Project Description:

This project aims to analyse user interactions on Instagram to extract actionable insights that can drive business growth. The data analysis focuses on user engagement and behaviour, providing the product, marketing, and development teams with critical information to enhance the platform's features and user experience.

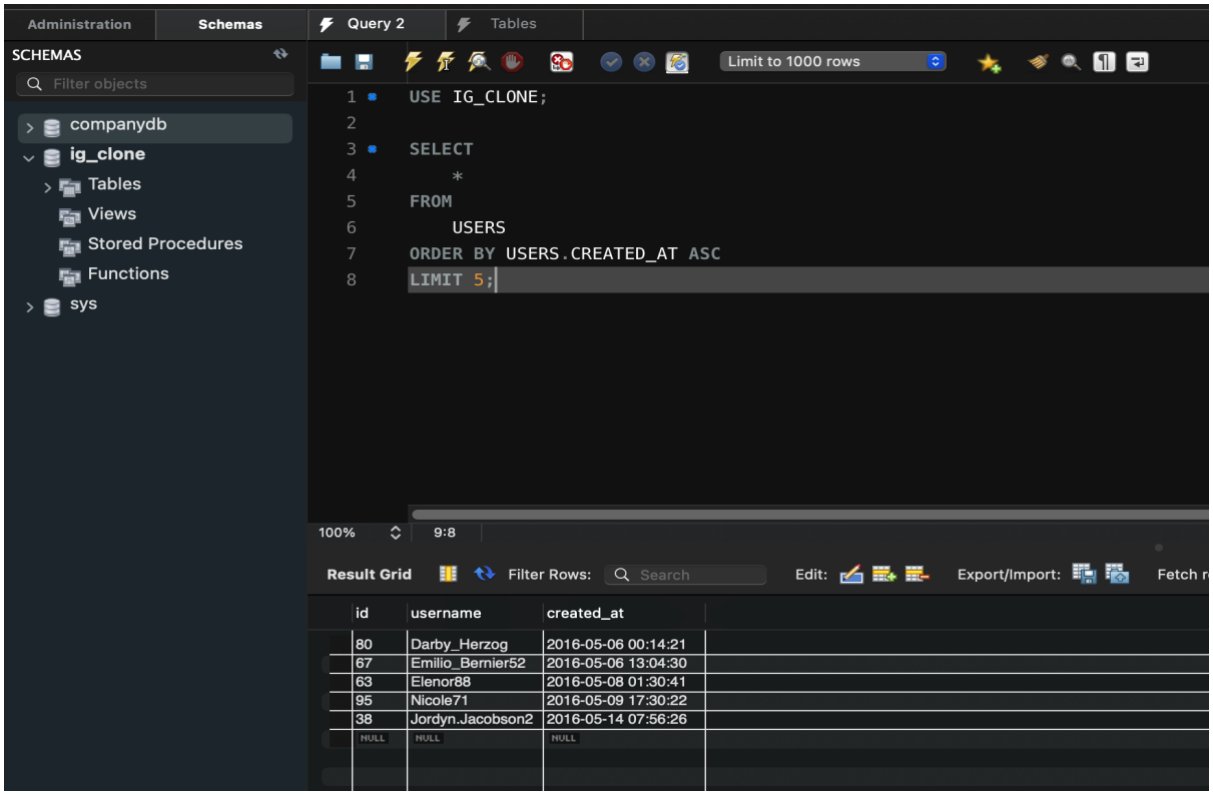
Approach:

I began by thoroughly understanding the questions posed by the management team. Using MySQL Workbench, I wrote SQL queries to retrieve the necessary data from the Instagram database. The approach involved:

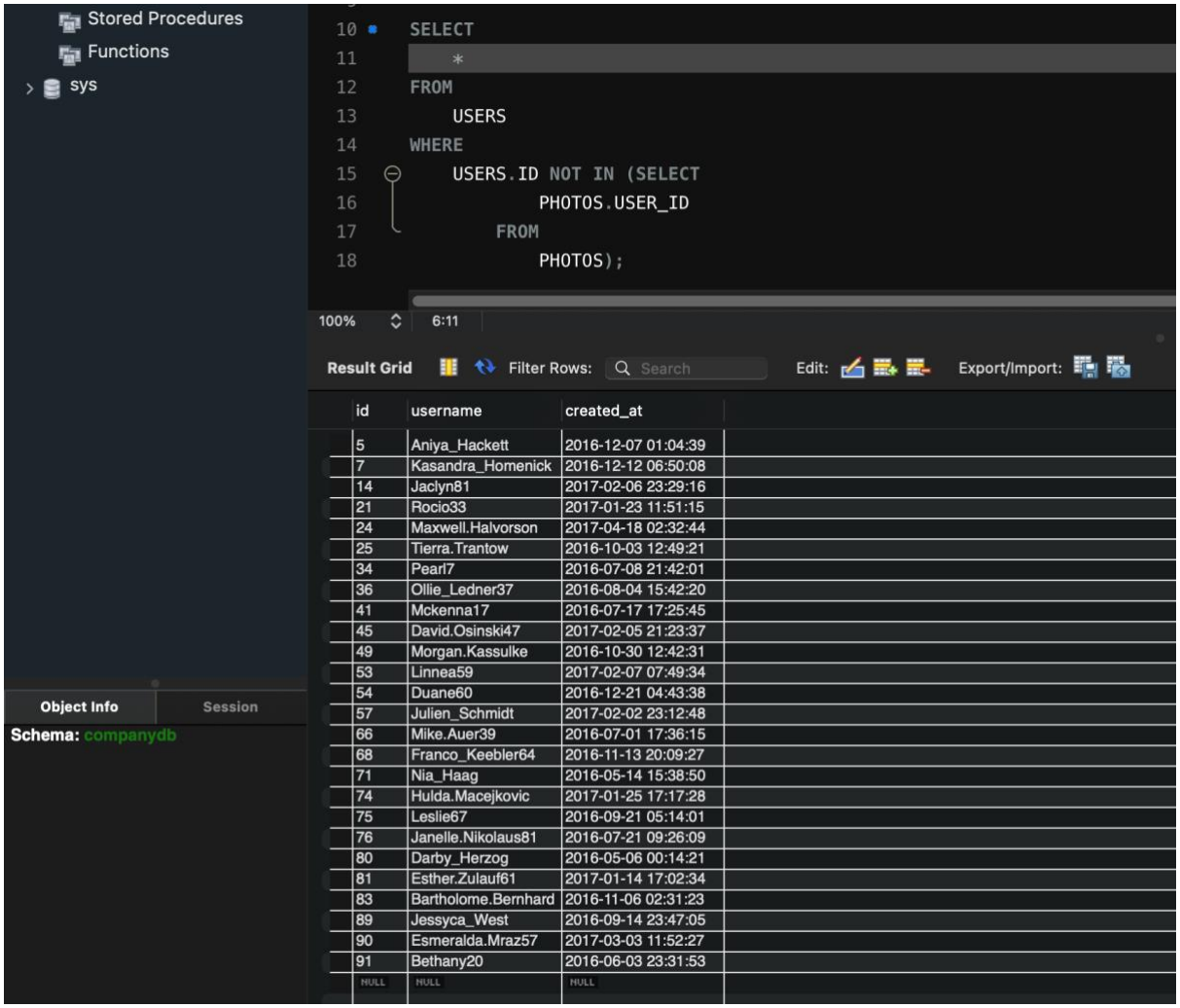
1. **Data Extraction:** Writing and executing SQL queries to extract relevant user data.
2. **Data Analysis:** Identifying patterns, such as the oldest users, users who never posted, the most popular hashtags, and the most active days on the platform.
3. **Insight Generation:** Summarizing the extracted data to answer the specific questions posed by the marketing and investor teams.

A) Marketing Analysis:

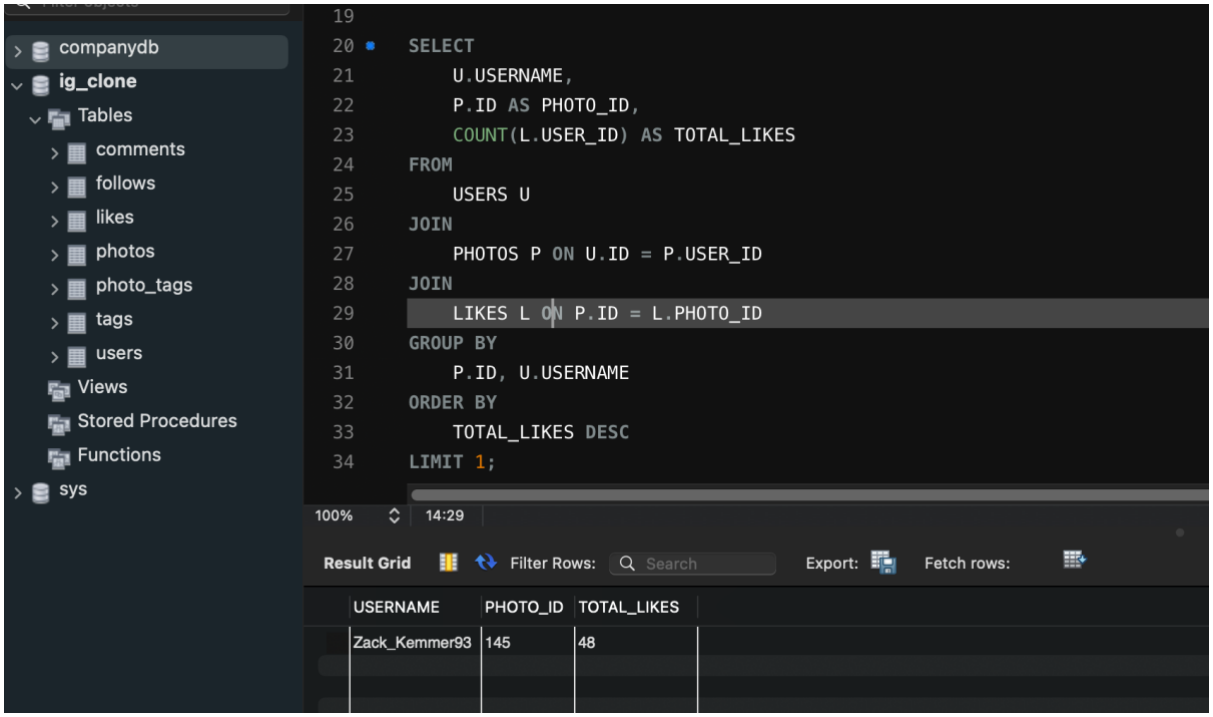
1. Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
   Your Task: Identify the five oldest users on Instagram from the provided database.
   - For this task I'll fetch data from users table in ascending order based on created_at field then I'll limit the row count to 5 as we only required 5 oldest user information.



2. Inactive User Engagement: 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.
   - For this task I will use 'NOT IN' operator with sub query. This will help me to fetch only those users whose userid is not present in photos table.
   - Alternative way would be to use left join this also fetch same data.(SELECT u.* FROM users u LEFT JOIN photos p ON u.id = p.user_id WHERE p.user_id IS NULL)



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.
   - I will be using multiple joins with users, photos and likes tables. To get total count on photo I will use “count” on user_id field present in likes table. Now as we need username, I will use group by to create group for username and photo_id. To get most liked photo I will use order by on total_likes and limit the result to 1with limit clause as we need only one winner.
   - Alternative approach would be to use with cluse to create temp table for count of likes for each phots, sort data in descending to total likes then limit the result to 1. Now join that temp table with users and photos table to get username and photo id. (WITH TOTAL_LIKES AS (SELECT COUNT(USER_ID) AS TOTAL_LIKES ,PHOTO_ID FROM LIKES GROUP BY PHOTO_ID ORDER BY TOTAL_LIKES DESC LIMIT 1) SELECT U.USERNAME,P.ID,L.TOTAL_LIKES FROM USERS U,PHOTOS P, TOTAL_LIKES L WHERE U.ID=P.USER_ID AND P.ID= L.PHOTO_ID;)



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.
   - For this task I will be using group by and count to get total count of tag_id for each tag_id then join photo_tag table with tag table to get the name of the tab. Order the total count in descending order limit the result to 5 to get most 5 used hashtags.

