

# Instagram User Analytics

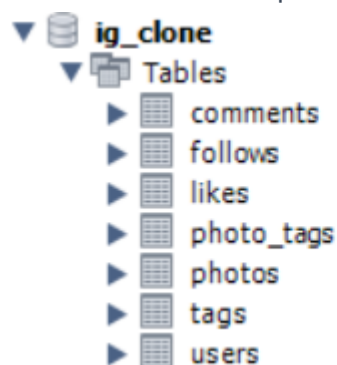
## Project Description:

This project is aimed at deriving insightful data to significantly influence decisions based on data driven metrics for instagram, in order to grow and enhance the potency of the platform. The Point of view for this project is simple i.e., to use data from the instagram database and analyse the same to reach a potential resolution for some of the key tasks and assignments dedicated for the marketing team which boosts the ability of this platform among the users and in-turn generate and influence revenue based on data for the stakeholders of the platform.

## Approach:

#Gather the required data under one database for analysis under MySQL Workbench.

Below pic explains the data required; Ig\_Clone being the database used and under the same tables with multiple attributes are listed.







1. Help the marketing team and sought a solution for the team where it's required to reward the most loyal user.

**Solution:** which can be simply achieved by Identifying the five oldest users on Instagram platform with the help of SQL queries from the database> Users> Created\_at.

>> Here to get the desired results, we have dense\_ranked the 'created\_at' attribute/column with an order by function in the Asc order(renamed the output as loyal\_user) and have limited our results to top 5 rows with the limit function.

>> The query and the output SS has been pasted below for more info, please refer the same.

```
4      #Task 1
5      # Identify the five oldest users on Instagram from the provided database.
6      •  select *,
7          dense_rank() over(order by created_at asc) As Loyal_user
8      from users
9      limit 0,5;
```

result Grid     Filter Rows: <input type="text"/>   Export:    Wrap Cell Content: 			
id	username	created_at	Loyal_user
80	Darby_Herzog	2016-05-06 00:14:21	1
67	Emilio_Bernier52	2016-05-06 13:04:30	2
63	Elenor88	2016-05-08 01:30:41	3
95	Nicole71	2016-05-09 17:30:22	4
38	Jordyn.Jacobson2	2016-05-14 07:56:26	5

2. Encouraging inactive users to start posting by sending them promotional emails.

**Solution:** Identify users who have never posted a single photo on Instagram and target those users.

>> To fetch this data we have joined 2 tables 'Photos' and 'Users'

>> Using count and group by function we've managed to count total post by each individual (Gathered ids of individuals who have posted) and termed the same as 'temp\_table' with CTE.

>> Now, using right join we've matched (users table and the above query results to filter id's which have not posted (null values) any posts)

>> This resulted in data where users have never posted on the platform.

>> Output is 26 rows of user/IDs who have never posted on the platform.

>> The query and the output SS has been pasted below for more info, please refer to the same.

```
5 #task 2 Users who never posted on instagram.
6 • with temp_table as
7   (select user_id, count(id) as total_posts
8    from photos
9    group by user_id)
10
11 select row_number() over() as `S.no`, users.id, users.username, temp_table.total_posts
12 from temp_table
13 right join users
14 on temp_table.user_id=users.id where user_id is null;
15
```

Result Grid			
Filter Rows: <input type="text"/>			
Export:			
Wrap Cell Content:			
S.no	id	username	total_posts
1	5	Aniya_Hackett	NULL
2	7	Kasandra_Homenick	NULL
3	14	Jadyn81	NULL
4	21	Rocio33	NULL
5	24	Maxwell.Halvorson	NULL
6	25	Tierra.Trantow	NULL
7	34	Pearl7	NULL
8	36	Ollie_Ledner37	NULL
9	41	Mckenna17	NULL
10	45	David.Osinski47	NULL
11	49	Morgan.Kassulke	NULL
12	53	Linnea59	NULL

Result 157 x

Output

Action Output

#	Time	Action	Message
1255	20:10:27	select *, dense_rank() over(order by created_at asc) As Loyal_user from users limit 0,5	5 row(s) returned
1256	20:21:51	with temp_table as (select user_id, count(id) as total_posts from photos group by user_id) #using right jo...	26 row(s) returned

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

**Solution:** Find the most liked photo on instagram by the users under our database/platform.

>> From the 'likes' table, we counted individual user (count(user\_id)) who liked how many of the photo\_id using groupby(photo\_id) function; i.e., total count of user\_id and group by photo id. From which we get the total likes on each photo id

>> Now from the photos table we matched the data using inner join with column(id).

#This gives us Total likes on each id and image url.

>> Finally to filter the MAX(total likes) we used WHERE function. As aggregated functions along with other columns cannot be yielded in a single select function

>> Output> id 145 has total likes of 48 on the image\_url <https://jarret.name> is the most liked photo.

