Project: Instagram User Analytics

(SQL Fundamentals)

Project Description:

As a data analyst working for Instagram product team, my role involved analyzing the raw data to derive insights which would help other departments (like marketing department) to make informed decisions which would help business to grow.

In this project, I have used SQL and MySQL Workbench as tools to analyze Instagram user data and answer the questions posed by the management team and Operations Team.

These insights will help the Product Manager, Marketing Manager and the rest of the team make informed decisions about the future direction of the Instagram app.

A) Marketing Analysis:

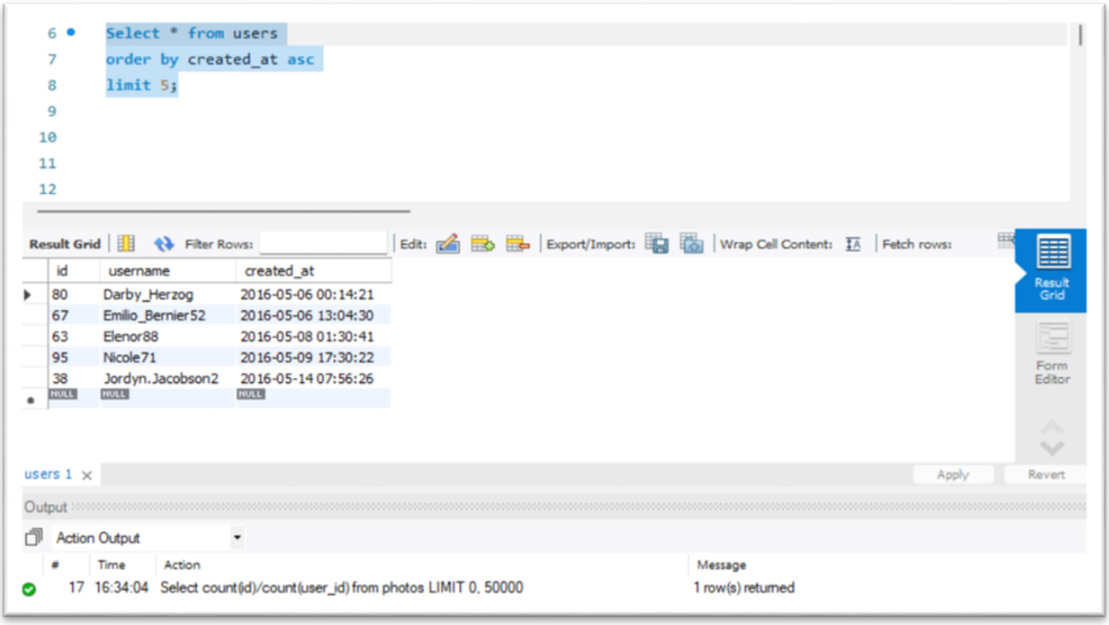
1) Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

Task: *Identify the five oldest users on Instagram from the provided database.*

Query:

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

Output:



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

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
NULL	NULL	NULL

```
#   Time      Action                                                 Message
17  16:34:04  Select count(id)/count(user_id) from photos LIMIT 0, 50000   1 row(s) returned
```

---

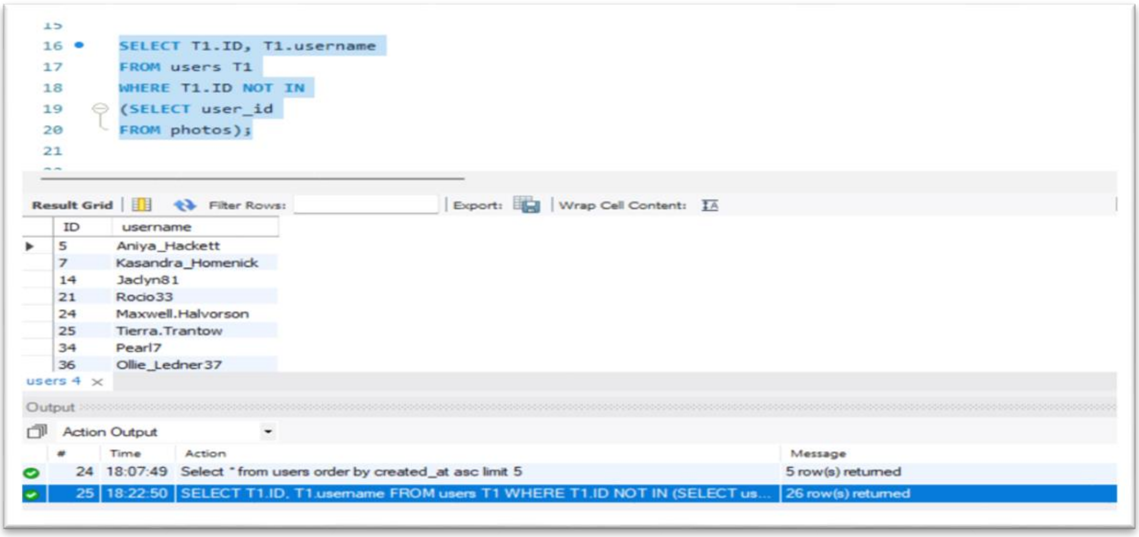
2) Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

