

Instagram User Analytics



Description

The project is about conducting user analysis on Instagram to provide insights for marketing, product, and development teams. The project aims to answer specific questions related to marketing and investor metrics.

The project will use the provided database to gather and analyze the required data to provide a detailed report to product team.



Approach : Since, it is complex to jot down the approach and insights of the whole project at once. Let us go through each question to be answered for the respective teams. Concluding the report with a dashboard of the project and knowledge gained.



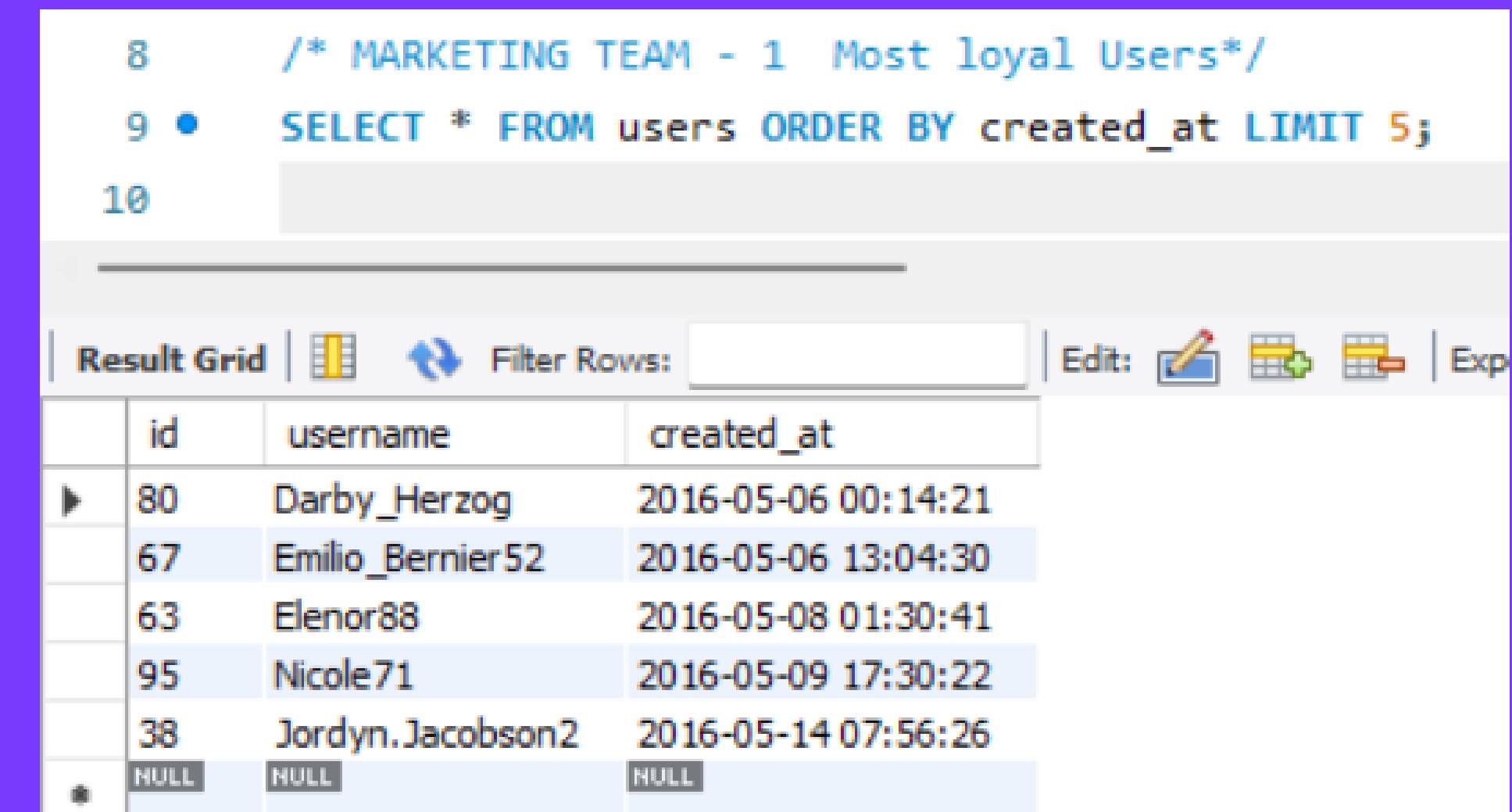
Tech-Stack Used: MySQL 8.0 CE



The questions are defined clearly and data provided is clean. Hence, the project starts with analysis addressing the problem questions.

