns id, username from users table (which is taken as an alias ‘u’) and also we are select photo id from photos table (which is taken as an alias s ‘p’) and we are also counting the user.id from likes table (which has taken as an alias ‘l’)
- Then the likes table this joined with selected columns, on condition photos id = likes photo_id, another condition is also checked using WHERE statement that users. id = photos user_id
- GROUP BY statement is used to group the result using id and username from users table

- And ORDER BY statement is used to sort the result set by likes_count which has been calculated through first query
- Then LIMIT is set to 1 to get the photo with highest likes

4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
Your Task: Identify and suggest the top five most commonly used hashtags on the platform

SQL Query:



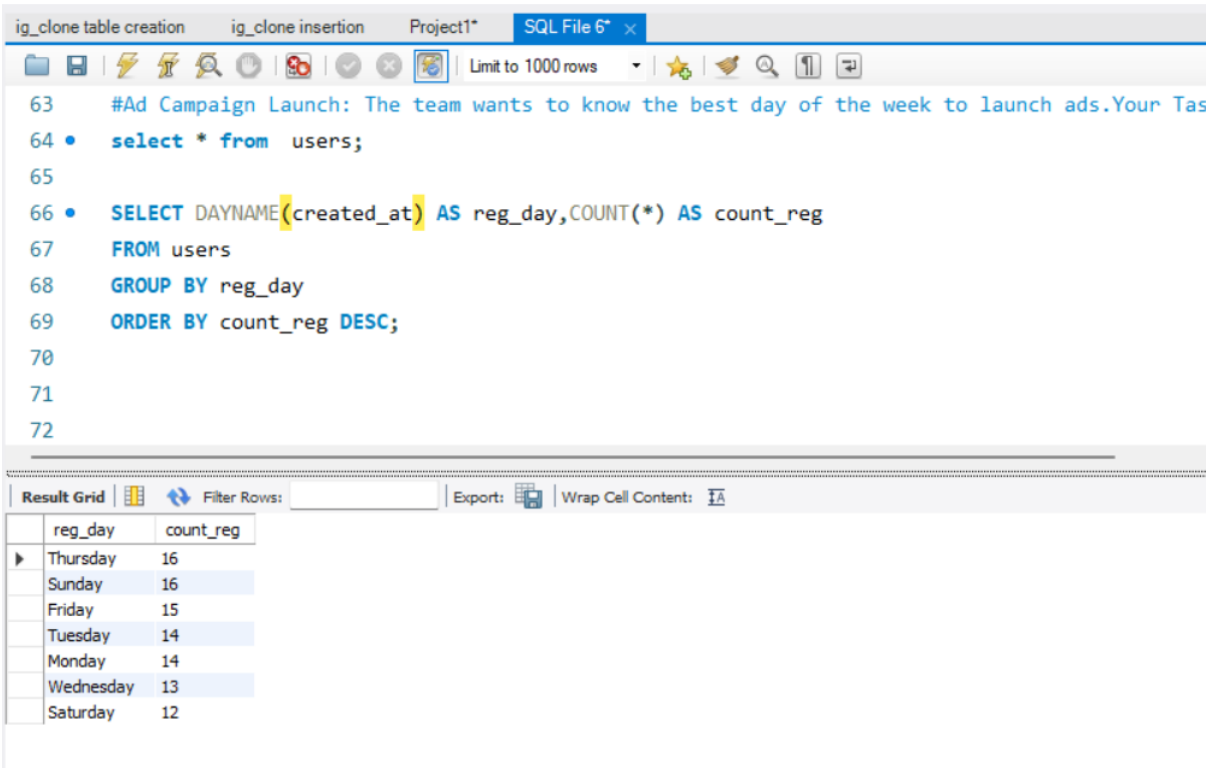
Explanation:

- To get the information about top 5 most used hashtags
- First, we'll get the tag id, tag name and count of the number of times tag is used using SELECT statement and tags, photo_tags table
- Then the rows are filtered on the condition id from tags table= tag_id from photo_tags table using WHERE statement
- Then the result is grouped on the basis of tag id and tag name
- And then ORDER BY statement is used on the count of tag and is set on DESC which sorts in descending order

- Then LIMIT statement is set 5 to get the top 5 most used tags

5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads. Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

SQL Query:



- To get the best day of the week to launch ads, we can get how many users have registered each day and decide the day with most users registered
- Here, 'DAYNAME' statement is used to get the name of the day from the given date which is given an alias reg_day and 'COUNT' statement is used to the number of users registered which is given an alias count_reg, And 'SELECT' statement is used to select both of these columns
- Then 'GROUP BY' is used to group the result based on the reg_day
- And 'ORDER BY' statement is used to sort the resultset based on count_reg descending order
- This will give us how many users have registered each day and now we can conclude that 'Thursday and Sunday' are the best days to launch the campaign

B)

1.User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
Your 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.