Task: *Identify users who have never posted a single photo on Instagram.*

Query:

```
Select T1.id, T1.username
from users T1
where T1.id not in
(select user_id from photos);
```

Output:



```sql
SELECT T1.ID, T1.username
FROM users T1
WHERE T1.ID NOT IN
(SELECT user_id
FROM photos);
```

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

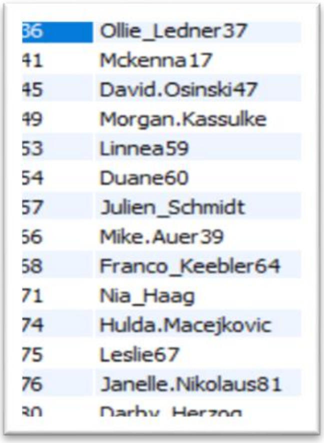
ID	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37

users 4 ×

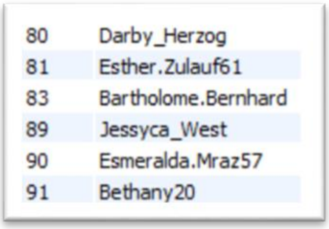
Output

Action Output

#	Time	Action	Message
24	18:07:49	Select * from users order by created_at asc limit 5	5 row(s) returned
25	18:22:50	SELECT T1.ID, T1.username FROM users T1 WHERE T1.ID NOT IN (SELECT us...	26 row(s) returned



36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
56	Mike.Auer39
58	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog



80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

---

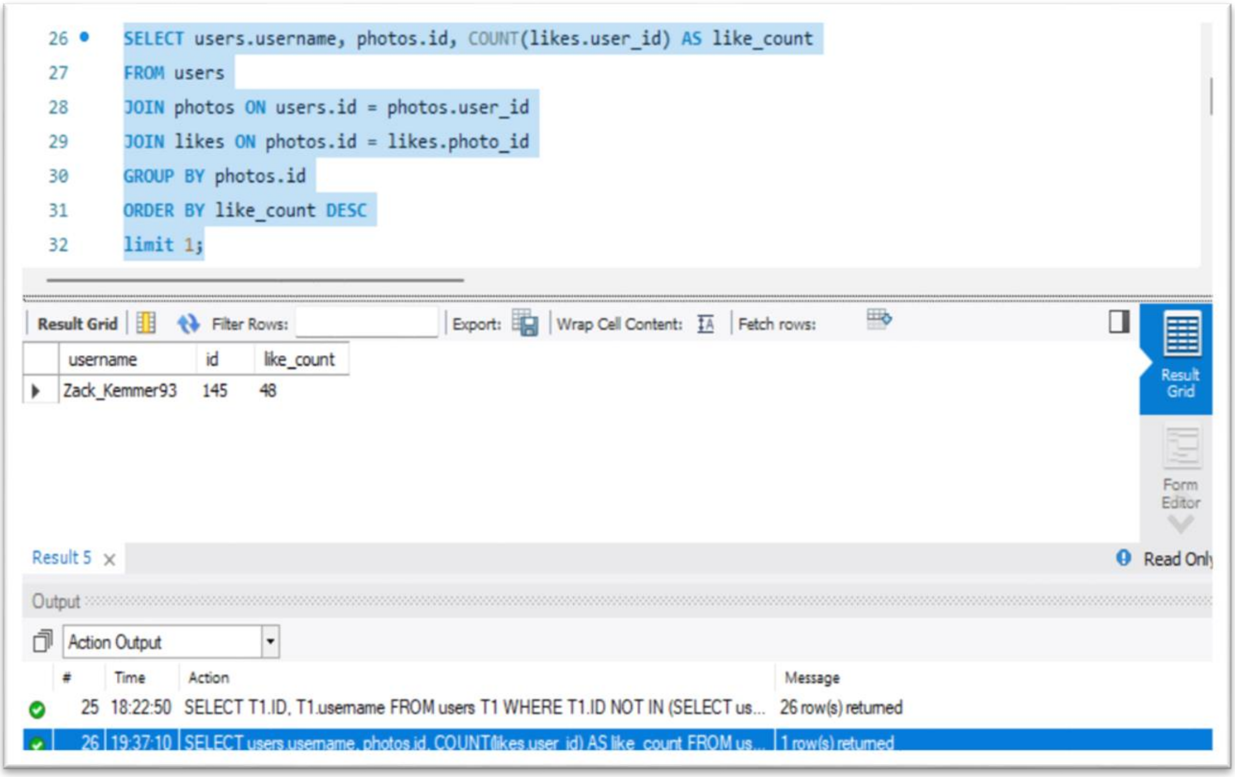
3) Contest Winner Declaration: The team has organized a contest where the user with <u>the most likes</u> on a single photo win.