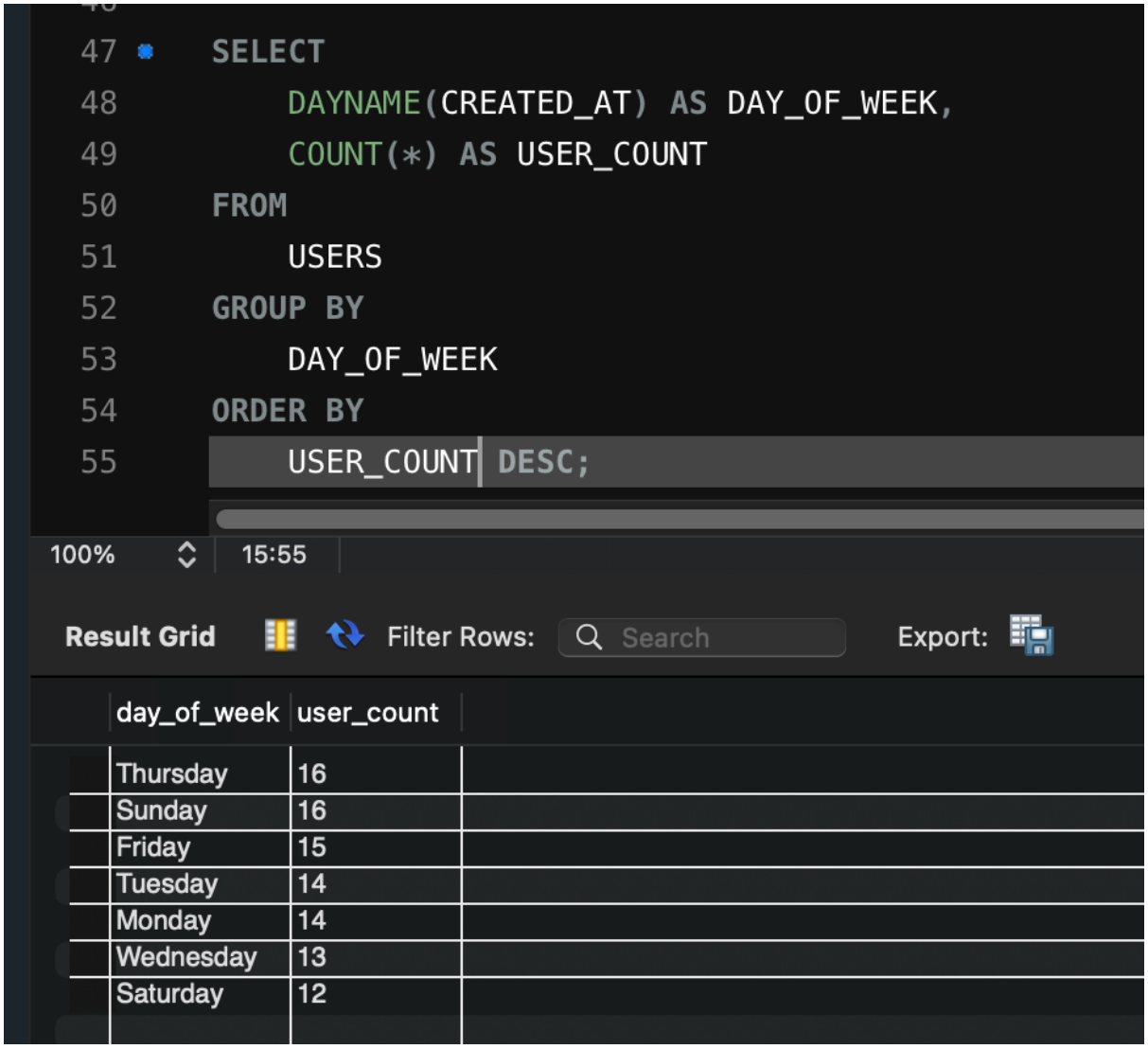
```
36  SELECT
37      tag_id, tag_name,COUNT(tag_id) total_count
38  FROM
39      photo_tags,
40      tags
41  WHERE
42      id = tag_id
43  GROUP BY tag_id
44  ORDER BY total_count DESC
45  LIMIT 5;
```

100% 6:41

Result Grid | Filter Rows: Search | Export: | Fet

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

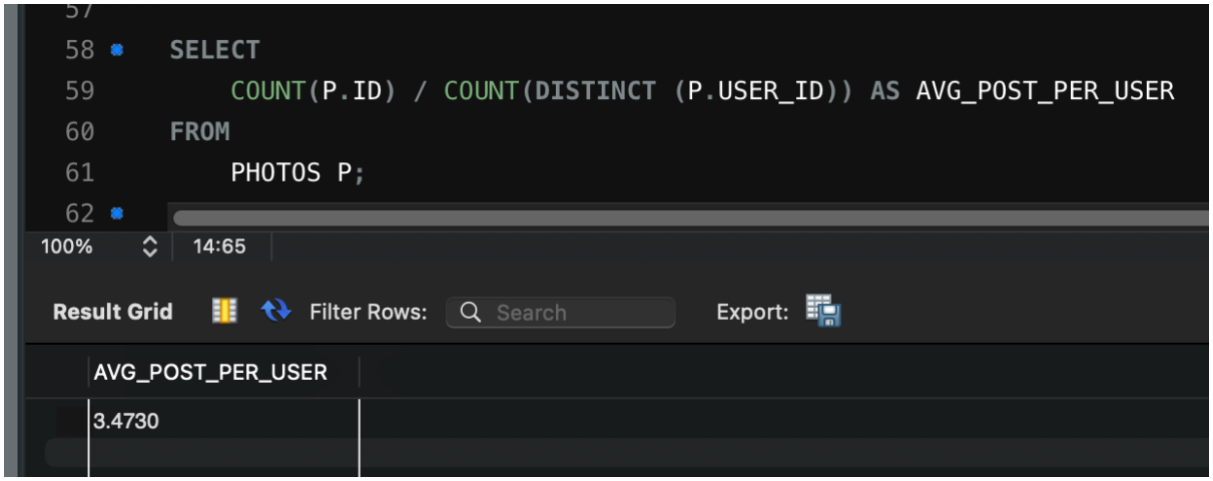
5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads.
   Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.
   - I will be using DAYNAME to extract the name of the day from created_at field as we required the name of the day, user count to sum up the number of users who created their account on that day. Use order by to sort the count in descending order.

