

A PROJECT ON INSTAGRAM USER ANALYTICS



By – Vansh

Mail ID – aroravansh11@gmail.com

SQL TASKS

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 provided database

Solution:

Query:

```
# Task Identify the five oldest users on instagram
```

```
use ig_clone;
```

```
select * from users;
```

```
SELECT
```

```
    id, username, MIN(created_at) AS oldest_user
```

```
FROM
```

```
    users
```

```
GROUP BY username , id
```

```
having oldest_user
```

```
ORDER BY oldest_user limit 5
```

```
;
```

Output:

Result Grid				Filter Rows:		E
	id	username	oldest_user			
▶	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			

Explanation:

Here,

- Firstly, I checked the user table data, there I found that for finding five oldest users I must arrange the table with oldest dates first
- Then after arranging the table, I set a limit of 5 so that I can get only first five rows from the table as the table is arranged in ascending order from oldest dates which are coming first
→ So, in first five rows I found five oldest users with their username, id, and oldest user dates

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.

Solution:

Query:

```
# Identify users who have never posted a single photo on Instagram.
```

```
select users.id,user_id, username, image_url
from users
left join photos
on users.id = photos.user_id
;
```

Output:



Result Grid				
		Filter Rows:	Export:	
	id	user_id	username	image_url
	4	4	Arely_Bogan63	https://maudie.org
	5	NULL	Aniya_Hackett	NULL
	6	6	Travon.Waters	http://annamae.name
	6	6	Travon.Waters	https://mac.org
	6	6	Travon.Waters	http://miracle.info
	6	6	Travon.Waters	http://emmet.com
	6	6	Travon.Waters	https://lisa.com
	7	NULL	Kassandra_Hom...	NULL
Result 23 x				

Query:

```
# Identify users who have never posted a single photo on Instagram.
```

```
select users.id, user_id, username, image_url
from users
left join photos
on users.id = photos.user_id
where image_url is null;
```

Output:

Result Grid			 Filter Rows:	<input type="text"/>	E
	id	user_id	username	image_url	
▶	5	NULL	Aniya_Hackett	NULL	
	7	NULL	Kassandra_Homenick	NULL	
	14	NULL	Jadyn81	NULL	
	21	NULL	Rocio33	NULL	
	24	NULL	Maxwell.Halvorson	NULL	
	25	NULL	Tierra.Trantow	NULL	
	34	NULL	Pearl7	NULL	
	36	NULL	Ollie_Ledner37	NULL	
	41	NULL	Mckenna17	NULL	
	45	NULL	David.Osinski47	NULL	
	49	NULL	Morgan.Kassulke	NULL	
	53	NULL	Linnea59	NULL	

Explanation:

Here,

- Firstly, I joined two tables those are “users” and “photos” using left join as I wanted that every user should be present in result set, so I used left join
- Then I found that many users are there who are having null values in their “image_url” column means they have not posted a single photo on Instagram

→ For uniquely presenting them I used a query i.e., “where image_url is null” which defined the result set very clearly containing only those users who have not posted a single photo on Instagram.

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.

Solution:

1st Query:

Determine the winner of the contest and provide their details to the team.

```
select * from likes;
```

```
select photo_id, count(photo_id) as count_value  
from likes  
group by photo_id  
order by count_value desc;
```

Output:

Result Grid			Filter Rows
	photo_id	count_value	
▶	145	48	
	127	43	
	182	43	
	123	42	
	30	41	
	52	41	
	61	41	
	147	41	
	174	41	
	192	41	
	256	41	
	13	40	

2nd Query:

So the 145 photo_id got more likes lets find the username attached with this photo_id who won

```
select id, image_url, user_id
from photos
where id=145;
```

Output:

Result Grid			
	id	image_url	user_id
▶	145	https://jarret.name	52
●	NULL	NULL	NULL