For Marketing Team, 5 oldest users -

- To find the top 5 loyal users, I queried the 'Users' database by sorting the table using the date accounts were created and limit the result set to 5 oldest users.
- Beside picture shows the result set of the 1st question by Marketing team to find the users who have been using the app for longest time.



The screenshot shows a MySQL query results interface. At the top, there is a code editor window containing the following SQL query:

```
8  /* MARKETING TEAM - 1 Most loyal Users*/
9 • SELECT * FROM users ORDER BY created_at LIMIT 5;
10
```

Below the code editor is a results grid. The grid has three columns: 'id', 'username', and 'created_at'. The data is as follows:

	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
	HULL	HULL	HULL

Inactive users (remind to post for the first time) -

- Using subquery, got all the users who have posted atleast once. Then, by WHERE clause, the data is filtered to only give the users in result set which are not listed in the previous subquery.

```
6  /* MARKETING TEAM - ##2## Remind Inactive Users to start posting */
7 •  SELECT username, id FROM users
8 WHERE
9   id NOT IN (SELECT DISTINCT
10    user_id
11   FROM
12    photos);
```

The report of the above query is presented in the dashboard at the end.

Contest winner (with most likes)

```
9 •   SELECT
10    users.id AS user_id,
11    username,
12    image_url,
13    photos.id AS photo_id
14  FROM
15    users
16      LEFT JOIN
17    photos ON photos.user_id = users.id
18  WHERE
19    photos.id = (SELECT
20      photo_id
21    FROM
22      likes
23    GROUP BY photo_id
24    ORDER BY COUNT(photo_id) DESC
25    LIMIT 1);
```

- Here, the query uses 3 tables to get the output. First, the photo with most likes is to be found. The subquery does that by giving the photo id with most likes.
- In the main query, I tried to filter the photos table by subquery result and joined the users. This provides the user details of the photo id.

	user_id	username	image_url	photo_id
▶	52	Zack_Kemmer93	https://jarret.name	145

Top 5 commonly used Hashtags

```
10  /* MARKETING TEAM - ## 4 ## Hashtag Researching */
11 • SELECT
12     id, tag_name, COUNT(tag_id) as HashTag_count
13  FROM
14      tags
15      JOIN
16          photo_tags ON tag_id = tags.id
17  GROUP BY tag_id
18  ORDER BY COUNT(tag_id) DESC
19  LIMIT 5;
```

	id	tag_name	HashTag_count
▶	21	smile	59
	20	beach	42
	17	party	39
	13	fun	38
	18	concert	24

- Initially, joined the 2 tables and grouped the total working set by tag ID using aggregate function.
- The above set is then sorted using the aggregate function to get the top most used hashtags.

Best Day to Launch AD Campaign

- This is a simple query with SQL server date function i.e. DAYNAME.
- I got the day from the date when the account was created and grouped.
- Now, sorted the data set using the aggregate function to find the number of accounts created on the days.

The image shows a screenshot of a SQL query and its results. On the left, a code editor displays a SQL script. On the right, a table shows the results of the query.

```
9  /* MARKETING TEAM - ##5## Launch AD Campaign */  
10 • SELECT  
11     DAYNAME(created_at) day_name, COUNT(created_at) Frequency  
12 FROM  
13     users  
14 GROUP BY day_name  
15 ORDER BY Frequency DESC;
```

day_name	Frequency
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

For investor metrics, user engagement (number of times average user posts) -

- Average posts a user posts can be obtained by considering all the users and the sum of all the posts each user make.
- Hence, we found the average by dividing one by the other. This is equivalent to total posts/total users.

```
9      /*  B) INVESTOR METRICS    */
10     /*    ##1## User Engagement   */
11 •  SELECT
12       COUNT(DISTINCT image_url) / COUNT(DISTINCT users.id) As avg_posts_per_user
13   FROM
14     photos JOIN users;
```

