Project Description

This Project involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

As a data analyst, first I will understand the project requirements and purpose and then I will understand the data which is provided in the project attachment and will perform analysis to get the meaningful insights.

Analysis: involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business.
For example, the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

Approach

1. **Understanding the Database Schema**: The first step was to understand the structure of the database and created the database ig_clone. The attachment provided the schema for several tables,
   including users, photos, comments, likes, follows, tags, and photo_tags.
   These tables represent different entities in a social media application and their relationships.

2. **Formulating SQL Queries**: Based on the questions, I formulated appropriate SQL queries to retrieve the required information from these tables. This involved using various SQL clauses such
   as SELECT, FROM, WHERE, GROUP BY, ORDER BY, and LIMIT.

3. **Analyzing Data**: Once I had the SQL queries, I executed the queries to retrieve the data and added additional functions to my SQL queries like SUM(), AVG(), COUNT() to get the aggregated and filtered data.

4. **Interpreting Results**: After retrieving the data, I interpreted the results to answer your questions. For example, to find out if users are still active and posting on Instagram, I suggested comparing the count of photos they have posted in different time periods.

5. **Providing Insights**: Based on the results of the data analysis, I provided insights that could be useful for decision-making. For example, I suggested scheduling an ad campaign on the day when most users register on Instagram to maximize your ad’s visibility.

Throughout this process, my goal was to provide clear, accurate, and helpful insights and answers to questions posed by management team.

Tech-Stack Used

The software and versions you used for the project:

MySQL Workbench 8.0 CE – To write SQL queries to analyze Instagram user data and answer questions posed by the management team

MYSQL Server 8.0 – To create and store database

Insights:

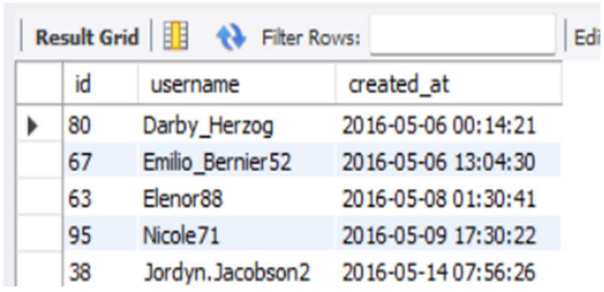
I will provide my insights in the form of answers for the questions posed by management team

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

SQL Query:

```
64 •  SELECT
65        *
66    FROM
67        users
68    ORDER BY created_at ASC
69    LIMIT 5;
```

Output: Below is the result for above query which shows 5 oldest users on Instagram along with the details like user id and date they created account on Instagram.



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

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.

SQL Query:

```
72 •  SELECT
73        username
74    FROM
75        users u
76            LEFT JOIN
77        photos p ON u.id = p.user_id
78    WHERE
79        p.user_id IS NULL;
```

Output:

Above query give the names of users who never posted single photo on Instagram in the same query if use COUNT() we will get number of users who never posted even a single photo and the result is 26.

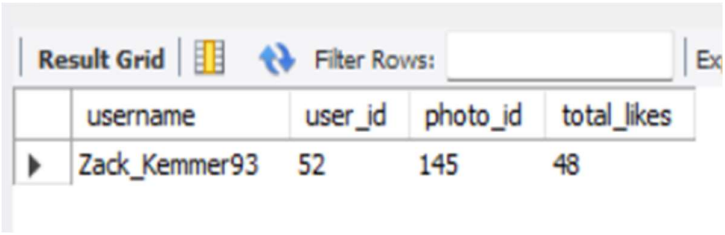
Aniya_Hackett	Tierra.Trantow	Morgan.Kassulke	Franco_Keebler64	Darby_Herzog
Kasandra_Homenick	Pearl7	Linnea59	Nia_Haag	Esther.Zulauf61
Jaclyn81	Ollie_Ledner37	Duane60	Hulda.Macejkovic	Bartholome.Bernhard
Rocio33	Mckenna17	Julien_Schmidt	Leslie67	Jessyca_West
Maxwell.Halvorson	David.Osinski47	Mike.Auer39	Janelle.Nikolaus81	Esmeralda.Mraz57
				Bethany20

