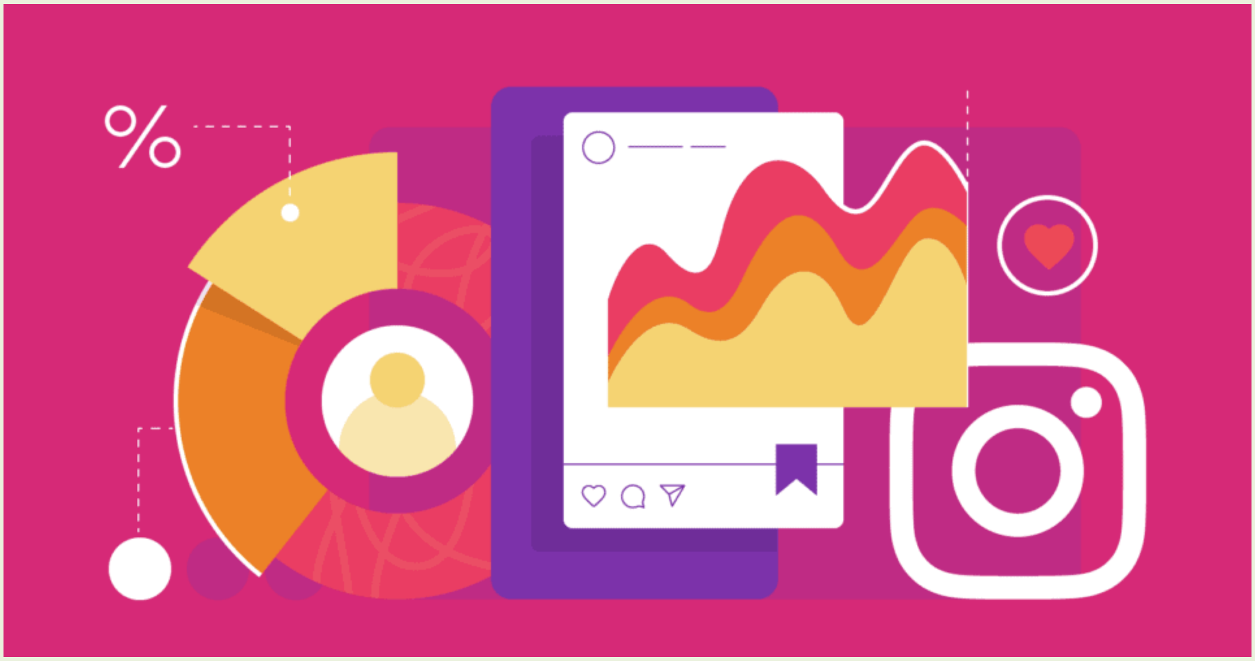
Mandatory Project

Instagram User Analytics



By T U Durga Prasad (DS-28C)

Instagram Analytics Process

Description:

The project's main goal is to offer insights into the available dataset. These insights are about giving the product, development, and marketing teams information on how users engage with the software or application so that they may evaluate the data and improve the software. I have given the team the necessary insights using MySQL queries, such as the top oldest users, the most popular hashtag, identifying bots or false accounts, etc.

Approach:

The data was carefully examined in order to identify the primary key and foreign key from different tables for joining the table. The data was then imported into tables in Workbench using Create Table and Insert Values.

Tech used:

