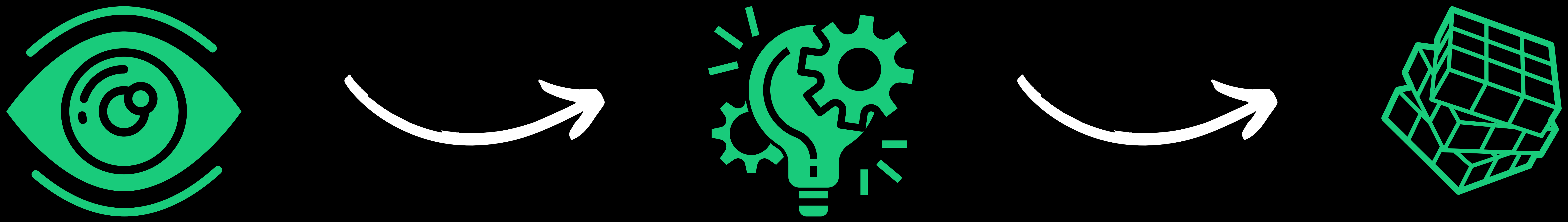
trainity

PROJECT 2

BY ASHISH KUMAR SAMANTARAY

PROJECT DESCRIPTION:

The project deals with the insights of the user activity across different photos, tracks activities such as comments, likes etc where we required to extract information based on market interest and investor's interest. So as to perform each task in Market Analysis or Investor Metrics, I have to analyze the question first, understand what is required for the data to be solved, run queries and get work done.



TECH STACK USED:

- XAAMP CONTROL PANEL v3.3.0 SUPPORTING MYSQL
- CANVA FOR CREATING PPT

I chose **XAAMP CONTROL PANEL** because previously I had already worked on Xaamp Admin supporting mySQL where I had performed some queries for a PG room finding site database.

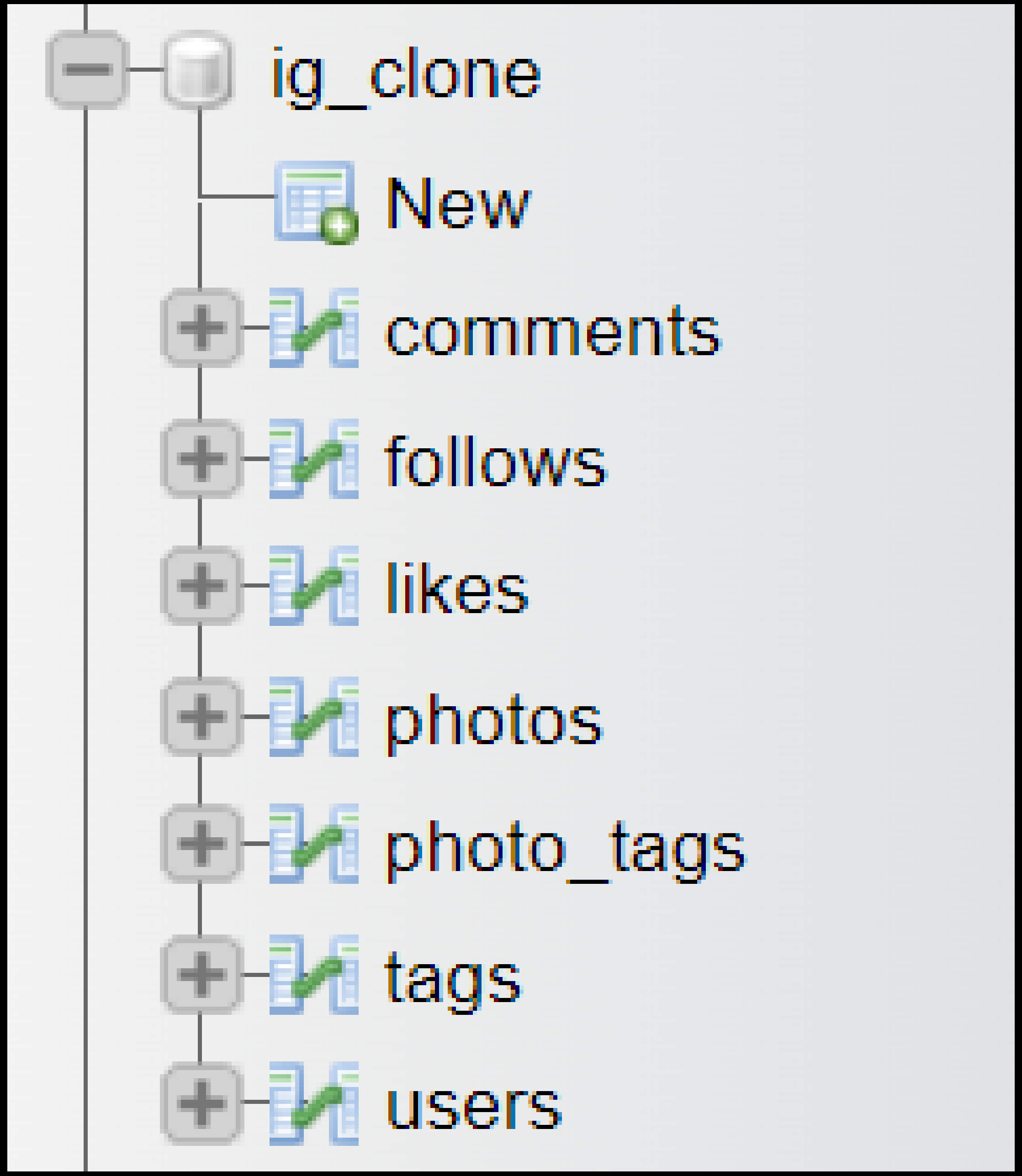
I chose **Canva** so as to make my PPT look more visually appealing.

