Instagram User Analytics

Project Description:

In order to help the Instagram product team and the product manager answer the questions posed by the management team, it is my project to find out the analytics and provide insights of the given dataset to the management team. MySQL Workbench was used to complete this project. For this project, I will need to find the 5 oldest users to find the users who have never posted a single photo on Instagram, the contest winner their details, the top 5 most commonly used hashtags on the platform, the day of the week do most users register on, average user posts on Instagram along with the total number of photos on Instagram/total number of users and finally, users (bots) who have liked every single photo.

Approach:

To approach this project, I began by fully comprehending the requirements and objectives of this project. Following this, I began executing the project with an adaptable mindset, exploring diverse approaches to accomplish the task. After evaluating various options, I selected the most suitable approach to continue with the project. I then proceeded to write and implement the queries to produce the correct, comprehensible output. Overall, I am confident with the project's outcome and consider it a success.

Tech-Stack Used:

MySQL Workbench with the latest released Version 8.0 Community Edition.

I used it because MySQL Workbench is a visual database design tool that combines SQL development, administration, database design, creation, and maintenance into a single development environment.

Insights:

I surmise that the dataset contains a lot of data, and that it is necessary to analyse it before continuing. I also infer that the data is distributed in a specific way, and that understanding the distribution is required before proceeding.

A. Marketing:

1. Rewarding Most Loyal Users:

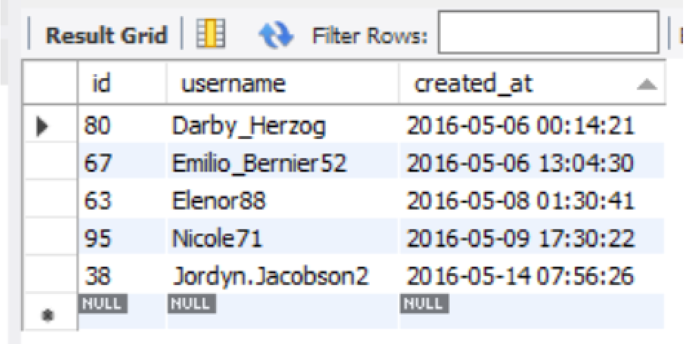
Here, I need to find the 5 oldest users of the Instagram from the given database.

For fulfilling the required query, I selected id, username, created_at column from the users table and order by created_at column in ascending order with limit of 5

SQL Statement :

```
SELECT
        id, username, created_at
FROM
        users
ORDER BY created_at ASC
LIMIT 5;
```

Result:



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

Conclusion:

Users having IDs 80, 67, 63, 95, and 38 are the oldest users on the platform.

2. Remind Inactive Users to Start Posting:

Here, I need to find the users who have never posted a single photo on Instagram.

For fulfilling the required query, I selected username from users table, then I used left join function to join users table with photos table on user.id = photos.user_id, then I used where function with condition photos.id is null.

SQL Statement:

```
SELECT
        username
FROM
        users
LEFT JOIN photos
        ON users.id = photos.user_id
WHERE photos.id IS NULL;
```

Result – User List

1	username
2	Aniya_Hackett
3	Kasandra_Homenick
4	Jaclyn81
5	Rocio33
6	Maxwell.Halvorson
7	Tierra.Trantow
8	Pearl7
9	Ollie_Ledner37
10	Mckenna17
11	David.Osinski47
12	Morgan.Kassulke
13	Linnea59
14	Duane60
15	Julien_Schmidt
16	Mike.Auer39
17	Franco_Keebler64
18	Nia_Haag
19	Hulda.Macejkovic
20	Leslie67
21	Janelle.Nikolaus81
22	Darby_Herzog
23	Esther.Zulauf61
24	Bartholome.Bernhard
25	Jessyca_West
26	Esmeralda.Mraz57
27	Bethany20

Conclusion:

This is list of users(usernames) who never posted a single photo on Instagram is given in the user list below. After seeing this data, we can conclude that a total of 26 people present in the database never posted any photos on the platform.