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

Query:

```
Select users.username, photos.id, count(likes.user_id)
as like_count from users
Join photos on users.id = photos.user_id
Join likes on photos.id = likes.photo_id
Group by photos.id
Order by like_count desc
limit 1;
```

Output:



```
SELECT users.username, photos.id, COUNT(likes.user_id) AS like_count
FROM users
JOIN photos ON users.id = photos.user_id
JOIN likes ON photos.id = likes.photo_id
GROUP BY photos.id
ORDER BY like_count DESC
limit 1;
```

username	id	like_count
Zack_Kemmer93	145	48

#	Time	Action	Message
25	18:22:50	SELECT T1.ID, T1.username FROM users T1 WHERE T1.ID NOT IN (SELECT us...	26 row(s) returned
26	19:37:10	SELECT users.username, photos.id, COUNT(likes.user_id) AS like_count FROM us...	1 row(s) returned

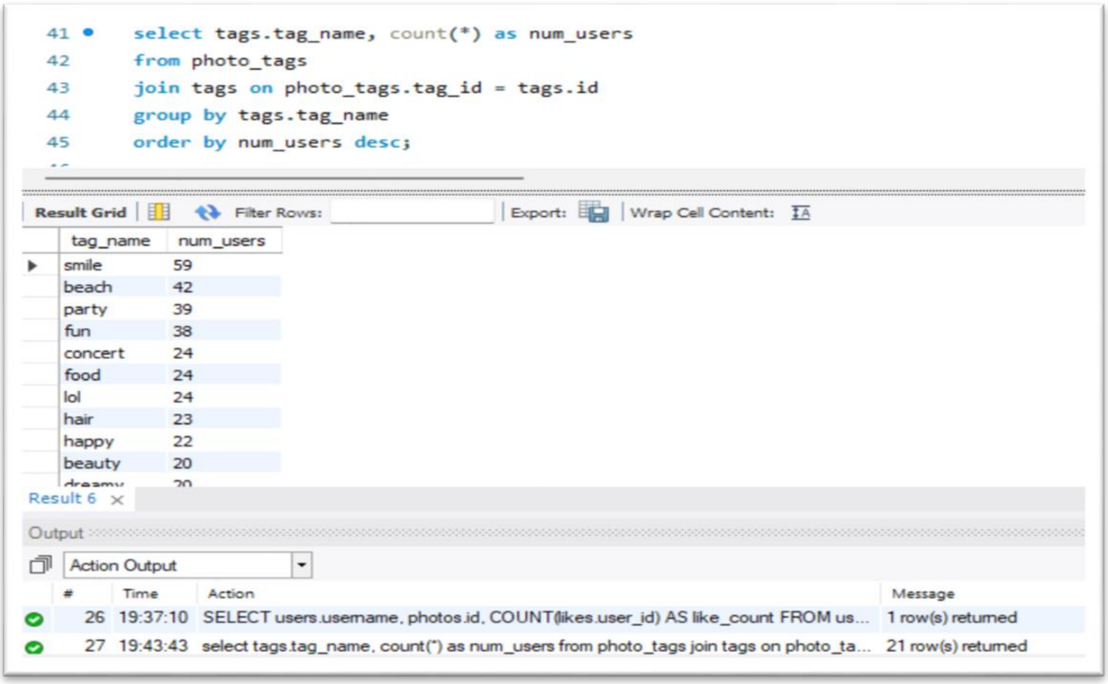
4) Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