SQL Query:

```
--
86
87
88 • SELECT AVG(posts_per_user) FROM(
89   SELECT user_id , COUNT(id) AS posts_per_user FROM photos
90   GROUP BY user_id)AS inner_loop;
91
92
```

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

AVG(posts_per_user)
3.4730

```
76
77   #User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making
78
79 • select * from users;
80 • select * from photos;
81
82 • select (select count(id) from users) as total_users , (select count(id) from  photos) as total_posts,
83   (select count(id) from users)/(select count(id) from photos) as posts_by_user;
84
```

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

total_users	total_posts	posts_by_user
100	257	0.3891

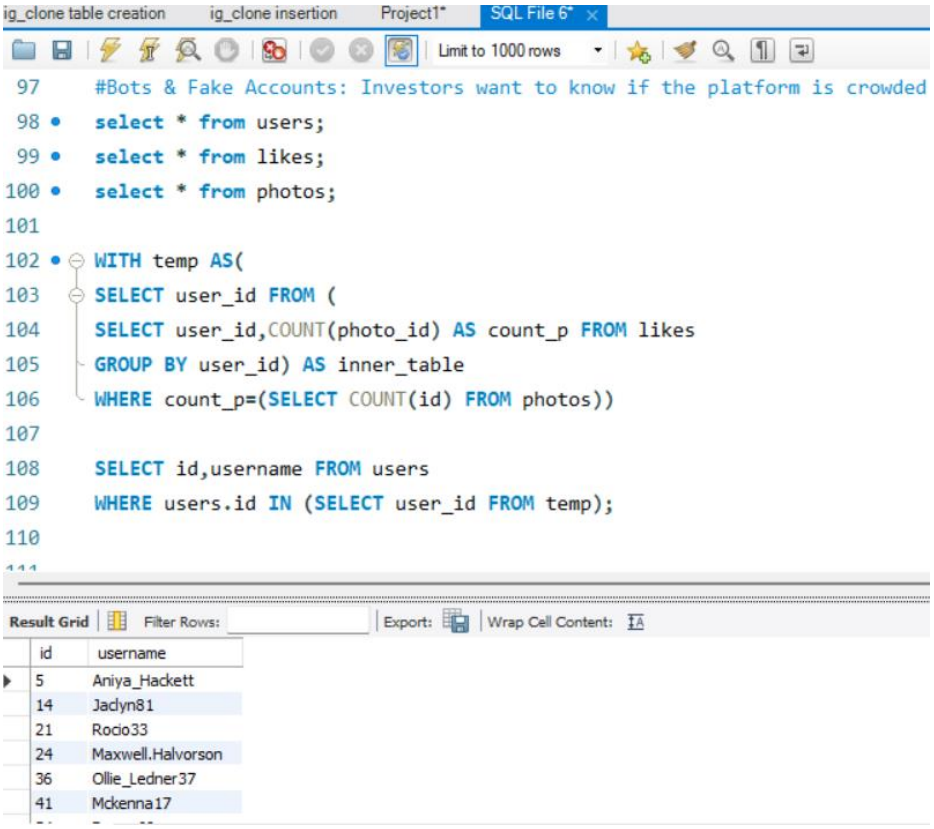
Explanation:

- To get the average posts per user, Firstly subquery is used where we are selecting user_id and count of the id from photos table using 'SELECT' statement which is then grouped by user_id and the table is given an alias inner_loop
- Then for the outer query average posts is calculated through the inner subquery count, this will provide the average number of posts per user

- Now for the second question which asks us to calculate total number of photos on Instagram divided by the total number of users.
- We solved this problem by counting the total number of users from users table as there could also be users who didn't post a single photo either, this is achieved using query "count(id)" from users table and alias is given as total_users
- Then to calculate the total number of posts, we can achieve this by using "count(id)" from photos table and alias is given as total_posts
- At last, we can divide the both results using '/' operator

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

SQL Query:



Result Grid | Filter Rows:

id	username
5	Aniya_Hackett
14	Jaclyn81
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

- To solve this problem, we can check for the users who liked every post
- So, first we'll calculate user_id, how many posts each user has liked using "COUNT" statement and alias is given as inner_table
- Then, we'll get user_id of users where count of liked posts is equal to the total numbers of posts on Instagram using 'COUNT' statement
- This will give us the user_id of the users who liked ever post and common table expression is created for it with name 'temp'
- Then the details of the user is taken where the user id matches the user_id from temp table
- This will give us all the user id and usernames of users who liked every single post on Instagram.

Tech-Stack-Used:

- MySQL Workbench was used as the main SQL analysis tool for this project.
- With its extensive SQL capabilities and easy-to-use interface, it was a great fit for managing the tasks listed in the project description.
- The program that was utilized was MySQL Workbench 8.0.

Insights:

A) Marketing Analysis:

Loyal User Reward:
By determining who the five oldest users are, the marketing team may honor and reward those who have consistently used Instagram.

Inactive User Engagement:
By identifying individuals who have never uploaded a photo, you can target this group with marketing communications to entice them to utilize the site more frequently.

Contest Winner Declaration:
Identifying and rewarding people who generate high interaction is made easier by basing the contest winner on the user who has the most likes on a particular photo.

Hashtag Research:
Finding the top five most popular hashtags might help a partner brand optimize their posts to get the greatest exposure and interaction.

Ad Campaign Launch:
The team may more effectively organize ad campaigns to reach a wider audience during peak registration periods by identifying the day of the week on which the majority of Instagram users register.

B) Investor Metrics:

User Engagement:
Determining the mean quantity of posts made by each user offers a measure for appraising

total user interaction. Furthermore, revealing the proportion of total photos to total users provides information on user activity levels.

Bots & Fake Accounts:

By locating people who have liked every single photo on the website, investors can detect possible bots and determine the legitimacy of user interaction, resulting in a more realistic portrayal of user activity.

Results:

- Learned how to use SELECT command to retrieve data from tables and WHERE clause that is used to filter data based on specified conditions.
- I learned how to group data using GROUP BY function and how to sort the data using ORDER BY function. And also learned other aggregated functions
- Understood how to join tables to combine data for analysis.