Bots & Fake Accounts

```
9   /*      ##2## Bots & Fake Accounts      */
10 • SELECT
11   *
12   FROM
13   USERS
14 WHERE
15   id IN (SELECT
16     user_id
17   FROM
18     Likes
19   GROUP BY user_id
20   HAVING COUNT(user_id) = (SELECT
21     COUNT(id)
22   FROM
23     PHOTOS));
```

- The result set is in the next slide.
- Here, I tried to follow the same approach which I used in the 2nd question.
- I tried to filter the data set by the subquery result set. Since, the count of each user ID appearing in likes table will be equal the total number of photos. (i.e one cannot like a single photo more than once).

Instagram

Instagram User Analytics

Bots & Fake Accounts

User Name	ID
Aniya_Hackett	5
Bethany20	91
Duane60	54
Jaclyn81	14
Janelle.Nikolaus81	76
Julien_Schmidt	57
Leslie67	75
Maxwell.Halvorson	24
Mckenna17	41
Mike.Auer39	66
Nia_Haag	71
Ollie_Ledner37	36
Rocio33	21

Top 5 Loyal Users

User Name	ID
Darby_Herzog	80
Elenor88	63
Emilio_Bernier52	67
Jordyn.Jacobson2	38
Nicole71	95

2.57

Average User Posts

257

Total no. of photos

100

Total number of users

Zack_Kemmer93

Contest Winner

52

User ID

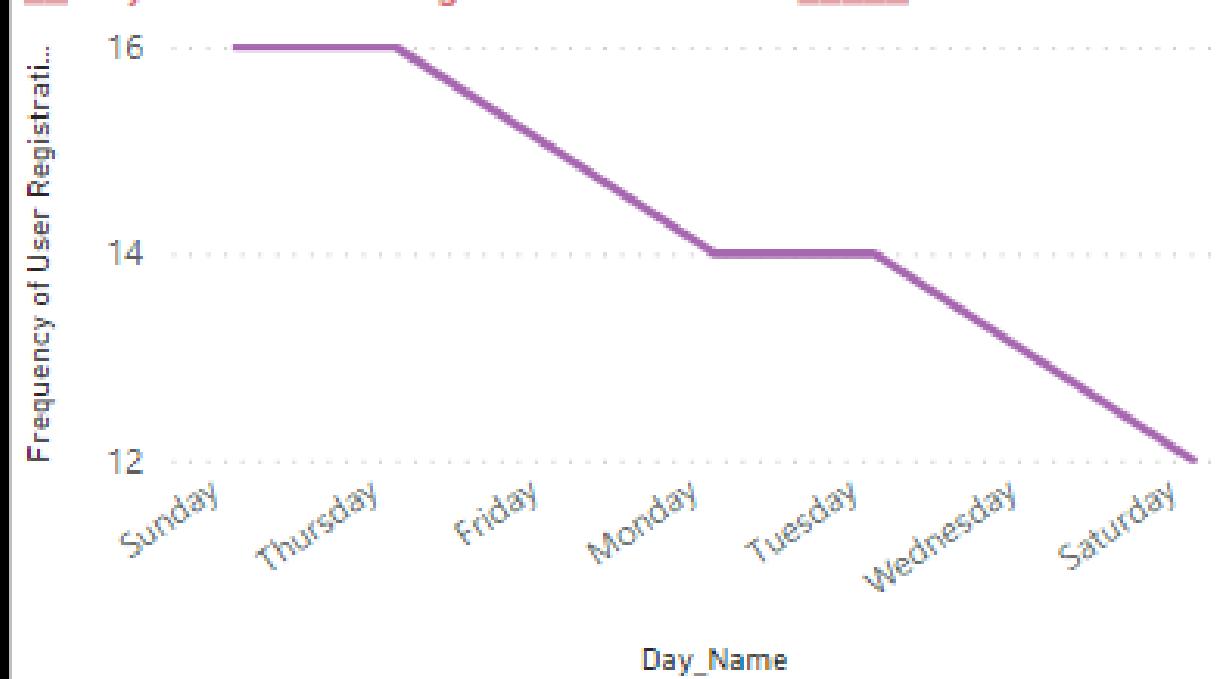
145

Photo ID

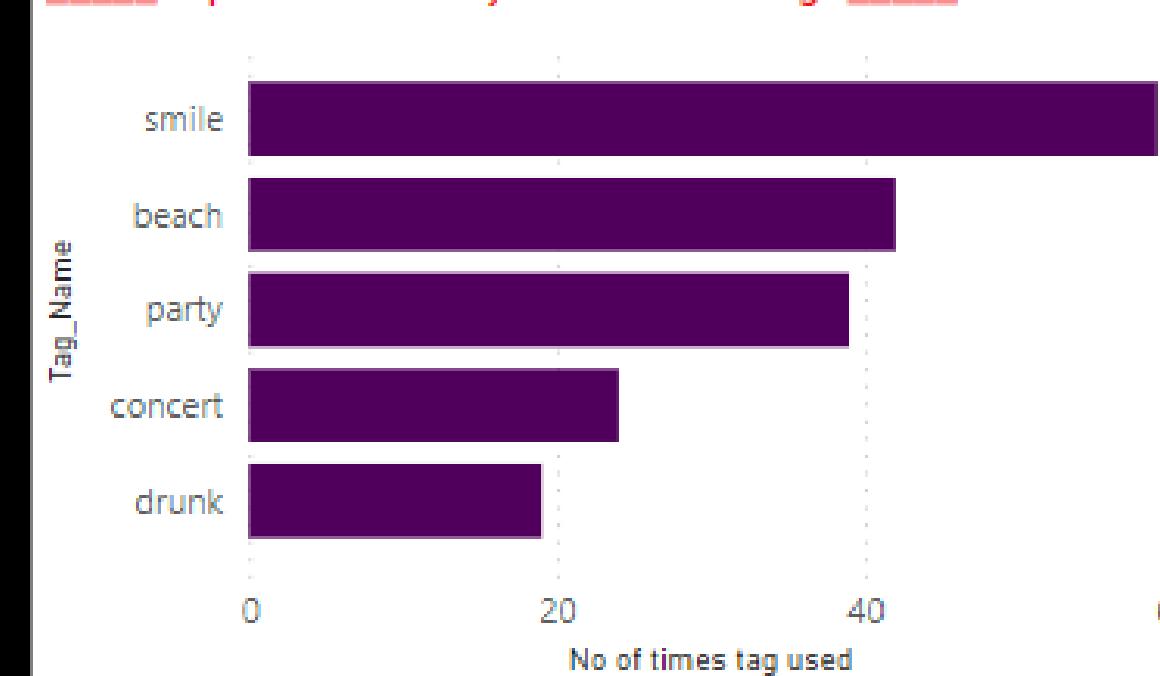
Inactive Users

User Name	ID
Aniya_Hackett	5