INSIGHTS AHEAD

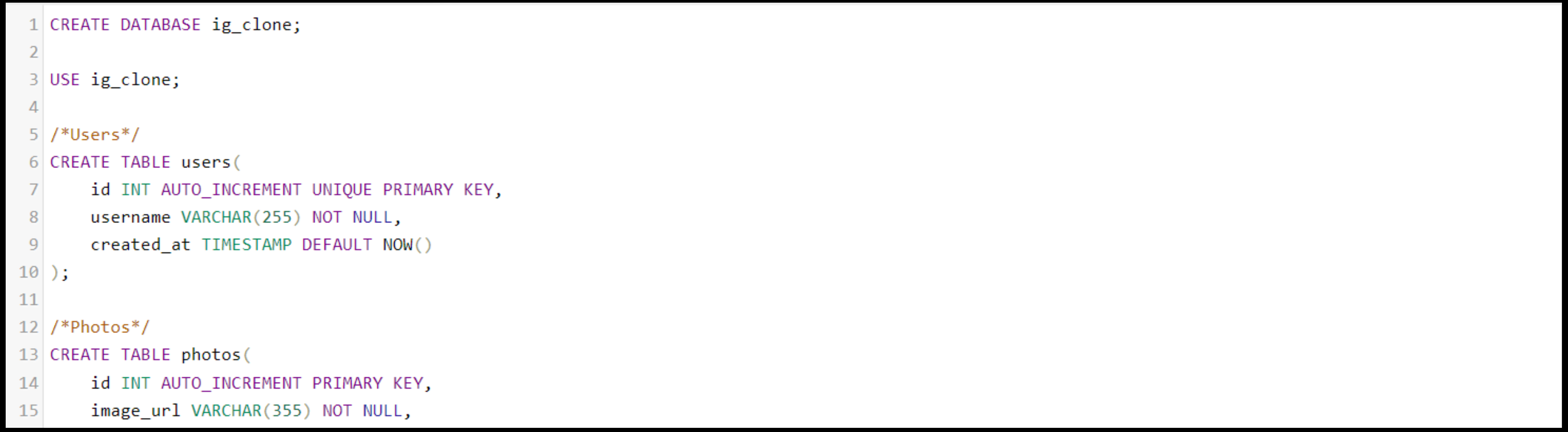
WITH QUERIES AND OUTPUT AND SUMMARY OF APPROACH

FIRST , A DATABASE WAS CREATED THROUGH THE SET OF QUERIES GIVEN IN THE DATASET ATTACHED IN THE PROJECT DESCRIPTION.

DATABASE:



QUERY BOX:



```sql
CREATE DATABASE ig_clone;

USE ig_clone;

/*Users*/
CREATE TABLE users(
    id INT AUTO_INCREMENT UNIQUE PRIMARY KEY,
    username VARCHAR(255) NOT NULL,
    created_at TIMESTAMP DEFAULT NOW()
);

/*Photos*/
CREATE TABLE photos(
    id INT AUTO_INCREMENT PRIMARY KEY,
    image_url VARCHAR(355) NOT NULL,
```

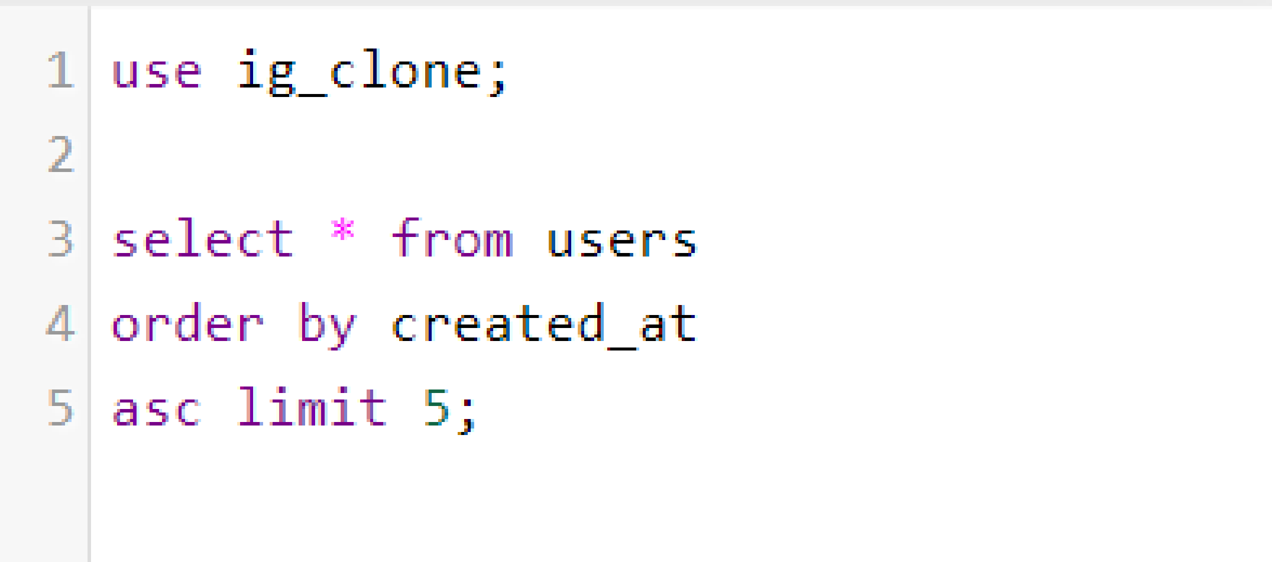
MARKETING ANALYSIS

TASK 1

Loyal User Reward:

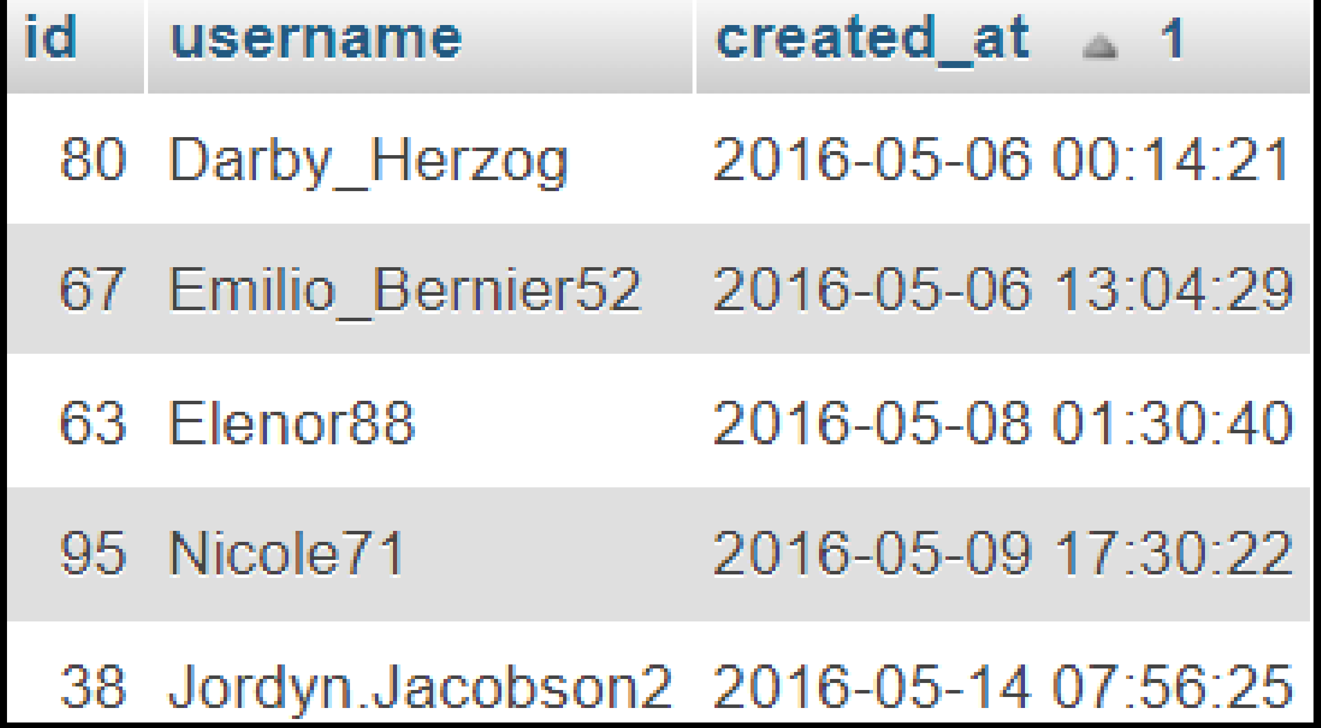
We were required to identify users who have never posted a single photo on Instagram for which we selected 5 users after ordering users by date of creation in ascending order.

QUERY BOX:



```
use ig_clone;

select * from users
order by created_at
asc limit 5;
```

OUTPUT::



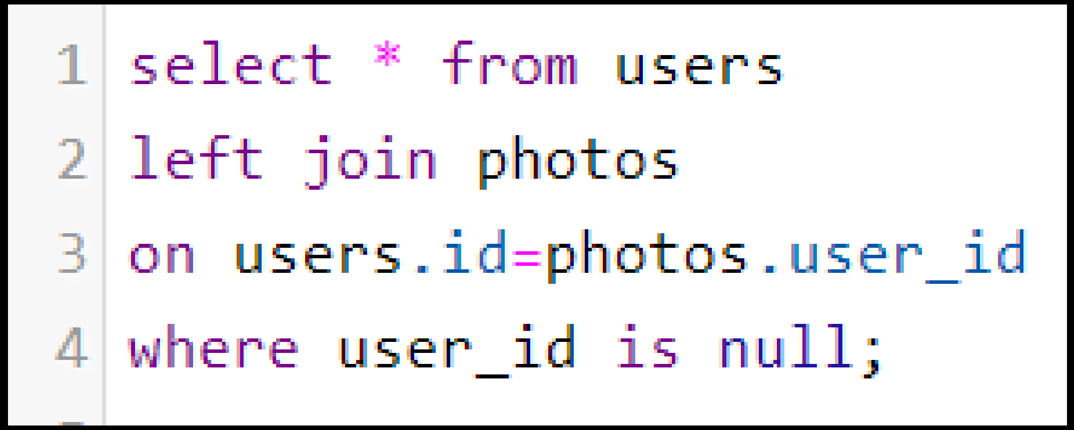
id	username	created_at ▲ 1
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:29
63	Elenor88	2016-05-08 01:30:40
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:25

TASK 2

Inactive User Engagement:

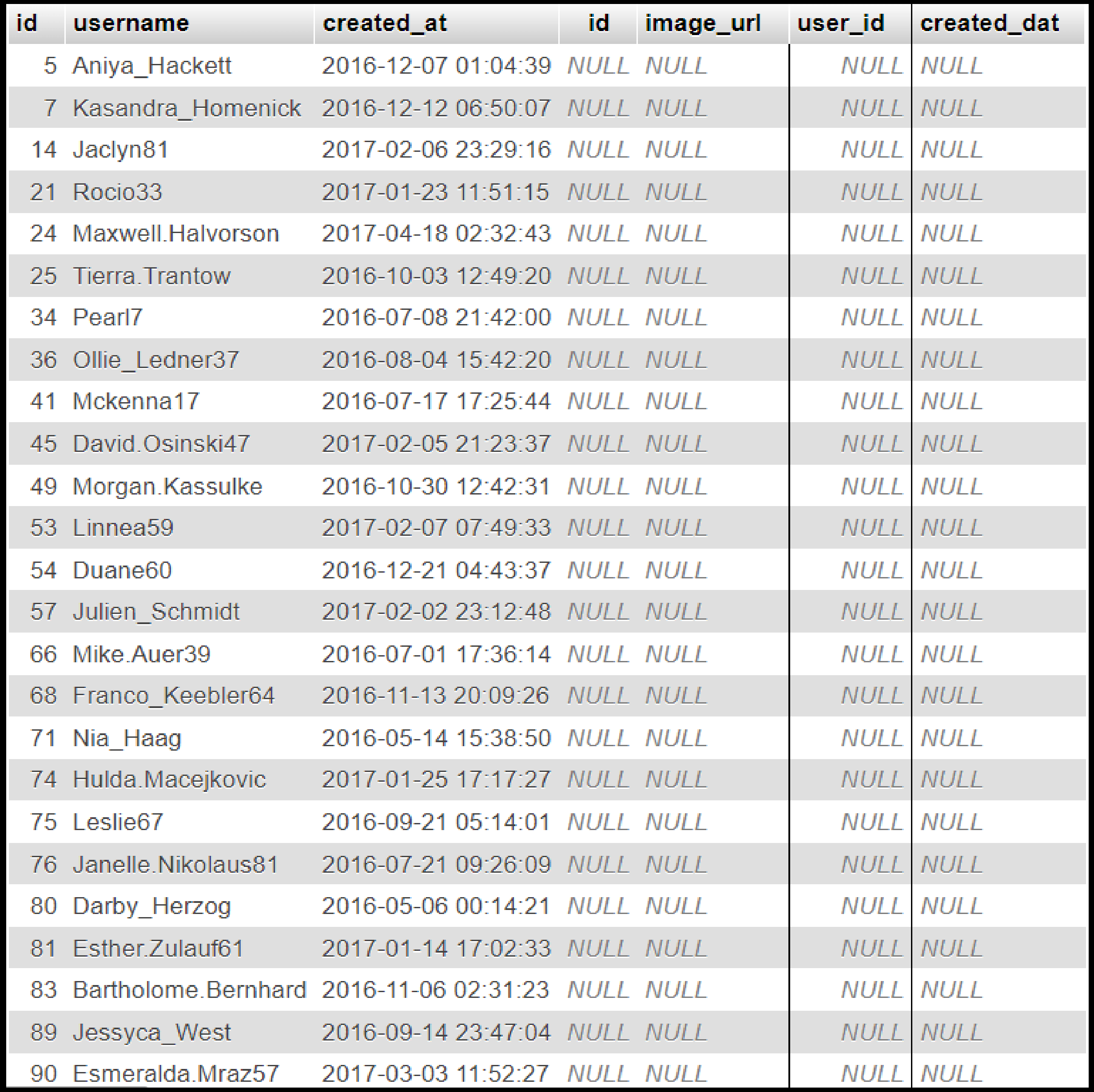
We were required to identify users who have never posted a single photo on Instagram for which we Lleft joined photos table on id of users in photos table and filtered the table where user id in photo is null as they have not posted any photo yet.

QUERY BOX:



```sql
select * from users
left join photos
on users.id=photos.user_id
where user_id is null;
```

OUTPUT:



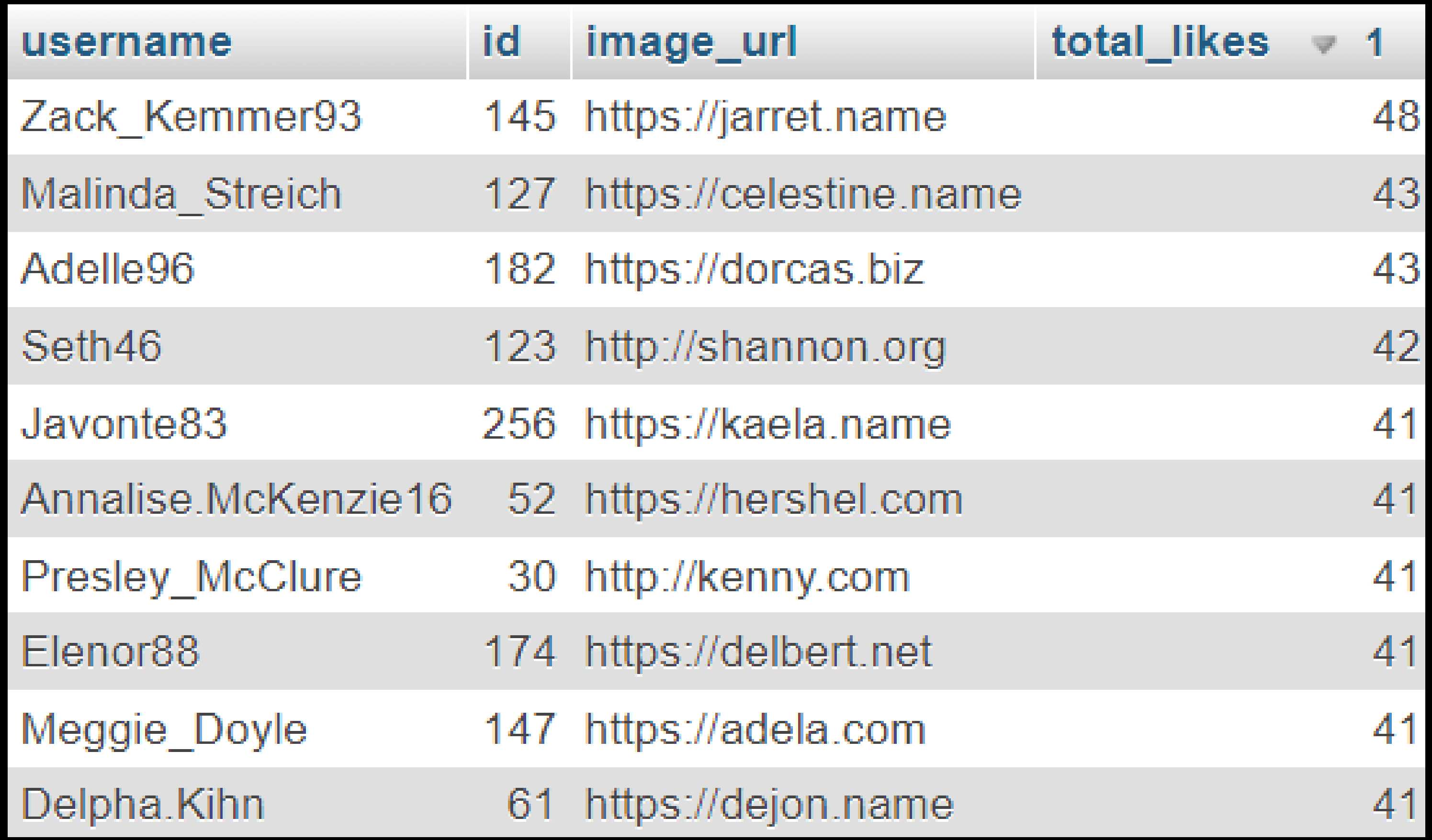
id	username	created_at	id	image_url	user_id	created_dat
5	Aniya_Hackett	2016-12-07 01:04:39	NULL	NULL	NULL	NULL
7	Kasandra_Homenick	2016-12-12 06:50:07	NULL	NULL	NULL	NULL
14	Jaclyn81	2017-02-06 23:29:16	NULL	NULL	NULL	NULL
21	Rocio33	2017-01-23 11:51:15	NULL	NULL	NULL	NULL
24	Maxwell.Halvorson	2017-04-18 02:32:43	NULL	NULL	NULL	NULL
25	Tierra.Trantow	2016-10-03 12:49:20	NULL	NULL	NULL	NULL
34	Pearl7	2016-07-08 21:42:00	NULL	NULL	NULL	NULL
36	Ollie_Ledner37	2016-08-04 15:42:20	NULL	NULL	NULL	NULL
41	Mckenna17	2016-07-17 17:25:44	NULL	NULL	NULL	NULL
45	David.Osinski47	2017-02-05 21:23:37	NULL	NULL	NULL	NULL
49	Morgan.Kassulke	2016-10-30 12:42:31	NULL	NULL	NULL	NULL
53	Linnea59	2017-02-07 07:49:33	NULL	NULL	NULL	NULL
54	Duane60	2016-12-21 04:43:37	NULL	NULL	NULL	NULL
57	Julien_Schmidt	2017-02-02 23:12:48	NULL	NULL	NULL	NULL
66	Mike.Auer39	2016-07-01 17:36:14	NULL	NULL	NULL	NULL
68	Franco_Keebler64	2016-11-13 20:09:26	NULL	NULL	NULL	NULL
71	Nia_Haag	2016-05-14 15:38:50	NULL	NULL	NULL	NULL
74	Hulda.Macejkovic	2017-01-25 17:17:27	NULL	NULL	NULL	NULL
75	Leslie67	2016-09-21 05:14:01	NULL	NULL	NULL	NULL
76	Janelle.Nikolaus81	2016-07-21 09:26:09	NULL	NULL	NULL	NULL
80	Darby_Herzog	2016-05-06 00:14:21	NULL	NULL	NULL	NULL
81	Esther.Zulauf61	2017-01-14 17:02:33	NULL	NULL	NULL	NULL
83	Bartholome.Bernhard	2016-11-06 02:31:23	NULL	NULL	NULL	NULL
89	Jessyca_West	2016-09-14 23:47:04	NULL	NULL	NULL	NULL
90	Esmeralda.Mraz57	2017-03-03 11:52:27	NULL	NULL	NULL	NULL