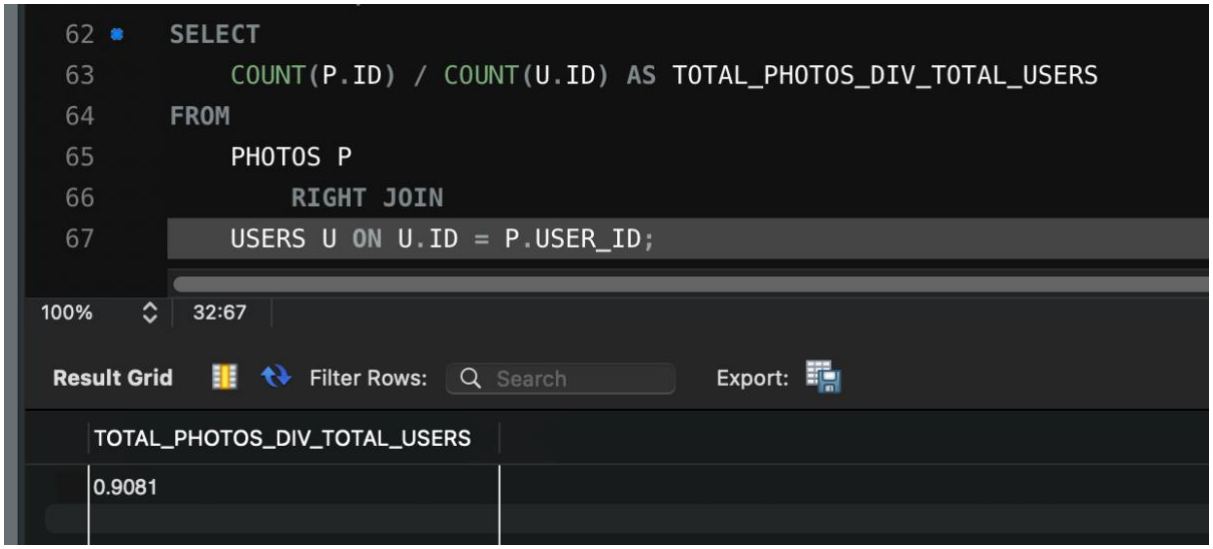


B) Investor Metrics:

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.