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:

```
84 •  SELECT users.username, photos.user_id, likes.photo_id, COUNT(*) as total_likes
85    FROM likes
86    JOIN photos ON likes.photo_id = photos.id
87    JOIN users ON photos.user_id = users.id
88    GROUP BY photos.user_id, likes.photo_id
89    ORDER BY total_likes DESC
90    LIMIT 1;
```

Output:



Result Grid | Filter Rows: | Ex

username	user_id	photo_id	total_likes
Zack_Kemmer93	52	145	48

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:

```
94 •  SELECT
95        tags.tag_name,
96        photo_tags.tag_id,
97        COUNT(*) AS value_occurance
98    FROM
99        photo_tags
100           JOIN
101       tags ON photo_tags.tag_id = tags.id
102   GROUP BY photo_tags.tag_id
103   ORDER BY value_occurance DESC
104   LIMIT 5;
```

Output: above query gives 5 top most tags commonly used

Result Grid | Filter Rows:

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

SQL Query:

```
108 •   SELECT DAYNAME(created_at) as day_of_week, COUNT(*) as total_users
109     FROM users
110     GROUP BY day_of_week
111     ORDER BY total_users DESC
112     ;
```

Output: As per below result we can see most of the users register on Thursday and Sunday. SQL listed Thursday on top of the list but Sunday being the weekend and has number of registrations equal to Thursday I am considering Sunday is the day on which team has to launch ads.

Result Grid | Filter Rows:

day_of_week	total_users
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

6. 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.*/

/*Calculate the average number of posts per user on Instagram*/

SQL Query:

```
117 •   SELECT user_id, COUNT(*) as total_posts
118     FROM photos
119     GROUP BY user_id
120     ORDER BY total_posts DESC;
```

Output: This query will list all the users and the average number posts posted by them

/*provide the total number of photos on Instagram divided by the total number of users.*/

SQL Query:

```
123 •   SELECT AVG(post_count) as average_posts_per_user
124     FROM (
125         SELECT user_id, COUNT(*) as post_count
126         FROM photos
127         GROUP BY user_id
128     ) as user_posts;
```

Output:

Result Grid | Filter Rows:

	average_posts_per_user
▶	3.4730

7. 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:

```
132 •   SELECT users.username, user_id
133     FROM likes join users on likes.user_id = id
134     GROUP BY user_id
135     HAVING COUNT(DISTINCT photo_id) = (SELECT COUNT(*) FROM photos);
136
```

Output: Above query gives the list of users who liked every single photo

username	user_id	username	user_id
Aniya_Hac	5	Duane60	54
Jaclyn81	14	Julien_Sch	57
Rocio33	21	Mike.Auer	66
Maxwell.H	24	Nia_Haag	71
Ollie_Ledn	36	Leslie67	75
Mckenna1	41	Janelle.Nik	76
		Bethany20	91

Result

This project definitely helps me rebrushing my SQL skills and I regained my confidence on using SQL for data analytics. I was relying mostly on Excel and Power BI for data analysis but with this project I feel SQL is easier to use.

When it comes to this project based on my analysis I can suggest to management team that we need launch more ads on Sundays in a week as most of the users are registering on Sundays in a week

And

In question number 6 on user engagement, it is found that the average posts posted per user is around 3- 4 which appears less but to find or analyse user engagement accurately I suggest to perform analysis to to compare the number of photos posted by each user in the last month with the number of photos posted in the previous month

Drive Link

Link for the SQL query file:

https://drive.google.com/file/d/1nfB07mnB77UtFwXyu8UGf5jpX4sgRD_3/view?usp=sharing