3. Declaring Contest Winner:

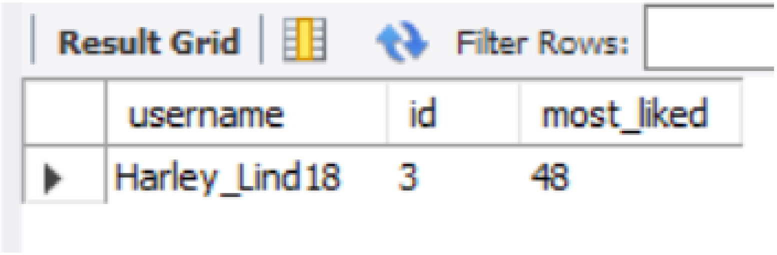
Here, I need to identify the contest winner and provide their details to the team.

For that, I have selected (username, id) from users, count(*) as most_liked from likes table the I use inner join function to join likes table with users table on users ON users.id=likes.user_id then I use group by on photo_id and then I use order by function on most_liked in descending order with limit 1.

SQL Statement:

```
SELECT
        users.username, users.id, COUNT(*) AS most_liked
FROM
        likes
INNER JOIN
users ON users.id = likes.user_id
GROUP BY photo_id
ORDER BY most_liked DESC
LIMIT 1;
```

Result:



username	id	most_liked
Harley_Lind18	3	48

Conclusion:

We learned that photo id 3, which gets the most likes of 48, belongs to the user ‘Harley_Lind18’. So, Harley_Lind18 is the winner of the contest.

4. Hashtag Resercher:

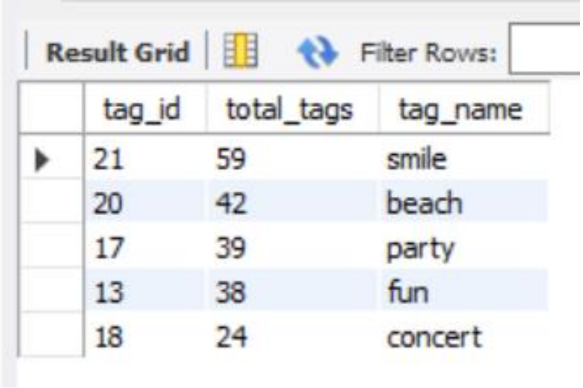
Here, I need to Identify and suggest the platform's top 5 most commonly used hashtags.

For that, I have selected tag_id, count (photo_id) as total_tags from photo tags table, and then I used join function to join tags table with photo tags table on
tags.id = photo_tags.tag_id, I used group by function on tag.id, and then I used order by function on total in descending order with limit 5.

SQL Statement:

```
SELECT
tag_id, COUNT(photo_id) AS total_tags,
tag_name FROM photo_tags
JOIN tags ON tags.id = photo_tags.tag_id
GROUP BY tag_id
ORDER BY total_tags DESC
LIMIT 5;
```

Result:



tag_id	total_tags	tag_name
21	59	smile
20	42	beach
17	39	party
13	38	fun
18	24	concert

Conclusion:

The top 5 popular tags used on Instagram are – smile, beach, party, fun, and concert, with total tag counts of 59, 42, 39, 38, 24, respectively.

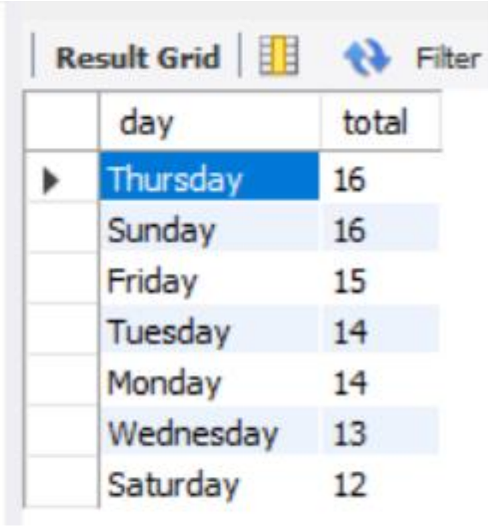
5. **Launch AD Campaign:**

Here, I need to Identify on what day of the week do most users register on and provide insights to the management team on when to schedule an ad campaign.

SQL Statement:

```
SELECT

DAYNAME(created_at) AS day,

COUNT(*) AS total

FROM users

GROUP BY day

ORDER BY total DESC;
```

Result:



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

CONCLUSION:

With the above query, we learned that most users (16) registered on Thursday and Sunday. So, either is the best day to schedule an ad campaign. Since Thursday is on top, we suggest moving forward with Thursday as the best day to schedule an ad campaign.