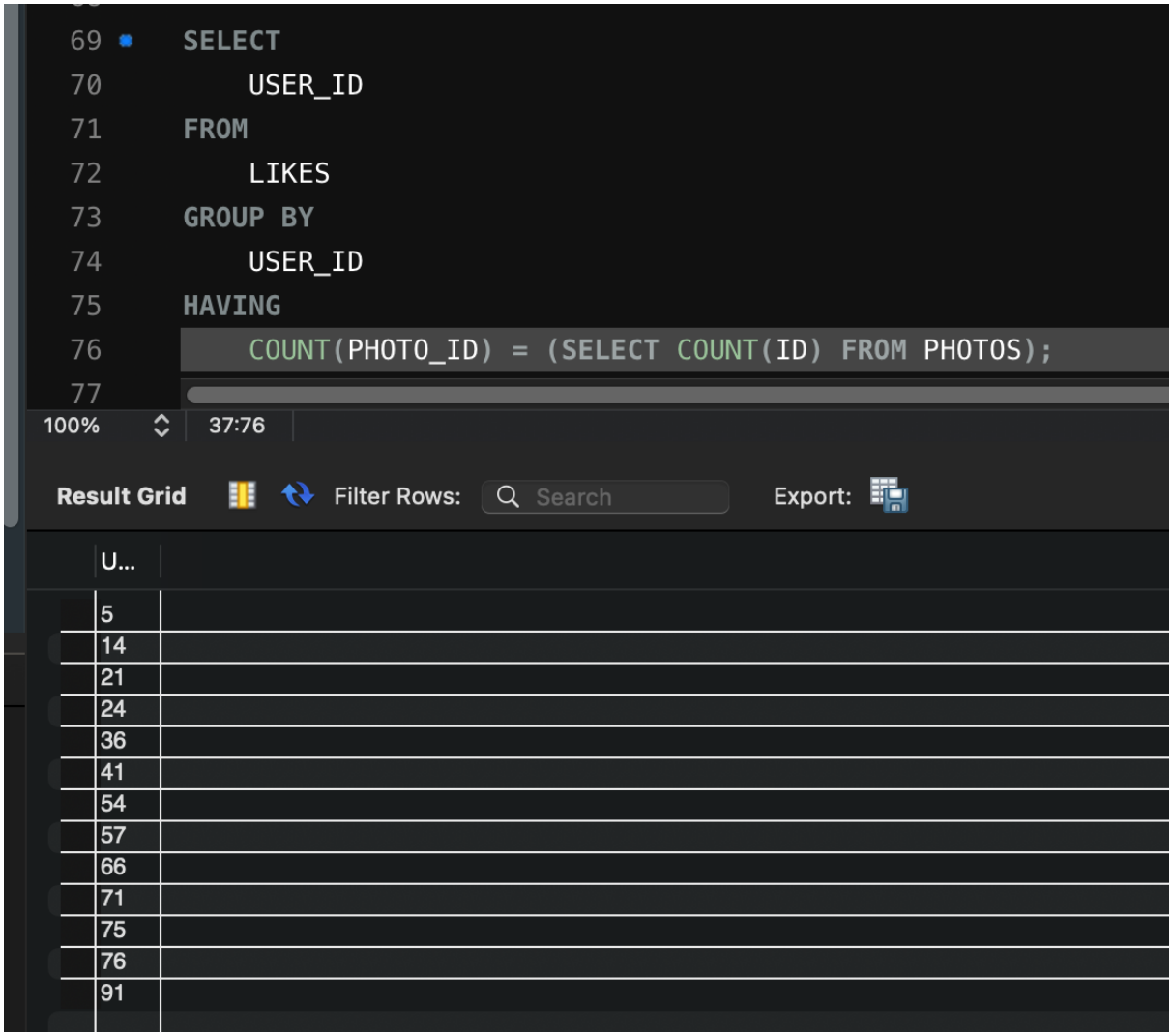
   - For first task I have divided total number of phots to total distinct users who posted on the Instagram which will give the average number of posts by user.
   - For second task I have to divide total number of post with total number of user present on Instagram irrespective of their posting status for that I used right join to include those users as well from users table who has not posted any photos.





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.

   - For this task I have used group by and having clause to filter out the user who has likes all the photos present in photos table. As I am using group by clause I cannot use where condition to filer the data, so I have used having clause for filtration.



Tech-Stack Used:

- **MySQL Workbench** (version 9.0.1): Chosen for its robust database management and data visualization features, essential for executing complex SQL queries and analysing large datasets.
- **SQL:** Utilized to perform data extraction and manipulation, allowing for precise querying and efficient handling of Instagram's user data.

Insights:

- **User Loyalty:** Identified the five oldest users on the platform, providing a basis for rewarding loyalty.
- **Engagement Patterns:** Found inactive users for targeted re-engagement campaigns and recognized the most popular hashtags to guide brand partnerships.
- **Contest Winner Declaration:** Zack_Kemmer93 is the contest winner with 48 total like on his post.
- **Hashtag Research:** Five most common hashtags are smile, beach, party, fun and concert.
- **Optimal Ad Timing:** Determined the most active days for user registrations, helping to optimize ad campaign launches. Here Thursday and Sunday would be the best of the week for new AD launch.
- **User Engagement:** From the above output we can see that on an average user is posting 3-4 phots and around 10% of users have not posted anything yet.
- **Bots Detection:** Detected potential bots by identifying users who liked every photo, addressing concerns about fake accounts.

Result:

The analysis provided the product team with clear insights that informed decisions on user engagement strategies, feature development, and marketing campaigns. The findings helped in identifying loyal users, optimizing ad spend, and improving user experience on Instagram, contributing to the platform's overall growth.