

PROJECT DESCRIPTION:

My project is about User analysis on Instagram to track how users engage and interact with our digital product (software or mobile application). I was given a database and I performed several SQL queries to see the trends and popularity of our product among users. I must provide insights for the marketing team and investors so that they can make future decisions regarding our product.

APPROACH:

Firstly, I analyzed their problem, performed relevant SQL queries, and got the desired results.

For this, we should understand MySQL and its function.

TECH-STACK USED

- MySQL Workbench 8.0 CE

INSIGHTS:

Working on this project helped me to clarify my logic regarding SQL queries. I got a deep knowledge of aggregate functions, Joins, SQL operators and sorting function in it. These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build In an app, track the success of the app by measuring user engagement and improve the user experience altogether while helping the business grow.

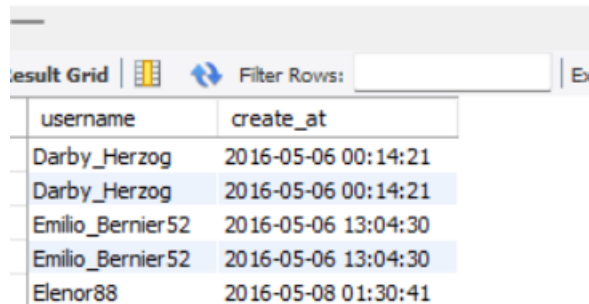
MARKETING ANALYSIS

The marketing team wants to launch a few campaigns and needs a better understanding of users and a few insights on how to increase user engagement for a better user experience.

1)**rewarding most loyal users:** I have found the five oldest users that have been using this platform for a very long time.

here we ordered the user table in ascending order and got 5 oldest users.

```
84    ## A) MARKETING ANALYSIS
85    #LOYAL USER REWARD
86    #five oldest users
87    • select username, create_at
88    FROM users
89    ORDER BY create_at ASC limit 5;
```



The screenshot shows a database query result grid with two columns: 'username' and 'create_at'. The results are ordered by 'create_at' in ascending order, showing the five oldest users. The first two rows are for 'Darby_Herzog' with a creation time of '2016-05-06 00:14:21'. The next two rows are for 'Emilio_Bernier52' with a creation time of '2016-05-06 13:04:30'. The final row is for 'Elenor88' with a creation time of '2016-05-08 01:30:41'.

username	create_at
Darby_Herzog	2016-05-06 00:14:21
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41

FIVE MOST LOYAL /OLDEST USERS ARE:

Darby_Herzog

Darby_Herzog

Emilio_Bernier52

Emilio_Bernier52

Elenor88

2) **Remind inactive users to start posting**: The team wants to encourage inactive users to start posting by sending them promotional emails.

To find the inactive user, I must see that user id whose photo id is NULL.

#INACTIVE USERS ENGAGEMENT

- ```
select username
FROM users
LEFT JOIN photos ON users.id=photos.user_id
WHERE photos.id IS NULL;
```

#### **# INACTIVE USERS LIST**

5-Aniya\_Hackett

7-Kasandra\_Homenick

14-Jaclyn81

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

89-Jessyca\_West

90-Esmeralda.Mraz57

91-Bethany20

#### **3.) CONTEST WINNER DECLARATION:**

The team has organized a contest where the user with the most likes on a single photo wins.

```
97 #CONTEST WINNER
98 • with base as(
99 select likes.photo_id, users.username,
100 COUNT(likes.user_id)
101 AS like_user
102 FROM ig_clone.likes likes
103 INNER JOIN Ig_clone.photos photos ON likes.photo_id= photos.id
104 INNER JOIN ig_clone.users users ON photos.user_id= users.id
105 GROUP BY likes.photo_id, users.username
106 ORDER BY like_user desc
107 LIMIT 1)
```

| Result Grid |               | Filter Rows: | Export: | Wrap Cell Content: |
|-------------|---------------|--------------|---------|--------------------|
| username    |               |              |         |                    |
|             | Zack_Kemmer93 |              |         |                    |