B) Investor Metrics:

1. User Engagement:

Here I need to find how many times does an average user posts on Instagram, and provide the total number of photos on Instagram/total number of users.

This problem can be solved in two parts:

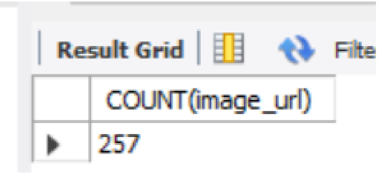
1.1. 1st Part: Finding the total no. of photos and total no. of users on the platform.

Here, I selected count command to count the numbers of images form the table photos.

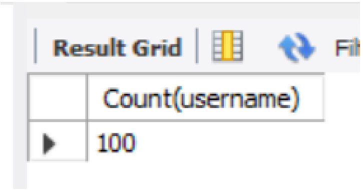
SQL statement:

```
SELECT COUNT(image_url)
FROM photos;

SELECT Count(username)
FROM users;
```

Result:



	COUNT(image_url)
▶	257



	Count(username)
▶	100

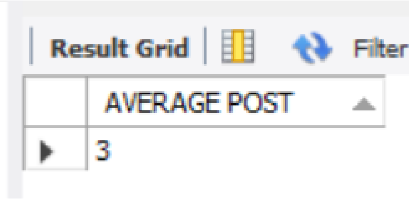
1.2. 2nd Part: Finding the average user post on the platform.

Here, I selected Floor command returns the largest integer value that is smaller than or equal to a number. Then I proceeded with count command of the distinct ids from the photos table and divided with the another selected another count command of distinct user_id from table photos and passed it to show output as average post.

SQL statement:

```
SELECT FLOOR (
(SELECT COUNT(DISTINCT id) FROM photos)
/
(SELECT COUNT(DISTINCT user_id) FROM photos))
AS "AVERAGE POST";
```

Result:



	AVERAGE POST
▶	3

Conclusion:

From the above result, we can conclude that an average posts per user is around 3.

2. **Bots & Fake Accounts:**

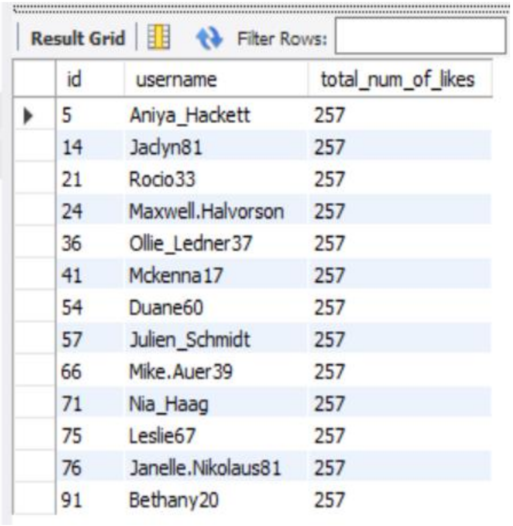
Here, I need to find and provide data on users (bots) who have liked every single photo on the site.

For that, I first selected user.id, username, and count function as total_likes_by_user from users table, then I used join function to join users table with likes table on users.id=likes.user_id, later I used group by function for user.id and having function with condition total_likes_by_user = select statement for count (*) from photos table.

SQL statement:

```
SELECT
users.id, username, COUNT(users.id) AS total_num_of_likes
FROM users
JOIN
        likes ON users.id = likes.user_id
GROUP BY users.id HAVING total_num_of_likes = (SELECT
COUNT(*)
FROM photos);
```

Result:



id	username	total_num_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

Conclusion:

There is a total of 257 photos available on the platform, and from the above result we can conclude that around 13 users come in the category of bots as they have liked every single picture available on the platform.

Result:

While working on the project on Instagram user analytics, I made significant progress. Understanding the fundamentals of SQL and learning how to use it effectively have been very helpful to me. I was able to examine Instagram users' data and their posts. This has made it easier for me to comprehend how Instagram users behave and interact with one another. It has significantly aided me in developing my abilities and knowledge. I now have a better understanding of how to analyse data using SQL. I've also improved my ability to concisely and clearly explain my findings to others. I believe that this project has greatly benefited me and improved my communication skills.

Drive Link:

https://drive.google.com/file/d/1ei9ds8efr8mhSrL1zeFtZ_IsoGPK9g2u/view?usp=sharing