>>The query and the output SS has been pasted below for more info, please refer to the same.

```
4      #Task 3
5      • With CTE_like as
6      (select photo_id, count(user_id) as Total_likes
7      from likes
8      group by photo_id),
9
10     RESULTS AS(
11     select id, Total_likes, image_url
12     from CTE_like
13     inner join photos
14     on CTE_like.photo_id = photos.id)
15
16     SELECT id, Total_likes, image_url
17     FROM RESULTS
18     WHERE TOTAL_LIKES = (SELECT MAX(TOTAL_LIKES) FROM RESULTS);
19
```

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

	id	Total_likes	image_url
▶	145	48	<a href="https://jarret.name">https://jarret.name</a>

4. One of our partner brand wants to know the most popular hashtags to use in their posts to reach the most people

**Solution:** Identify and suggest the top five most commonly used hashtags on the platform simply by using basic Sql queries.

>> Used count on photo\_id and grouped by/distributed them by tag\_id to find total 'Hashtag\_used\_count' per tag\_id

>> Later inner joined photo\_tags table(CTE\_hash) and tags table by column id of hashtags to fetch tag\_name.

>> From the above results table, dense\_ranked 'hashtag\_used\_count' in desc order and limited them to top 5; to fetch the top 5 used hashtags

>> **Smile, beach, party, fun, food were the top 5 used hashtags.**

>>The query and the output SS has been pasted below for more info, please refer to the same.

```

4      #task 4
5      • With CTE_hash as (
6          select count(photo_id) as Hastag_used_count, tag_id
7          from photo_tags
8          group by tag_id),
9
10     RESULTS AS(
11         select Hastag_used_count, tag_id, tag_name
12         from CTE_hash
13         inner join tags
14         On CTE_hash.tag_id = tags.id)
15
16     select tag_id, Hastag_used_count, tag_name,
17         dense_rank() over(order by Hastag_used_count desc) as top_5_hashtags
18     from Results
19     limit 0,5;
20

```

Result Grid				
Filter Rows:				
Export:				
Wrap Cell Content:				
tag_id	Hastag_used_count	tag_name	top_5_hashtags	
21	59	smile	1	
20	42	beach	2	
17	39	party	3	
13	38	fun	4	
5	24	food	5	

Result 29 x

5. Launch an ad campaign for which we need to figure out the best day to launch the same.

**Solution:** Simply be achieved on the day of the week when most users register on Instagram will be the schedule for the ad campaign.

>> We select the dayname from the timestamp and count the total users along side.

>> we then group those total users by dayname(day of the week); we get total no of id created on each day of the week

>> limit the same to a value of 1, we get Thursday when most users register on Instagram And we schedule for the ad campaign Thursday.

>>The query and the output SS has been pasted below for more info, please refer to the same.

```

13      #task5
14      • SELECT DAYNAME(created_at) AS day_of_week, COUNT(id) as user_count
15      FROM users
16      group by dayname(created_at)
17      order by user_count Desc
18      limit 1;

```

Result Grid		
Filter Rows:		
Export:		
Wrap Cell Content:		
Fetch rows:		
day_of_week	user_count	
Thursday	16	

## B) Investor Metrics:

6. As the Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

**Solution:** Simply find the average number of posts per user on Instagram. Additionally, find the total number of photos on Instagram divided by the total number of users. Which will provide an idea for the Stakeholders

>> We right joined table Users and photos with common attribute as 'ID'

>> We counted (image\_url) and counted(distinct user\_id) and divided the same  $\text{count}(\text{image\_url})/\text{count}(\text{distinct user\_id})$  to fetch Avg\_Post\_Per\_User

>> Avg\_Post\_Per\_User = 3.5 rounded to 1 decimal.(#3 as whole number)

>> For the total number of photos on Instagram divided by the total number of users; we Counted (image\_url) and (distinct id) and divided the same  $\text{count}(\text{image\_url})/\text{count}(\text{distinct id})$  to fetch the desired result: 2.6 rounded to decimal of 1.

>> Finally printed the same results under one select statement.

>>The query and the output SS has been pasted below for more info, please refer to the same.

```
37
38 #Task B 1
39
40 with cte_results as (
41     select right_table.id, right_table.username, left_table.user_id, left_table.id as img_id, left_table.image_url
42     from users right_table
43     left join photos left_table
44     on right_table.id = left_table.user_id
45
46     select round(count(image_url)/count(distinct user_id),1) as Avg_Post_Per_User,
47     round(count(image_url)/count(distinct id),1) as `Total_Posts/total_Users` from cte_results;
48
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
Avg_Post_Per_User	Total_Posts/total_Users			
3.5	2.6			

7. Provide details about fake and dummy accounts to the investors.

**Solution:** Spot the fake accounts/bots

>> From the likes table, we counted(distinct photo\_id) under CTE\_results.

>> Later Using grouped it by user\_id : This gives us each user liking total no of photos

>> Later with having clause, we counted  $\text{count}(\text{distinct photo\_id}) = (\text{select act\_id from CTE\_results})$ ; resulting in user\_id which have liked all 257 id in total.

>> Resulting in 13 unique ids which are potentially bots.

>>The query and the output SS has been pasted below for more info, please refer to the same.

