PROJECT-2

Instagram User Analytics

Submitted by Shiksha Lahre

This project is made with the help of a dataset provided in the assignment. We used SQL and MySQL Workbench 8.0 to run the queries on the dataset to extract the required information regarding marketing analysis and investor matrices.

Project Description: This project is about extracting some significant information by examining Instagram User Analytics. We have been provided with a dataset, where Users information such as, date of account created, photos liked by them/of them, their comments, tags used, etc. have been listed. Purpose of this project is to track user engagement, and getting valuable insights from them, and then answering those questions raised by the management team.

Tech-Stack used: MySQL Workbench 8.0 CE. I have used this platform due to its diverse advantages and functions:
1) Create and manipulate models
2) Create models from a target database or import sql files
3) Convert ER Diagrams into sql statements and push them to a sql server
4) Create and edit tables and insert data
5) Database migration, Performance monitoring

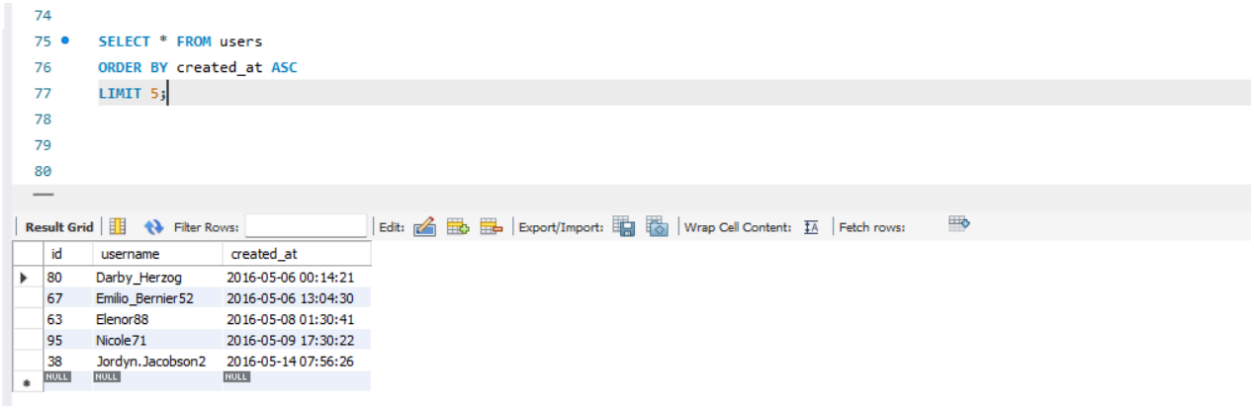
Approach: I have carefully examined the tables created in the database. They consist of data of:

- 100 instagram users
- 257 image URLs
- 1000 comments made
- Number of likes
- Followers and followee
- Types of tags

Thus, an approach for answering every question is possible with different statements and functions used in SQL.

MARKETING ANALYSIS

1. **User Engagement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
   Your 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.



```
74
75 •  SELECT * FROM users
76    ORDER BY created_at ASC
77    LIMIT 5;
78
79
80
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows:

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

Approach: Here, to line up the earliest dates, we used ORDER BY command, where we fetch earliest dates. We also used a LIMIT restriction to get only 5 data.

2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
Your Task: Identify users who have never posted a single photo on Instagram.

Approach: We used subquery here, where collecting distinct user ids who have never posted a single picture from table PHOTOS. This was an independent query. Then we connected these user ids to table USERS and fetched all the person's names.

```sql
select *
from users
where id not in (select distinct user_id from photos);
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

id	username	created_at
5	Aniya_Hackett	2016-12-07 01:04:39
7	Kasandra_Homenick	2016-12-12 06:50:08
14	Jadyn81	2017-02-06 23:29:16
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
25	Tierra.Trantow	2016-10-03 12:49:21
34	Pearl7	2016-07-08 21:42:01
36	Ollie_Ledner37	2016-08-04 15:42:20
41	Mckenna17	2016-07-17 17:25:45
45	David.Osinski47	2017-02-05 21:23:37
49	Morgan.Kassulke	2016-10-30 12:42:31
53	Linnea59	2017-02-07 07:49:34
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
68	Franco_Keebler64	2016-11-13 20:09:27
71	Nia_Haag	2016-05-14 15:38:50
74	Hulda.Macejkovic	2017-01-25 17:17:28
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
80	Darby_Herzog	2016-05-06 00:14:21
81	Esther.Zulauf61	2017-01-14 17:02:34
83	Bartholome.Bernhard	2016-11-06 02:31:23
89	Jessyca_West	2016-09-14 23:47:05
90	Esmeralda.Mraz57	2017-03-03 11:52:27
91	Bethany20	2016-06-03 23:31:53
NULL	NULL	NULL