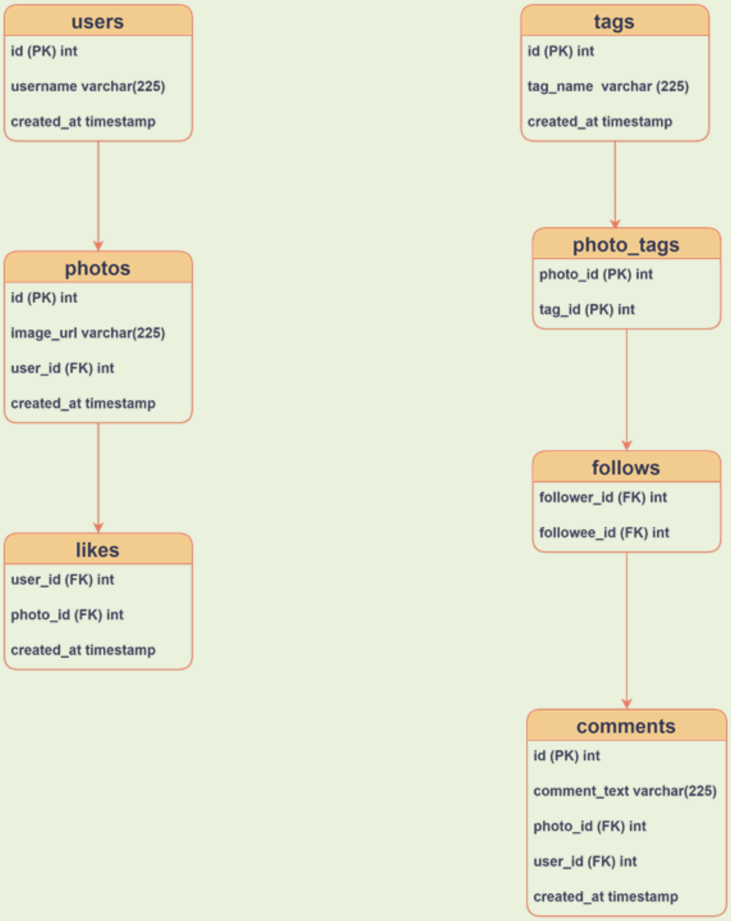
MySQL workbench

Insights:

By using basic functions like aggregation (sum, count), group by, order by the quires is written for getting a required insight. For desired outputs sometimes the tables have to been joined using inner join or left join.

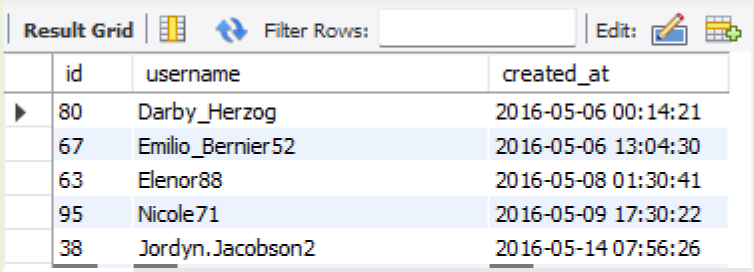
Tasks:

1. Create an ER diagram or draw a schema for the given database.



2. We want to reward the user who has been around the longest, Find the 5 oldest users.
   SELECT * FROM users ORDER BY created_at LIMIT 5;
   Function used: order by, Limit
   Order by: Helps in sorting the table ascending or descending.
   Limit: returns the records from the table according to the specified limit value.

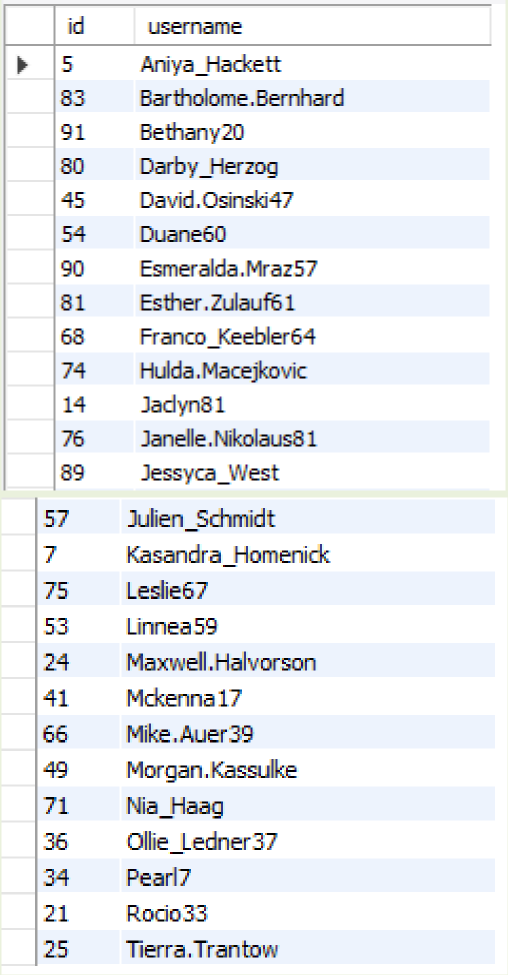
   Output:



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

3. To target inactive users in an email ad campaign, find the users who have never posted a photo.
   select users.id, username from users
   left join photos on users.id = photos.user_id
   where photos.id is null;
   Function used: Left join, we know that by using left join the null values can be found in second table.