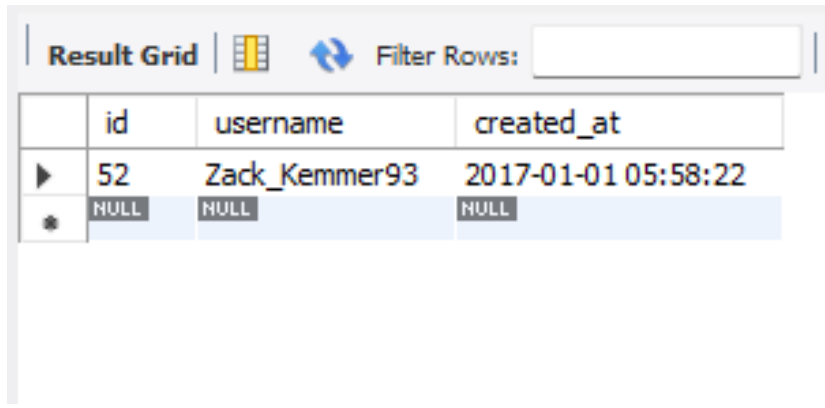
3rd Query:

Now we got the user_id so its now easy to find the winner

```
select * from users;
```

```
select id,username,created_at
from users
where id=52;
```

Output:



	id	username	created_at
▶	52	Zack_Kemmer93	2017-01-01 05:58:22
*	NULL	NULL	NULL

Explanation:

Here,

- **In the First Query** I searched for “photo id” with highest likes by using count function with order by clause and arranged the “count_value” in descending order so that the highest liked “photo_id” will be shown first, all this is done from the likes table.
- **In Second Query** with the help of result of 1st query i.e., with photo id I searched for the user id to whom that photo id belongs from the photos table and used where clause for directly getting the important information.
- **In the Third Query** now I got the “user_id” from the “Second Query” so its easy to find the username with the help of user id which I searched in users table as shown in the query and got desired output
→ So, The Winner is Id no 52 i.e., “**Zack_kemmer93**”

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 used hashtags on the platform

Solution:

1st Query:

```
# Identify and suggest the top five most commonly used hashtags on the platform.
```

```
select tag_id, count(tag_id) as tagcount
from photo_tags
group by tag_id
order by tagcount desc limit 5 ;
```

Output:

Result Grid		
	tag_id	tagcount
▶	21	59
	20	42
	17	39
	13	38
	18	24

2nd Query:

```
# we found five tag_id 21,20,17,13,18 which are mostly used hashtags on the platform
```

```
select id, tag_name
from tags
where id in ("21","20","17","13","18");
```

```
# so these are the tags which are mostly used hastags on the platform
```

Output:

Result Grid		
	id	tag_name
▶	13	fun
	17	party
	18	concert
	20	beach
	21	smile
⊙	NULL	NULL

Explanation:

Here,

- **In First Query** I have searched for tagcount through which I got to know that which “tag_id” is mostly used and all this I have done using “photo_tags” table and with the help of group by and order by clause as mentioned above in the first query.
- After getting “tag_id” from the result set of first query, its easy to find that what is the tag name of that “tag_id” and this is done using “tags” table, in this I have used in operator with where clause for uniquely defining the hastags.
→ So, the Hastags which are mostly used on the platform are
1. Smile
2. Beach
3. Party
4. Fun 5. Concert.

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.

Solution:

Query:

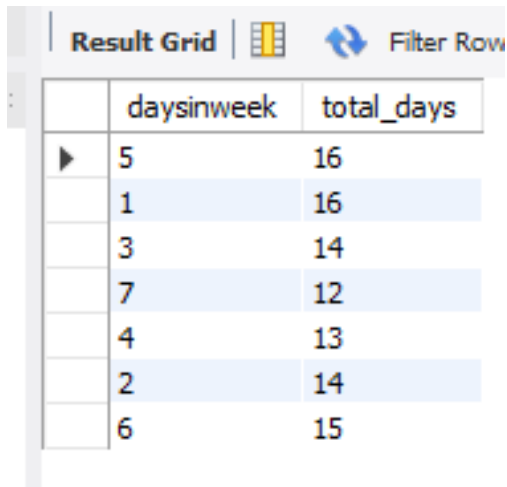
```
# Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.
```

```
select * from users;
```

```
select dayofweek(created_at)as daysinweek,  
count(dayofweek(created_at) )as total_days  
from users  
group by daysinweek;
```

```
# Here we got 2 days we can go with either 5 i.e., Thursday or with 1 i.e, Sunday
```

Output:



	daysinweek	total_days
▶	5	16
	1	16
	3	14
	7	12
	4	13
	2	14
	6	15

Explanation:

Here,

- **For calculating days in week**, I have used a function i.e., dayofweek which provides values as 1 for Sunday, 7 for Saturday and remaining numbers also goes with sequence, All the calculations are done using users table and “created_at” column
- **Count function** is used for counting the repetition for finding the accurate day for ad campaigns
→ **So**, we got two values 5(Thursday) and 1(Sunday) we can go with either one of them there is no issue as most of the users register on these two dates.