3) **Hashtag Research:** A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

```sql
SELECT tag_id, COUNT(*) AS tag_count FROM photo_tags GROUP BY tag_id ORDER BY tag_count DESC limit 5;

CREATE TABLE tags_count (
    tags varchar(50),
    tag_id int,
    tag_count int
);
insert into tags_count(tags, tag_id, tag_count)
values ('smile', 21, 59),('beach', 20, 42),('party', 17, 39),('fun', 13, 38),('concert', 18,24);
select *from tags_count limit 5;
```

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

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

Result Grid | Form Editor

Approach: At first we focused on tag ids and counted it. There were a total 21 ids out of which, id no, 21, 20, 17, 13, 18 were mostly used. Hence, we tabled tags are in its descending usage.

Now, from the table TAGS, we know the tag_id of each tag, hence we created a new table where the name of the tag, tag_id and count of each tag is used.

4) Ad Campaign Launch: The team wants to know the best day of the week to launch ads. Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

```
104 •   SELECT DAYOFWEEK(created_at) AS day_of_week, COUNT(*) AS registration_count
105     FROM users
106     GROUP BY day_of_week
107     ORDER BY registration_count DESC;
108
109
110
```

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

day_of_week	registration_count
5	16
1	16
6	15
3	14
2	14
4	13
7	12

Approach: We used the function DAYOFWEEK on the created date (column name: created_at) in the table USER. Here we see that on day 1st and 5th, maximum people have registered. Hence, we can schedule our campaign on any of these days.

INVESTOR MATRICES

1. User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
   Your 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.

Approach: we used the aggregate function AVG in table LIKES, to calculate the average post done by each number of users. Maximum post is done by user 46, and least post is done by user 61, (136 and 116) posts respectively.

```sql
SELECT user_id, avg(photo_id) as average_post_by_user
from likes
group by user_id
order by average_post_by_user desc;
```

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

user_id	average_post_by_user
46	139.9773
18	139.9146
38	139.4353
13	138.7957
39	138.2472
65	137.8542
97	136.7536
32	135.9341
72	135.8941
95	135.2791
96	135.1837
69	133.7010
93	133.6923
12	133.4805
60	132.4651
55	132.3974
48	132.2000
84	131.9467
31	131.8977
78	131.8193
63	131.6988
10	131.6207
79	130.9067
22	130.3187
20	129.5402
44	129.0233

2. Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.
Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

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

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

user_id
5
14
21
24
36
41
54
57
66
71
75

Approach: This query groups the likes by column user_id and then checks if the count of distinct photo_id for each user is equal to the total count of photos in the Photos table. Users who have liked every single photo will match this condition.

Insights: In the course of the project, my role as a data analyst was centered on extracting meaningful insights using SQL and MySQL Workbench. Key findings included identifying the five oldest users on Instagram, pinpointing users who never posted a photo, determining and communicating the contest winner's details, and analyzing user registration patterns to advise optimal ad campaign scheduling. Notably, the identification of potential bots enhances platform integrity. These insights equip the marketing team and investors with valuable information for strategic decision-making, fostering a deeper understanding of user dynamics and engagement on Instagram.

Result: The significance of SQL in this project cannot be overstated, as it served as the backbone for extracting meaningful insights and facilitating informed decision-making. SQL, with the assistance of MySQL Workbench, enabled me to efficiently query and manipulate the database to address specific business questions. The identification of the oldest users, determination of inactive users, and analysis of contest winners heavily relied on SQL's querying capabilities. Moreover, SQL's aggregations and filtering functionalities played a pivotal role in uncovering user registration patterns, guiding the optimal scheduling of ad campaigns. The language's versatility was particularly evident in spotting potential bots through anomaly detection in liking behavior. In essence, the proficiency in SQL not only streamlined data analysis but also empowered me to derive actionable insights, showcasing its indispensable role in data analytics and its direct impact on the success of the project.