Kasandra_Homenick	7
Jaclyn81	14
Rocio33	21
Maxwell.Halvorson	24
Tierra.Trantow	25
Pearl7	34
Ollie_Ledner37	36
Mckenna17	41
David.Osinski47	45
Morgan.Kassulke	49
Linnea59	53
Duane60	54
Julien_Schmidt	57
Mike.Auer39	66
Franco_Keebler64	68
Nia_Haag	71
Hulda.Macejkovic	74
Leslie67	75
Janelle.Nikolaus81	76
Darby_Herzog	80
Esther.Zulauf61	81
Bartholome.Bernhard	83
Jessyca_West	89
Esmeralda.Mraz57	90
Bethany20	91

Day wise User registrations count



Top 5 commonly used HashTags



My Insights

- In the previous slide, I jotted down the top 5 loyal users, the inactive users who did not make a single post, the one Contest winner, top 5 hashtags to have high reach.
- The Marketing team had a special request to provide insights on when to schedule a campaign. According to the report, Thursday and Sunday are the days in the week most users register on the platform.
- For Investor Metrics, the fake accounts list is provided on the dashboard. The average posts an user post (including inactive users) is 2.57. The Average posts of an user who is active is 3.473. The latter number is obtained by not considering the 26 inactive users. I excluded this information on the dashboard since it may lead to confusion.
- Here, my journey on this project comes to an end.

Thank
you

