INSTAGRAM USER ANALYTICS

SQL Fundamentals

PROJECT DESCRIPTION

- This project helps to analyses the raw data to create useful insights. Various database management tools can be used to extract useful insights and even visualize them. This enables a way to increase efficiency of a platform.
- It helps us to know the interaction of the users with the product and calculates the success of the product.
- With the database given, we can find every solution to the problems or questions asked by the client using SQL.

APPROACH

- We have SQL to execute the given problems in this project.
- We have to create database using SQL queries.
- Using the database, we can use DDL (Data Definition Language), DCL (Data Control Language), DML (Data Manipulation Language) to create table and insert values to it, etc.
- Later we can use Sorting Functions, Aggregate Functions, Operators or Joins to create a query which solves the problems.

TECH-STACK USED

MySQL Workbench v8.0.30.0 was used during project execution in order to query the database. I have MySQL workbench Client to create database and get the output to the queries. The ease of access and setup, troubleshooting support as well as the GUI made it a good tool for the project.

INSIGHTS

- Using the given database, we can use DDL (Data Definition Language), DCL (Data Control Language), DML (Data Manipulation Language) to create table and insert values to it, etc.
- DML is used to insert, delete and update the table, DCL is used to Grant or revoke, DDL is used to Create, Drop, Alter, Truncate the data to the table.
- Later we can use Sorting Functions, Aggregate Functions, Operators or Joins to create a query which solves the problems.
- Sorting Functions include GROUP BY, HAVING, LAST, LIMIT, TOP.
- Aggregate Function includes MIN, MAX, COUNT, AVG, SUM.
- Operators include Arithmetic operators (+, -, *, /, %), Comparison operators (==,!=, <=, >=, <, >), Logical operators (AND, ALL, ANY, EXISTS, IN, NOT, OR).

QUESTIONS:

A) Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

1. **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.
   Your Task: Find the 5 oldest users of the Instagram from the database provided.

   Query:

   SELECT * FROM users

   ORDER BY created_at ASC

   LIMIT 5;

```
100 rows in set (0.12 sec)

mysql> 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 |
+----+------------------+---------------------+
5 rows in set (0.07 sec)

mysql>
```

2. **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.
   Your Task: Find the users who have never posted a single photo on Instagram.

   Query:

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

```
mysql> select username from users left join photos on users.id=photos.user_id where photos.id is null;
+---------------------+
| username            |
+---------------------+
| Aniya_Hackett       |
| Kasandra_Homenick   |
| Jaclyn81            |
| Rocio33             |
| Maxwell.Halvorson   |
| Tierra.Trantow      |
| Pearl7              |
| Ollie_Ledner37      |
| Mckenna17           |
| David.Osinski47     |
| Morgan.Kassulke     |
| Linnea59            |
| Duane60             |
| Julien_Schmidt      |
| Mike.Auer39         |
| Franco_Keebler64    |
| Nia_Haag            |
| Hulda.Macejkovic    |
| Leslie67            |
| Janelle.Nikolaus81  |
| Darby_Herzog        |
| Esther.Zulauf61     |
| Bartholome.Bernhard |
| Jessyca_West        |
| Esmeralda.Mraz57    |
| Bethany20           |
+---------------------+
26 rows in set (0.00 sec)
```

3. **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
   Your Task: Identify the winner of the contest and provide their details to the team

   Query:

   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=likes.photo_id GROUP BY photos.id ORDER BY total_likes DESC LIMIT 1;

```
mysql> 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=likes.photo_id GROUP BY photos.id ORDER BY total_likes D
ESC LIMIT 1;
+----------+----+----------------+-------------+
| username | id | image_url      | total_likes |
+----------+----+----------------+-------------+
| Kaley9   | 30 | http://kenny.com |          41 |
+----------+----+----------------+-------------+
1 row in set (0.01 sec)

mysql>
```

   So, the winner of the contest is Kaley9.

   If u want to find the remaining winners of the competition or top 15 people we can use the following query.

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=likes.photo_id GROUP BY photos.id ORDER BY total_likes DESC LIMIT 15;

```
mysql> 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=lik
es.photo_id group by photos.id order by total_likes desc limit 15;
+--------------------+----+----------------------+-------------+
| username           | id | image_url            | total_likes |
+--------------------+----+----------------------+-------------+
| Zack_Kemmer93      | 52 | https://hershel.com  |          41 |
| Jayson65           | 61 | https://dejon.name   |          41 |
| Kaley9             | 30 | http://kenny.com     |          41 |
| Alexandro35        | 13 | https://fred.com     |          40 |
| Ressie_Stanton46   | 62 | https://rigoberto.net|          39 |
| Seth46             | 44 | http://golden.org    |          39 |
| Mike.Auer39        | 66 | http://lionel.net    |          39 |
| Harley_Lind18      |  3 | http://vicky.biz     |          38 |
| Arely_Bogan63      |  4 | http://oleta.net     |          38 |
| Eveline95          | 23 | http://madison.net   |          38 |
| Karley_Bosco       | 69 | http://harvey.info   |          38 |
| Kasandra_Homenick  |  7 | https://selina.name  |          38 |
| Mckenna17          | 41 | https://adella.net   |          38 |
| Yazmin_Mills95     | 37 | http://camille.name  |          38 |
| Elenor88           | 63 | http://megane.biz    |          38 |
+--------------------+----+----------------------+-------------+
15 rows in set (0.01 sec)

mysql>
```

4. **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
   Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform

   Query:

   SELECT tags.tag_name,count(*) AS hashtag_count FROM tags JOIN photo_tags
   ON tags.id=photo_tags.tag_id
   GROUP BY tags.tag_name ORDER BY hashtag_count
   DESC LIMIT 5;

```
mysql> select tags.tag_name,count(*) as hashtag_count from tags join photo_tags on tags.id=photo_tags.tag_id group by tags.tag_name order by hashtag_count DES
C limit 5;
+----------+---------------+
| tag_name | hashtag_count |
+----------+---------------+
| smile    |            59 |
| beach    |            42 |
| party    |            39 |
| fun      |            38 |
| concert  |            24 |
+----------+---------------+
5 rows in set (0.00 sec)

mysql>
```

5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs.

Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

Query:

SELECT MAX(WEEKDAY(created_at)) AS day FROM users;

```
mysql> select  max(weekday(created_at)) as day from users;
+------+
| day  |
+------+
|    6 |
+------+
1 row in set (0.00 sec)

mysql>
```

The WEEKDAY() function returns the weekday number for a given date.
Note: 0 = Monday, 1 = Tuesday, 2 = Wednesday, 3 = Thursday, 4 = Friday, 5 = Saturday, 6 = Sunday.

So, it's ***SUNDAY*** where most users register on and it is the day to schedule an ad campaign.

B) Investor Metrics: Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

1. **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts

   Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.

Note: This is the average of all the posts in the Instagram posted by users.

Query:

SELECT ROUND((SELECT COUNT(*)FROM photos)/(SELECT COUNT(*) FROM users),2) AS avg_user_posts;

```
mysql> SELECT ROUND((SELECT COUNT(*)FROM photos)/(SELECT COUNT(*) FROM users),2) as avg_user_posts;
+----------------+
| avg_user_posts |
+----------------+
|           2.57 |
+----------------+
1 row in set (0.00 sec)

mysql>
```

Note: this is the average of all posts based on users (grouped by username) in the Instagram.

Query:

SELECT users.username,AVG(photos.id) AS average_posts FROM photos JOIN users ON users.id=photos.user_id GROUP BY users.username;

```
mysql> select users.username,avg(photos.id) as average_posts from photos join users on users.id=photos.user_id group by users.username;
+----------------------+---------------+
| username             | average_posts |
+----------------------+---------------+
| Kenton_Kirlin        |        3.0000 |
| Andre_Purdy85        |        7.5000 |
| Harley_Lind18        |       11.5000 |
| Arely_Bogan63        |       15.0000 |
| Travon.Waters        |       19.0000 |
| Tabitha_Schamberger11|       23.5000 |
| Gus93                |       27.5000 |
| Presley_McClure      |       31.0000 |
| Justina.Gaylord27    |       35.0000 |
| Dereck65             |       39.5000 |
| Alexandro35          |       44.0000 |
| Billy52              |       48.5000 |
| Annalise.McKenzie16  |       52.5000 |
| Norbert_Carroll35    |       56.0000 |
| Odessa2              |       58.0000 |
| Hailee26             |       59.5000 |
| Delpha.Kihn          |       61.0000 |
| Kenneth64            |       62.0000 |
| Eveline95            |       68.5000 |
| Josianne.Friesen     |       77.0000 |
| Darwin29             |       80.0000 |
| Dario77              |       82.5000 |
| Jaime53              |       88.5000 |
| Kaley9               |       93.5000 |
| Aiyana_Hoeger        |       95.0000 |
| Irwin.Larson         |       97.5000 |
| Yvette.Gottlieb91    |      102.0000 |
| Lennie_Hartmann40    |      105.5000 |
| Yazmin_Mills95       |      107.0000 |
| Jordyn.Jacobson2     |      108.5000 |
| Kelsi26              |      110.0000 |
| Rafael.Hickle2       |      111.0000 |
| Maya.Farrell         |      113.0000 |
| Janet.Armstrong      |      117.0000 |
| Seth46               |      121.5000 |
| Malinda_Streich      |      125.5000 |
| Harrison.Beatty50    |      130.0000 |
| Granville_Kutch      |      133.0000 |
| Gerard79             |      135.0000 |
| Mariano_Koch3        |      139.0000 |
| Zack_Kemmer93        |      144.0000 |
| Meggie_Doyle         |      147.0000 |
| Peter.Stehr0         |      148.0000 |
| Aurelie71            |      152.5000 |
```

Note: total number of photos on Instagram.

Query:

SELECT COUNT(*) FROM photos;

```
mysql> SELECT COUNT(*)FROM photos;
+----------+
| COUNT(*) |
+----------+
|      257 |
+----------+
1 row in set (0.00 sec)
```

Note: total number of users.

Query:

SELECT COUNT(*) FROM users;

```
mysql> SELECT COUNT(*)FROM users;
+----------+
| COUNT(*) |
+----------+
|      100 |
+----------+
1 row in set (0.00 sec)
```

2. **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts
   Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

Query:

SELECT users.username,users.id,COUNT(user_id) AS total_likes FROM users JOIN photos ON users.id=photos.user_id JOIN likes on likes.user_id=users.id and likes.photo_id=photos.id group by users.id HAVING total_likes=(SELECT COUNT(*) FROM photos);

According to the database I have created the bots are empty.

```
mysql> select users.username,users.id,count(users.id) as total_likes from users join photos on users.id=photos.user_id join
likes on likes.user_id=users.id and likes.photo_id=photos.id group by users.id having total_likes=(select count(*) from phot
os);
Empty set (0.01 sec)
```

RESULT:

This project helped me to learn the methods on how to solve the query. I have learnt some new ways to solve the problems. Even though it took more time to solve all the problems as a beginner I'm sure I can solve some queries easily now.

THANK YOU

DONE BY

JEEVIKA K