TASK 3

Contest Winner Declaration:

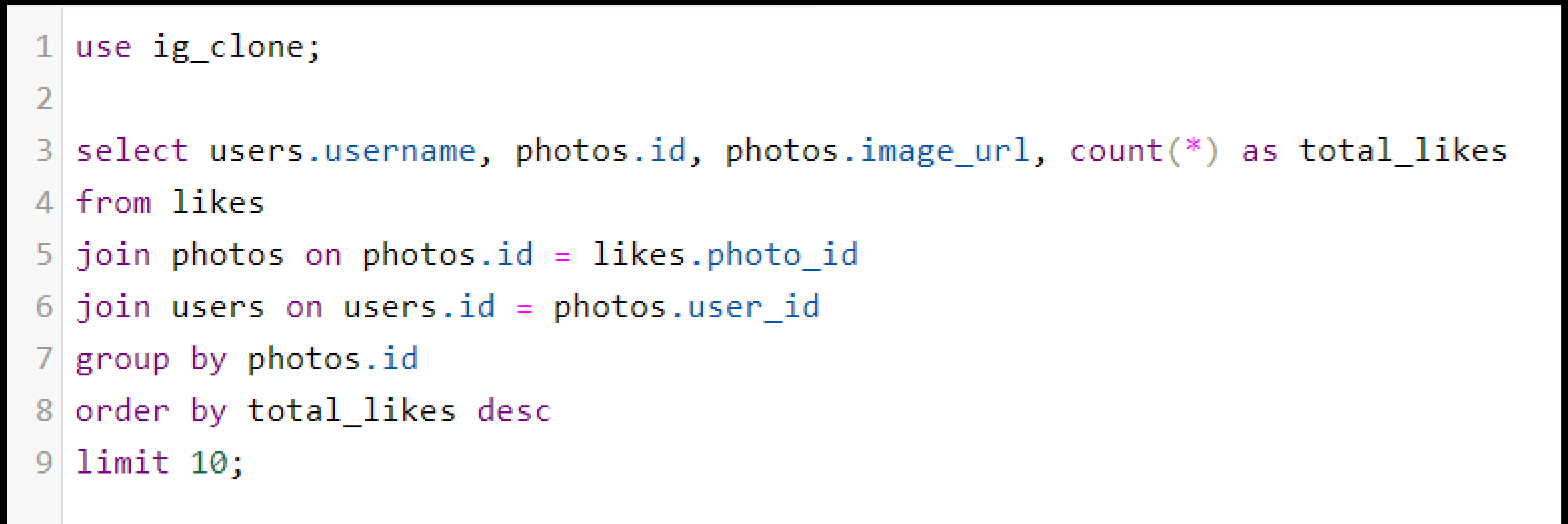
We were required to identify the users with the most likes on a single photo and determine the winner of the contest and provide their details to the team for which we have to join the table of photos and likes so as to determine which are the photo with most of the likes.