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.



Solution:

(P.T.O)

Query:

```
# 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.  
  
select count(photos.id) as total_photos,  
       count(users.id) as total_users,  
       (count(photos.id)/count(users.id)) as resultset  
from photos  
Left join users  
on photos.id = users.id
```

Output:

Result Grid   Filter Rows: <input type="text"/>			
	total_photos	total_users	resultset
▶	257	100	2.5700

Explanation:

Here,

- For Calculation the average number of posts per user on Instagram I used count function for finding the total number of posts/photos and users
- Then I used Divide Operator (/) for dividing them and getting the result set which shows the average.
- For Extracting data from multiple tables I used Left join.

2. Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

Your Task: Identify users who have liked every single photo on the site, as this is not typically possible for a normal user.

Solution:

Query:

```
# Identify users (potential bots) who have liked every single photo on the site,  
  
select l.user_id,u.username  
from (  
    select user_id, count(Distinct image_url) as total_photos  
    from photos  
    group by user_id ) as p  
join (  
    select user_id, count(Distinct photo_id) as liked_photos  
    from likes  
    group by user_id ) as l  
on  
    p.user_id = l.user_id  
join users u on u.id = p.user_id  
where p.total_photos = l.liked_photos;
```

Output:

Result Grid				Filter
	user_id	username		

Explanation:

Here,

- Firstly, in this I have used count function for counting the total photos and number of likes per user in which distinct function I used for getting only unique values
- Then I used “join” for finding out whether a user has liked the photo,
- Where is used for comparing the total number of distinct photos with the number of distinct photos liked by each user
- **So**, if they match it means the user has liked every single photo on the site.

→Result set shows that there is no such user who liked every photo That means the data is bot free

(P.T.O)

PROJECT DESCRIPTION

In this project I assumed me as a data analyst working with the product team at **Instagram**. My role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

The purpose of this project is to gain knowledge of practical work

The approach I followed is like first I understood that I have been provided with some tables database through which I derived some insights and on the basis of that data I had to solve the questions.

Approach

The approach is very simple

- First, I observed the data very carefully, after observing the description I started solving the questions
- Then at the time of solving the questions
 1. Firstly, I understood the question's sense that what I have to exactly find.
 2. Then with my understanding I solved the question and provided a brief explanation.

TECH STACK USED

- **MYSQL WORKBENCH 8.0 CE**

WHY I USED MYSQL WORKBENCH 8.0 CE

I used “MYSQL Workbench 8.0 CE” because I only learned database skills on this software, so for me this software is the best to go with.

Insights

- This project provided me with valuable skills which helped me to increase my data analytics skills as I faced challenges while solving problems.
- I learned from MySQL Software how we can solve a problem by sorting data and arranging it collectively.
- Through this project I analyzed that how user trends and patterns change over time
- I was engaged in tracking likes, comments and other forms of interaction

All the above insights I obtained from this project and this project really helped me to keep my first step in the field of data analytics.

Result

Finally, I have successfully accomplished my tasks, and this is my first main project in the field of data analytics which I will consider as an achievement for me,

This project really benefited me in terms of **Analytical Skills** as in each and every task I had to first analyze the problem that “**what actually I had to find**” and then after analyzing the problem I had to find the solution which again benefited me in terms of “**Problem Solving Skills**”.

And one more thing Because of this project submission format that is “pdf” format My Report Writing skills are also improving.

So, Sir/ma'am

If you feel like I should improve something in this report or project submission skills kindly update me with the feedback

on Gmail – aroravansh11@gmail.com

Thank You.