   Output:



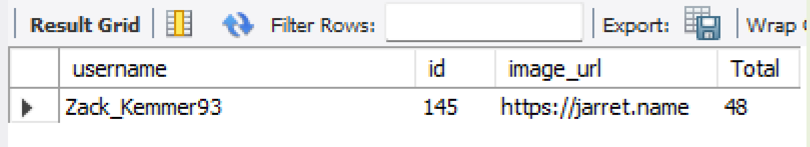
id	username
5	Aniya_Hackett
83	Bartholome.Bernhard
91	Bethany20
80	Darby_Herzog
45	David.Osinski47
54	Duane60
90	Esmeralda.Mraz57
81	Esther.Zulauf61
68	Franco_Keebler64
74	Hulda.Macejkovic
14	Jaclyn81
76	Janelle.Nikolaus81
89	Jessyca_West
57	Julien_Schmidt
7	Kasandra_Homenick
75	Leslie67
53	Linnea59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

4. Suppose you are running a contest to find out who got the most likes on a photo. Find out who won?
   select username, ph.id, ph.image_url, count(*) as Total From photos ph
   Inner Join likes l on ph.id=l.photo_id
   Inner Join users u on u.id= ph.user_id
   Group by ph.id
   Order by total desc
   Limit 1;
   Functions Used: Group by,order by ,inner join

Inner join: Combining the two tables

Group by: arranging the identical data into groups.

Output:

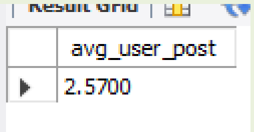


	username	id	image_url	Total
▶	Zack_Kemmer93	145	https://jarret.name	48

5. The investors want to know how many times does the average user post.

```
SELECT (SELECT COUNT(*)FROM photos)/(SELECT COUNT(*) FROM users)
 as avg_user_post;
```

Output:

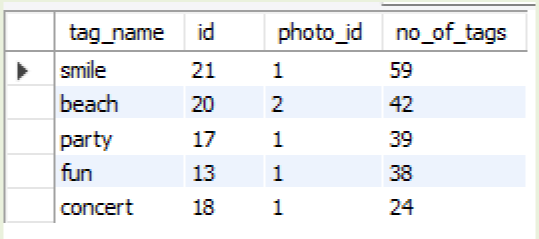


	avg_user_post
▶	2.5700

6. A brand wants to know which hashtag to use on a post, and find the top 5 most used hashtags.

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

Output:

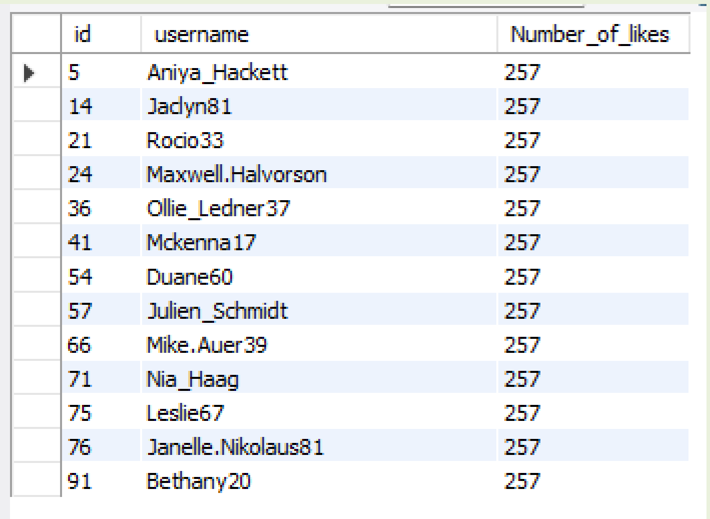


	tag_name	id	photo_id	no_of_tags
▶	smile	21	1	59
	beach	20	2	42
	party	17	1	39
	fun	13	1	38
	concert	18	1	24

7. To find out if there are bots, find users who have liked every single photo on the site.

```
select id,username, count(*) as Number_of_likes from users
INNER JOIN likes
on likes.user_id= users.id
Group By likes.user_id
Having Number_of_likes = (select Count(*) from photos);
```

Output:

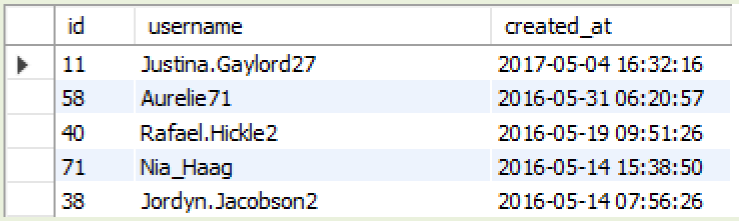


	id	username	Number_of_likes
▶	5	Aniya_Hackett	257
	14	Jaclyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna17	257
	54	Duane60	257
	57	Julien_Schmidt	257
	66	Mike.Auer39	257
	71	Nia_Haag	257
	75	Leslie67	257
	76	Janelle.Nikolaus81	257
	91	Bethany20	257

8. Find the users who have created Instagram id in may and select top 5 newest joinees from it?

```
SELECT * FROM users
WHERE MONTH(created_at) = 5
ORDER BY created_at DESC
LIMIT 5;
```