OUTPUT:



username	id	image_url	total_likes ▼ 1
Zack_Kemmer93	145	https://jarret.name	48
Malinda_Streich	127	https://celestine.name	43
Adelle96	182	https://dorcas.biz	43
Seth46	123	http://shannon.org	42
Javonte83	256	https://kaela.name	41
Annalise.McKenzie16	52	https://hershel.com	41
Presley_McClure	30	http://kenny.com	41
Elenor88	174	https://delbert.net	41
Meggie_Doyle	147	https://adela.com	41
Delpha.Kihn	61	https://dejon.name	41

QUERY BOX:



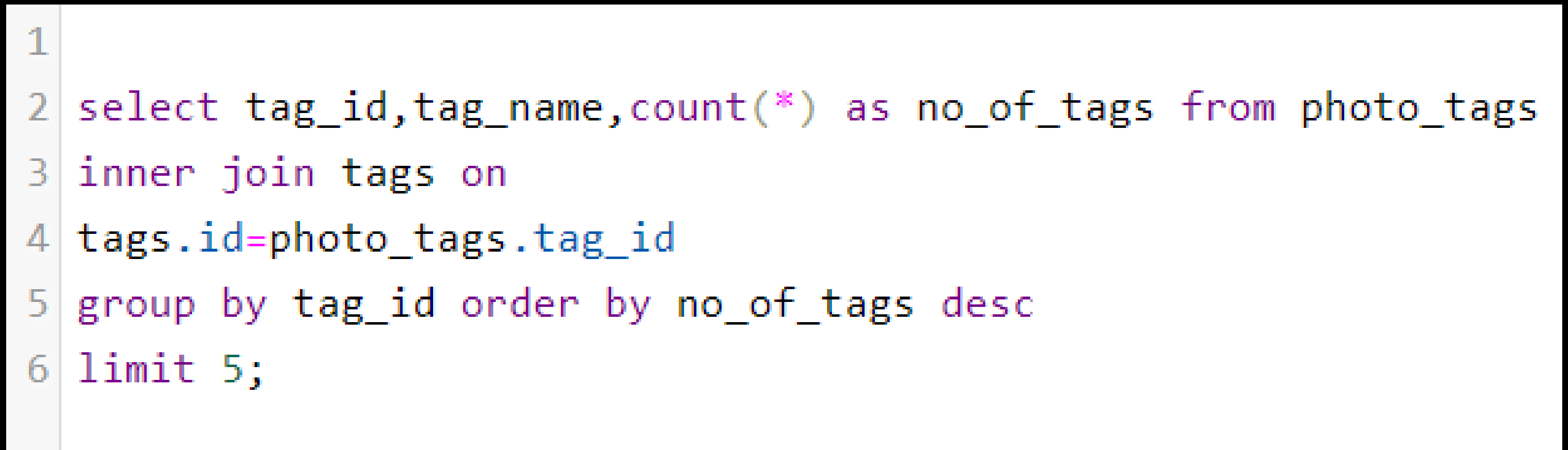
```
use ig_clone;

select users.username, photos.id, photos.image_url, count(*) as total_likes
from likes
join photos on photos.id = likes.photo_id
join users on users.id = photos.user_id
group by photos.id
order by total_likes desc
limit 10;
```

TASK 4

Hashtag Research:

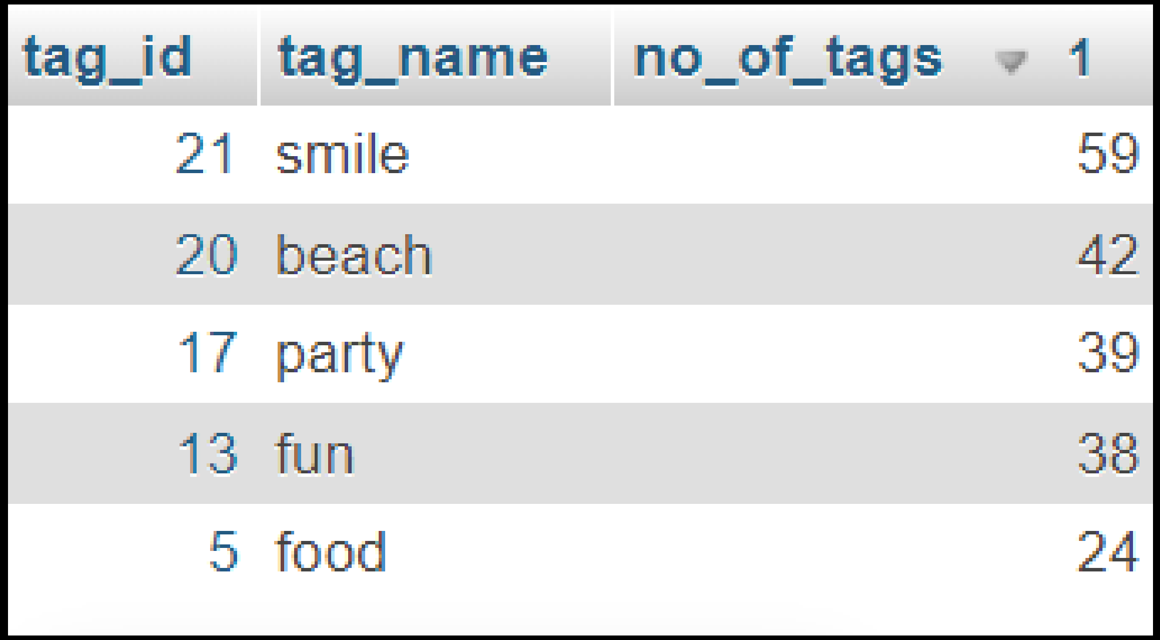
We were required to identify dentify and suggest the top five most commonly used hashtags on the platform for which we joined the photo_tags and tags table and then counted the no of photos the hashtags are used in the photos.