***Task:** Identify and suggest the top five most used hashtags on the platform.*

Query:

```
Select tags.tag_name, count (*) as num_users
from photo_tags
join tags on photo_tags.tag_id = tags.id
group by tags.tag_name
order by num_users desc;
```

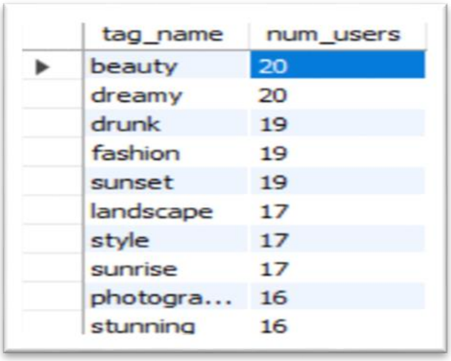
Output:



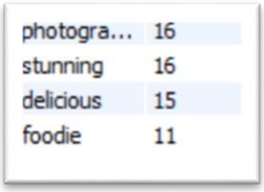
```
select tags.tag_name, count(*) as num_users
from photo_tags
join tags on photo_tags.tag_id = tags.id
group by tags.tag_name
order by num_users desc;
```

tag_name	num_users
smile	59
beach	42
party	39
fun	38
concert	24
food	24
lol	24
hair	23
happy	22
beauty	20
dreamy	20

#	Time	Action	Message
26	19:37:10	SELECT users.username, photos.id, COUNT(likes.user_id) AS like_count FROM us...	1 row(s) returned
27	19:43:43	select tags.tag_name, count(*) as num_users from photo_tags join tags on photo_ta...	21 row(s) returned



tag_name	num_users
beauty	20
dreamy	20
drunk	19
fashion	19
sunset	19
landscape	17
style	17
sunrise	17
photogra...	16
stunning	16



photogra...	16
stunning	16
delicious	15
foodie	11

5) **Ad Campaign Launch:** The team wants to know the best day of the week to launch ads.

Task: *Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.*

Query:

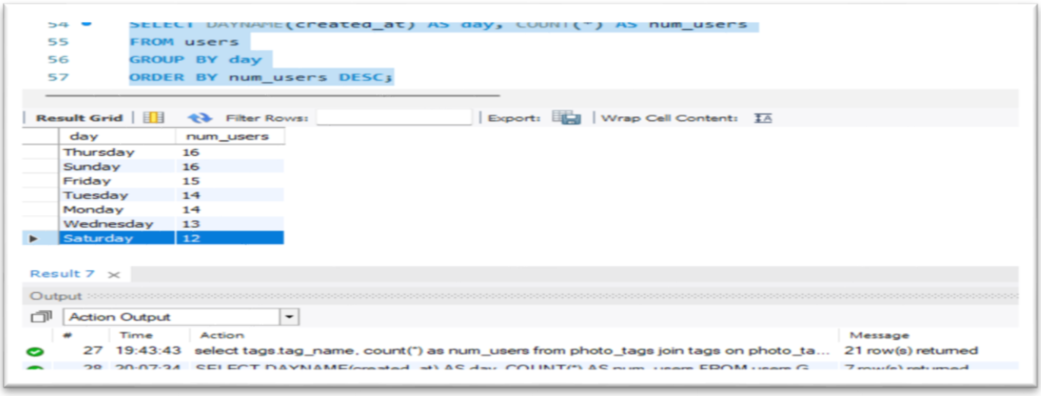
Select dayname (created_at) AS day, count (*) AS num_users

from users

Group by day

Order by num_users desc;

Output:



```
SELECT DAYNAME(created_at) AS day, COUNT(*) AS num_users
FROM users
GROUP BY day
ORDER BY num_users DESC;
```

day	num_users
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

```
#   Time      Action                                                                              Message
27  19:43:43  select tags.tag_name, count(*) as num_users from photo_tags join tags on photo_ta...  21 row(s) returned
28  20:07:24  SELECT DAYNAME(created_at) AS day, COUNT(*) AS num_users FROM users G...              7 row(s) returned
```

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.