# Zack kemmer93 is the contest winner

**4.) HASHTAG RESEARCH:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

#Most popular hashtags are:

```
110 #HASHTAG RESEARCH
111 • select t.tag_name,
112 COUNT(p.photo_id)
113 AS num_tags
114 from ig_clone.photos_tags p
115 INNER JOIN ig_clone.tags t
116 ON p.tag_id= t.id
117 GROUP BY tag_name
118 ORDER BY num_tags DESC
119 LIMIT 5;
```

| Result Grid |          | Filter Rows: |
|-------------|----------|--------------|
| tag_name    | num_tags |              |
| smile       | 59       |              |
| beach       | 42       |              |
| party       | 39       |              |
| fun         | 38       |              |
| concert     | 24       |              |

Result 7 x

# 'SMILE' Is the most popular hashtag

**5.) AD CAMPAIGN LAUNCH:** The team wants to know the best day of the week to launch ads.

```
123 #AD CAMPAIGN LAUNCH
124 • Select dayname(create_at)
125 AS day,
126 COUNT(*) AS total
127 FROM users
128 GROUP BY day
129 ORDER BY TOTAL desc;
```

| Result Grid |       | Filter Rows: |
|-------------|-------|--------------|
| day         | total |              |
| Thursday    | 32    |              |
| Sunday      | 32    |              |
| Friday      | 30    |              |
| Tuesday     | 28    |              |
| Monday      | 28    |              |

# The best day to launch the campaign is either Thursday or Sunday.

### **INVESTORS METRICS:**

1) **USER ENGAGEMENT:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

```
136 • SELECT (select COUNT(*)
137 FROM photos) /
138 (select COUNT(*)
139 FROM users)
140 as avg;
```

| Result Grid |  | Filter Rows: |
|-------------|--|--------------|
| avg         |  |              |
| 1.2850      |  |              |

# The average number of posts per user on Instagram is 1.2850

- 2) **BOTS AND FAKE ACCOUNTS**: Investors want to know if the platform is crowded with fake and dummy accounts.

**#IDENTIFYING USERS THAT HAVE LIKED ALL THE PHOTOS ON THE APP:**

```
#B) BOTS AND FAKE ACCOUNTS
FINDING ACCOUNTS WHO HAVE LIKED EVERY SINGLE PICTURE ON THE APP

#first, calculating total number of photos on instagram
• SELECT COUNT(*)
 FROM ig_clone.photos;

#there are 257 total photos on instagram

#now finding accounts that have liked every single one of these 257 photos
• WITH photo_count AS(
 SELECT user_id,
 COUNT(photo_id) AS num_like
 FROM ig_clone.likes
 GROUP BY
 user_id
 ORDER BY
 NUM_LIKE DESC
)
SELECT *
FROM photo_count
WHERE num_like= (SELECT COUNT(*) FROM ig_clone.photos);
```

**#Following are the user IDs of BOTS and fake accounts:**

|   | user_id | num_like |
|---|---------|----------|
| ▶ | 21      | 257      |
|   | 71      | 257      |
|   | 5       | 257      |
|   | 66      | 257      |
|   | 41      | 257      |
|   | 14      | 257      |
|   | 57      | 257      |
|   | 24      | 257      |
|   | 76      | 257      |
|   | 75      | 257      |
|   | 54      | 257      |
|   | 91      | 257      |
|   | 36      | 257      |

**APPROACH:** For this project, I have used My SQL to extract the required data from the given database using the Join function, subqueries, Aggregation, where condition, Group by, Distinct and other functions required.

keeping the Primary key and foreign key in consideration provided all the reports asked by the marketing department and Investor metrics department.

### **RESULTS AND INSIGHTS:**

From this project, I got an idea about how as a business or data analyst we work on real-time data to make any data-driven decision.

One thing I infer about this project is the dataset provided was very limited and small with respect to Rows and columns, But still, it was a very good experience working on such kind of project.

It helped me a lot to understand the analysis process well and to provide insights for the best decision possible.