QUERY BOX:



```

select tag_id,tag_name,count(*) as no_of_tags from photo_tags
inner join tags on
tags.id=photo_tags.tag_id
group by tag_id order by no_of_tags desc
limit 5;
```

OUTPUT:



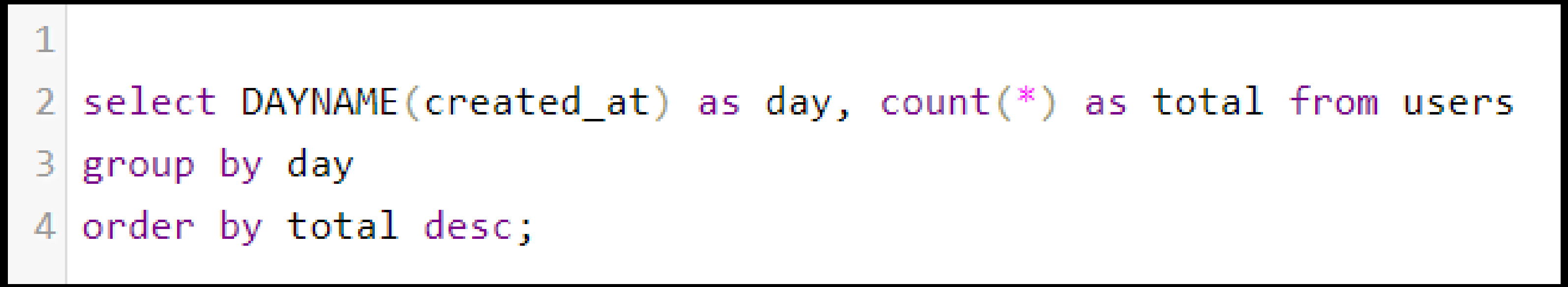
tag_id	tag_name	no_of_tags ▼ 1
21	smile	59
20	beach	42
17	party	39
13	fun	38
5	food	24

TASK 5

Ad Campaign Launch:

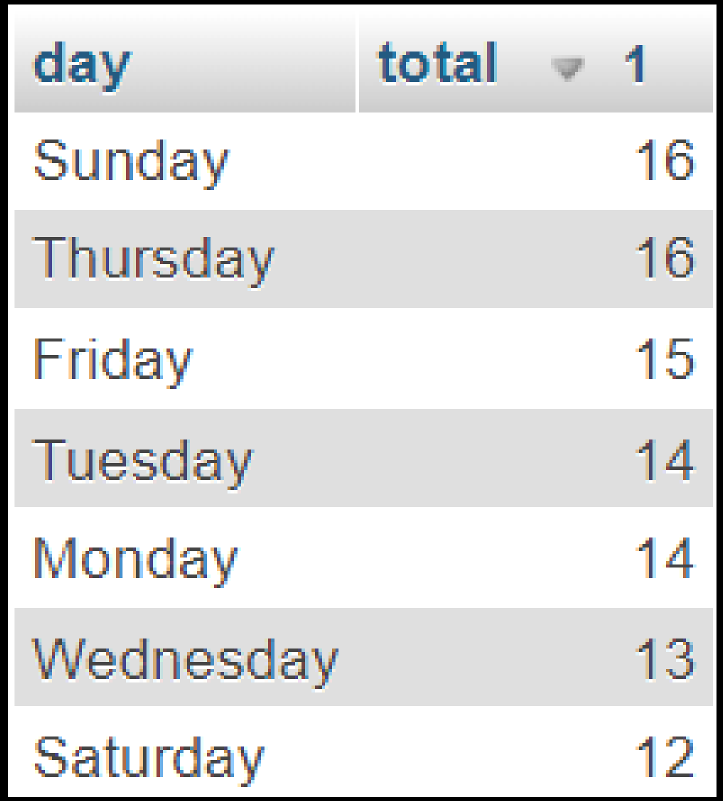
We were required to determine the day of the week when most users register on Instagram according to which we can launch an ad campaign so that most number of people could get to see the ad released for which we simply counted the no of users registered in a particular day of a week and then order them in descending order so as to identify the greatest no of aactivities in a day.

QUERY BOX:



```

select DAYNAME(created_at) as day, count(*) as total from users
group by day
order by total desc;
```

OUTPUT:



day	total ▼ 1
Sunday	16
Thursday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

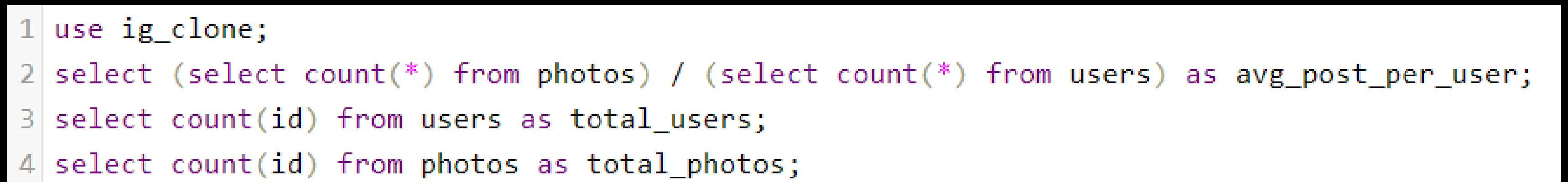
INVESTORS METRICS

TASK 1

User Engagement:

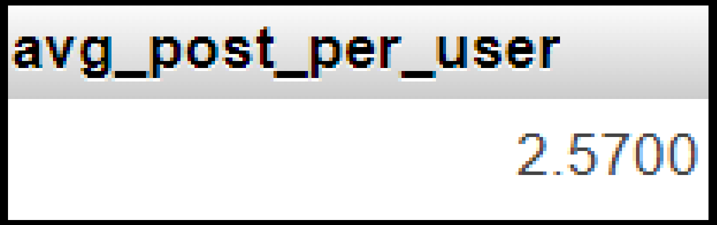
We were required to calculate the average number of posts per user on Instagram and also, provide the total number of photos on Instagram divided by the total number of users for which we created a temporary field of total no of rows in users and total no of rows in photos and divided them to get average . Further to manually verify this, we also extracted total no of users and total no of photos separately.