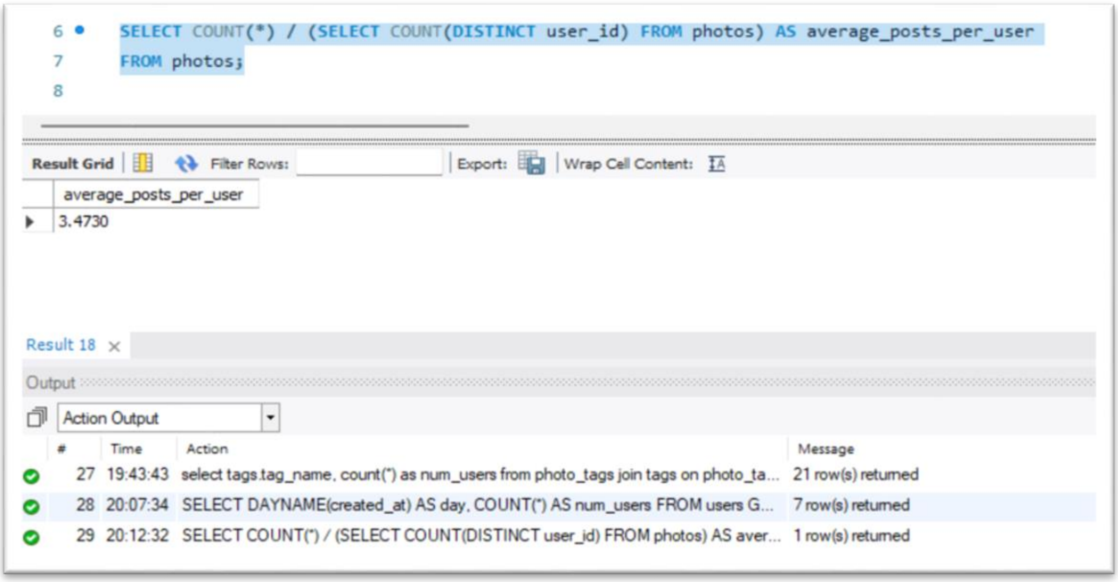
Task: *Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.*

- The average number of posts per user on Instagram

Query:

```
Select count (*) / (Select count(distinct user_id)
        from photos) as average_posts_per_user from photos;
```

Output:



```
SELECT COUNT(*) / (SELECT COUNT(DISTINCT user_id) FROM photos) AS average_posts_per_user
FROM photos;
```

```
average_posts_per_user
3.4730
```

```
#   Time      Action                                                                  Message
27  19:43:43  select tags.tag_name, count(*) as num_users from photo_tags join tags on photo_ta...  21 row(s) returned
28  20:07:34  SELECT DAYNAME(created_at) AS day, COUNT(*) AS num_users FROM users G...      7 row(s) returned
29  20:12:32  SELECT COUNT(*) / (SELECT COUNT(DISTINCT user_id) FROM photos) AS aver...      1 row(s) returned
```

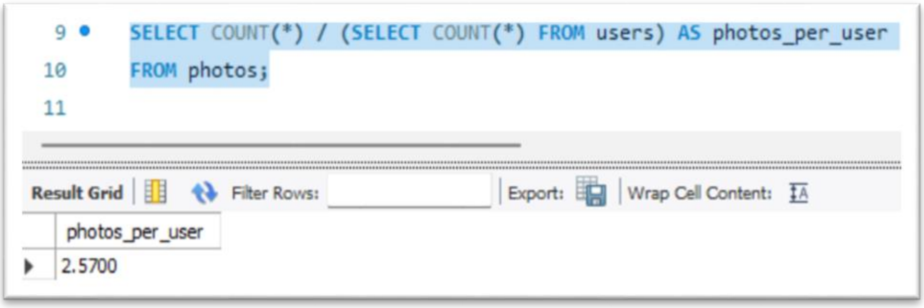
- The total number of photos on Instagram divided by the total number of users.