```

27
28
29
30 #Bots
31 #Task B 2
32 • SELECT * FROM ig_clone.likes;
33
34 • With CTE_results As (
35     select count(distinct photo_id) as act_id from likes
36 )
37     select user_id from likes
38     group by user_id
39     having count(distinct photo_id) = (select act_id from CTE_results);
40

```

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

user_id
5
14
21
24
36
41
54
67

Result 158 x

Output

Action Output

#	Time	Action	Message
✓ 1277	21:12:21	With CTE_results As ( select count(distinct photo_id) as act_id from likes ) select user_id from likes grou...	13 row(s) returned
✓ 1278	21:12:26	With CTE_results As ( select count(distinct photo_id) as act_id from likes ) select user_id from likes grou...	13 row(s) returned

**NOTE: Please find all the results separately listed under Results sub-head**

## Tech-Stack Used

MySQL Workbench 8.0.38 Community edition.

MySQL is a popular open-source database management system known for its high performance and reliability. It's cost-effective, making it ideal for startups, students and small projects, and supports scalability for growing applications. With a user-friendly interface and extensive documentation with a heavy user base and wide adoption one can find support and resources about the same all across the programmers(SQL) community, as it's accessible for both beginners and advanced users(paid edition/licensed).Overall, MySQL is a solid choice for a range of applications, from small websites to large systems.

## Insights:

Working on this project has provided valuable hands-on experience, significantly boosting my confidence in using SQL and crafting queries. It's clear that simply observing an instructor isn't enough to truly learn a programming language; relying solely on theoretical knowledge can be limiting. The real challenge—and growth—comes when you actively apply the concepts you've learned throughout the journey. This is exactly how one can learn any programming i.e., through applying the knowledge gained throughout the journey.

Apart from the assigned task i have noticed

1. Significant data insights where i can single handedly point out Photo\_id 13, 157,247 Were trending photos of their time since they are the most commented photos across the data/platform with each photo having 39 comments.(Consider the fact that the algorithm promotes engagements leading to trending photos)

	photo_id	comments_per_photo
▶	13	39
	157	39
	247	39

2. I have also noticed 13 user\_ids; are the most active users across the platform since they have commented the most(257 comments each). A list of users have been provided for the same.

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

3. Also worked on the highest followed account.  
But per the data all accounts are following each other leading to all individuals having 99 followers(just a catch here 😊)

	follower_id	following_users
▶	2	99
	3	99
	4	99
	5	99
	6	99
	8	99
	9	99
	10	99
	11	99
	12	99
	13	99
	14	99
	15	99

## Result(outputs for the same have been pasted below)

Task1:

ID	Username	created_at	Loyal_user
80	Darby_Herzog	2016-05-06 00:14:21	1
67	Emilio_Bernier52	2016-05-06 13:04:30	2
63	Elenor88	2016-05-08 01:30:41	3
95	Nicole71	2016-05-09 17:30:22	4
38	Jordyn.Jacobson2	2016-05-14 07:56:26	5

Task2:

S.no ID Username

1	5	Aniya_Hackett
2	7	Kasandra_Homenick
3	14	Jaclyn81
4	21	Rocio33
5	24	Maxwell.Halvorson
6	25	Tierra.Trantow
7	34	Pearl7
8	36	Ollie_Ledner37
9	41	Mckenna17
10	45	David.Osinski47
11	49	Morgan.Kassulke
12	53	Linnea59
13	54	Duane60
14	57	Julien_Schmidt
15	66	Mike.Auer39
16	68	Franco_Keebler64
17	71	Nia_Haag
18	74	Hulda.Macejkovic
19	75	Leslie67
20	76	Janelle.Nikolaus81
21	80	Darby_Herzog



22 81 Esther.Zulauf61

23 83 Bartholome.Bernhard

24 89 Jessyca\_West

25 90 Esmeralda.Mraz57

26 91 Bethany20

**Task 3:**

ID likes img\_url

145 48 https://jarret.name

**Task 4:**

Tag\_id used\_count name rank

21 59 smile 1

20 42 beach 2

17 39 party 3

13 38 fun 4

5 24 food 5

**Task 5:**

DAY user\_count

Thursday 16

**Task 5:**

**B) Investor Metrics Task1:**

	Avg_Post_Per_User	Total_Posts/total_Users
	3.5	2.6

**B) Investor Metrics Task2:**

Bots/user\_id

5

14

21

24

36

41

54

57

66

71

75

76

91

=====

**All tasks outlined in our project plan have been successfully completed. We have retrieved the necessary marketing data from our database, and the results have been published. Additionally, key investor metrics have been extracted, revealing a minimal presence of fake users—only 13 unique IDs appear to be potentially bots out of a sample size of 100.**

**In terms of engagement, these 13 users have shown significant activity on Instagram, collectively making 257 comments each. This suggests that they are actively interacting with content, which could inform our marketing strategies moving forward.**

**Moreover, our analysis indicates that Thursdays are the best day for launching marketing campaigns, as this day consistently attracts the highest number of registered users. Leveraging this insight could enhance our ability to convert leads into successful outcomes for our advertising efforts.**

**In conclusion, the insights derived from our data not only affirm the quality of our user engagement but also provide strategic direction for optimising our marketing campaigns moving forward.**

=====