Output:

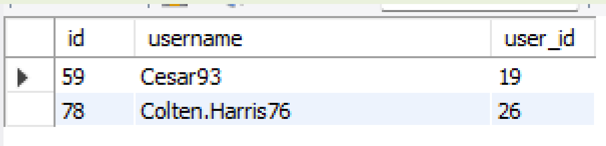


	id	username	created_at
▶	11	Justina.Gaylord27	2017-05-04 16:32:16
	58	Aurelie71	2016-05-31 06:20:57
	40	Rafael.Hickle2	2016-05-19 09:51:26
	71	Nia_Haag	2016-05-14 15:38:50
	38	Jordyn.Jacobson2	2016-05-14 07:56:26

9. Can you help me find the users whose name starts with c and ends with any number and have posted the photos as well as liked the photos?

```
select users.username, photos.user_id from users
inner join photos on users.id = photos.id
inner join likes on photos.user_id = likes.user_id
where username regexp '^C.*[0-9]$'
GROUP BY users.id, users.username
HAVING COUNT(DISTINCT photos.id) > 0 AND COUNT(DISTINCT likes.photo_id) > 0;
```

Output:

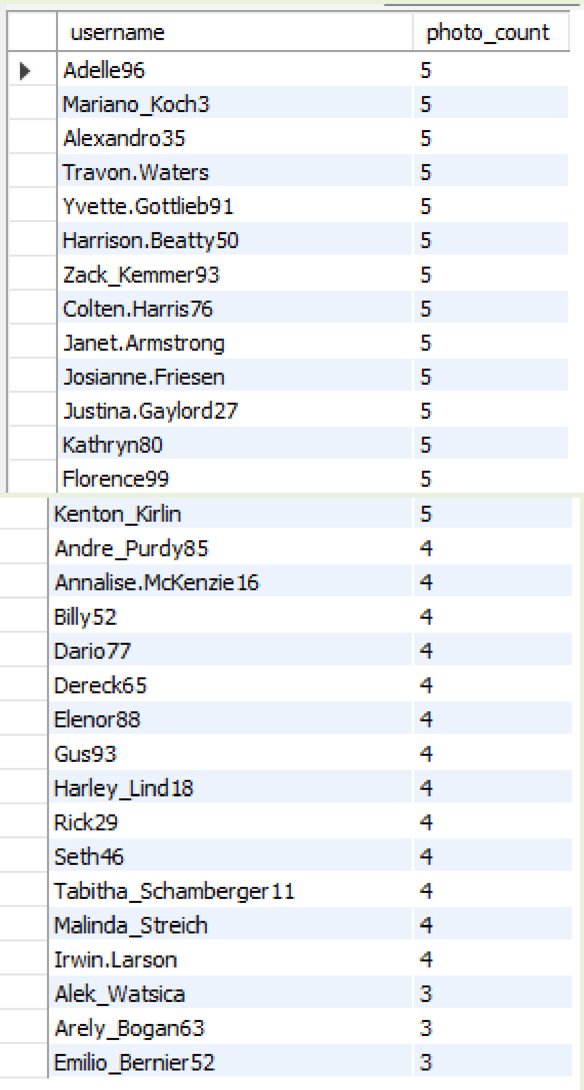


	id	username	user_id
▶	59	Cesar93	19
	78	Colten.Harris76	26

10. Demonstrate the top 30 usernames to the company who have posted photos in the range of 3 to 5.

```
SELECT u.username, COUNT(p.user_id) AS photo_count
FROM users u
JOIN photos p ON u.id = p.user_id
GROUP BY u.id, u.username
HAVING COUNT(p.user_id) BETWEEN 3 AND 5
ORDER BY photo_count DESC
LIMIT 30;
```

Output:



username	photo_count
Adelle96	5
Mariano_Koch3	5
Alexandro35	5
Travon.Waters	5
Yvette.Gottlieb91	5
Harrison.Beatty50	5
Zack_Kemmer93	5
Colten.Harris76	5
Janet.Armstrong	5
Josianne.Friesen	5
Justina.Gaylord27	5
Kathryn80	5
Florence99	5
Kenton_Kirlin	5
Andre_Purdy85	4
Annalise.McKenzie16	4
Billy52	4
Dario77	4
Dereck65	4
Elenor88	4
Gus93	4
Harley_Lind18	4
Rick29	4
Seth46	4
Tabitha_Schamberger11	4
Malinda_Streich	4
Irwin.Larson	4
Alek_Watsica	3
Arely_Bogan63	3
Emilio_Bernier52	3

Conclusion:

This project is all the performance of the social media app and analyzing the user data.