  <u>Query:</u>

  Select count (*) / (select count (*) FROM users) as

  photos_per_user from photos;

  <u>Output:</u>



```
SELECT COUNT(*) / (SELECT COUNT(*) FROM users) AS photos_per_user
FROM photos;
```

photos_per_user
2.5700

---

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

***Task**: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user*

<u>Query:</u>

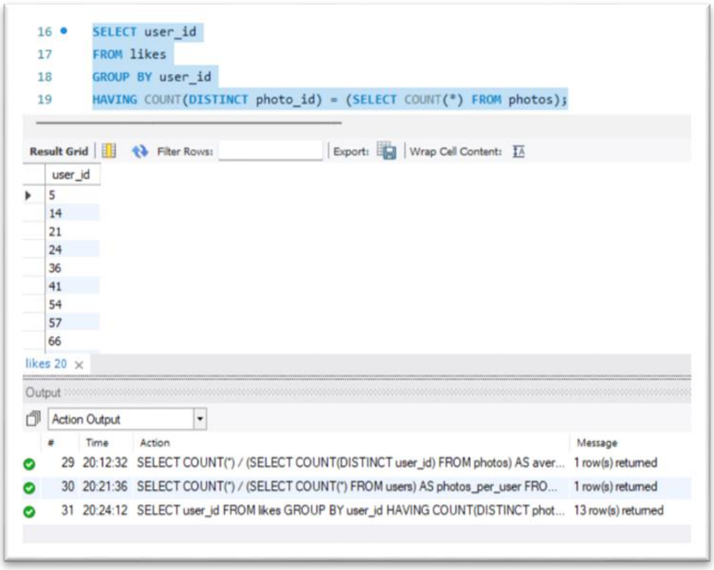
Select user_id

From likes

Group by user_id

Having count(distinct photo_id) = (Select count(*) from photos);

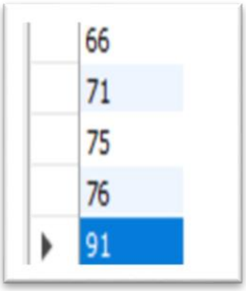
<u>Output:</u>



```sql
SELECT user_id
FROM likes
GROUP BY user_id
HAVING COUNT(DISTINCT photo_id) = (SELECT COUNT(*) FROM photos);
```

```
user_id
5
14
21
24
36
41
54
57
66
```

```
#   Time      Action                                                                   Message
29  20:12:32  SELECT COUNT(*) / (SELECT COUNT(DISTINCT user_id) FROM photos) AS aver...  1 row(s) returned
30  20:21:36  SELECT COUNT(*) / (SELECT COUNT(*) FROM users) AS photos_per_user FRO...    1 row(s) returned
31  20:24:12  SELECT user_id FROM likes GROUP BY user_id HAVING COUNT(DISTINCT phot...   13 row(s) returned
```



```
66
71
75
76
91
```

Tech-Stack Used:

MySQL Workbench: MySQL Workbench is a cross-platform tool, meaning it's available for Windows, macOS, and Linux. This can be advantageous if you work in environments with multiple operating systems.

Insights Derived:

1)Users Metrics:

- ♦ **Number of Users:**
  – From 2016 May to 2017 May total 100 users have joined Instagram

- ♦ **Users Joined:**
  – Maximum Number of users have joined in Month of October.
  – Maximum Number of users have joined on Thursday and Sunday.

- **Bot Detection:**
  - Unusual patterns in activity, such as excessive posting or repetitive interactions, may indicate bot behavior.
  - There are 20 users who like each photo posted on Instagram which is not possible for real users to do.

---

2)Engagement Metrics:

- **Number of Photos Posted:**
  - Total No. of Photos Posted in year 2016-2017 = 257 Photos
  - Average No. of Photos Posted per Users = 2.57 Photos

- **Top 5 Users posted most photos on Instagram:**

User_Id	Username	No. of Photos Posted
23	Eveline95	12
88	Aurelie71	11
59	Cesar93	10
86	Delfina_VonRueden68	9
58	Clint27	8

- **Top 5 Images which got max no. of likes**

Photo_id	Image_url	No. of Likes	Tags Used
145	http://kenny.com	48	Fun, Party, Concert, Drunk, Smile
127	http://shannon.org	43	Smile
182	https://celestine.name	43	Beauty, Dreamy, Beach, Smile
123	https://jarret.name	42	Style, Hair, party

30	https://dorcas.biz	41	Food, Foodie, delicious

♦ **Top 5 mostly used Tags:**

Tag name	No. of times used
smile	59
beach	42
party	39
fun	38
concert	24

We can conclude that mostly people take photos when they are in social gathering or in beach which is joyful and social atmosphere

---

Conclusion:

The analysis provided a comprehensive understanding of the user's Instagram accounts. This Project will help the Marketing Team and Operations team to make well informed decisions that can further enhance Performance Metrics, increase User engagement and a more targeted approach to content creation.