QUERY BOX:

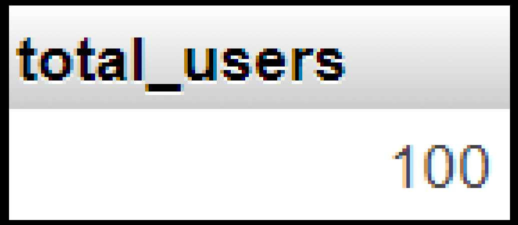


```
use ig_clone;
select (select count(*) from photos) / (select count(*) from users) as avg_post_per_user;
select count(id) from users as total_users;
select count(id) from photos as total_photos;
```

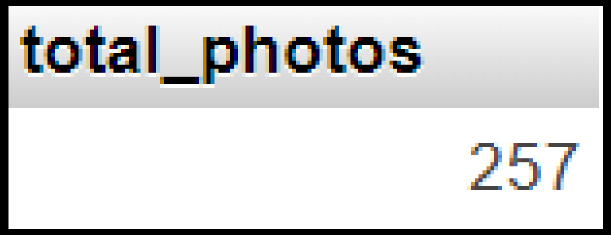
OUTPUT:



avg_post_per_user
2.5700



total_users
100



total_photos
257

MANUAL CALCUALTION:

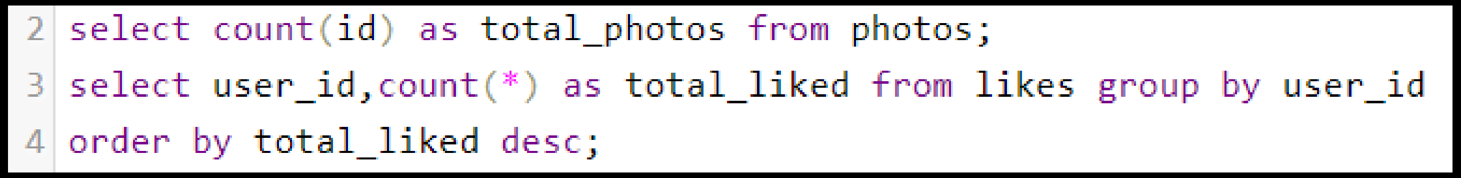
TOTAL_PHOTOS/TOTAL_USERS
=257/100
=2.57

TASK 2

Bots & Fake Accounts:

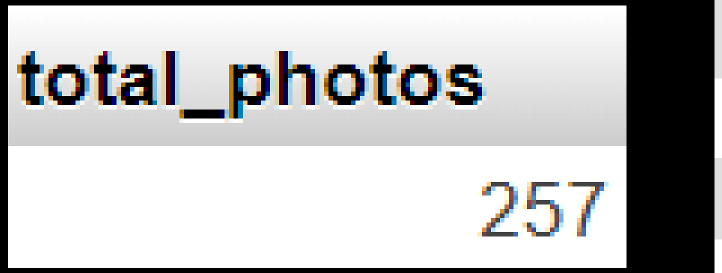
We were required to identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user. For this, we need to see how many photos were there in total. So we got 257, which means the user which has liked 257 photos has been a bot. Hence we grouped the no of photos liked according to the user id and listed down the userid in the form of table.

QUERY BOX:

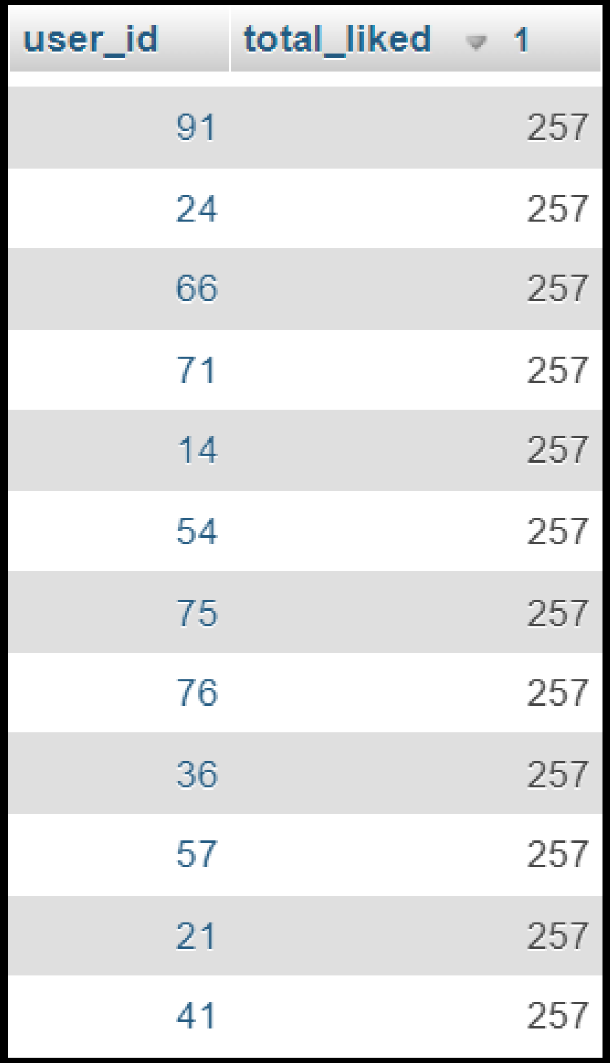


```
select count(id) as total_photos from photos;
select user_id,count(*) as total_liked from likes group by user_id
order by total_liked desc;
```

OUTPUT:



total_photos
257



user_id	total_liked ▼ 1
91	257
24	257
66	257
71	257
14	257
54	257
75	257
76	257
36	257
57	257
21	257
41	257

RESULT:

I successfully handled a practical situation and learned to run queries and provide insites for market analysis or investors metrics. Also, I was able to learn a lot of new queries and also learned to relate tables, join them and get desired work to be done.

Overall, I feel like I have mastered the basics which feels not more than an actual data scientist.