

Project – 2

Instagram User Analytics

Project Description

The aim of this project is to analyze Instagram user engagement and take insights through which we will be able to take informed decisions about future directions of Instagram. By using MySQL workbench and SQL queries we will be able to handle tasks and answer key questions.

Approach

- First loaded the provided dataset into MySQL workbench.
- Then reviewed the table schema and the keys associated to it to understand the relationship between each table and columns.
- Further we analyzed the problem statement given and wrote SQL queries for each task and repeated the process of writing the SQL queries until we found an accurate answer.
- Verified accuracy of queries by validating results against expected patterns.

Tech-Stack Used

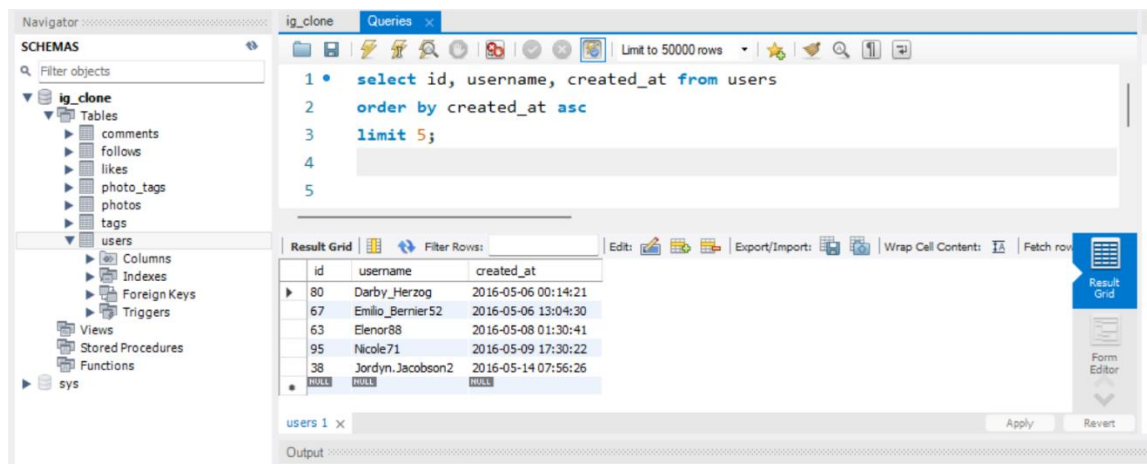
- Used MySQL workbench Version 8.0.41.0.
- Choose MySQL workbench as it can be used for database setup, querying, and performing commands for filtering, aggregating and joining data to have insights.

Insights

A) Marketing Analysis:

1. Loyal User Reward

We found 5 most loyal users by accessing users table and arranging them in ascending order and by printing the 5 users who have made account first from the given dataset.



The screenshot shows a database query interface with a sidebar on the left displaying a schema tree for 'ig_clone'. The main area shows a query editor with the following SQL code:

```
1 • select id, username, created_at from users
2   order by created_at asc
3   limit 5;
```

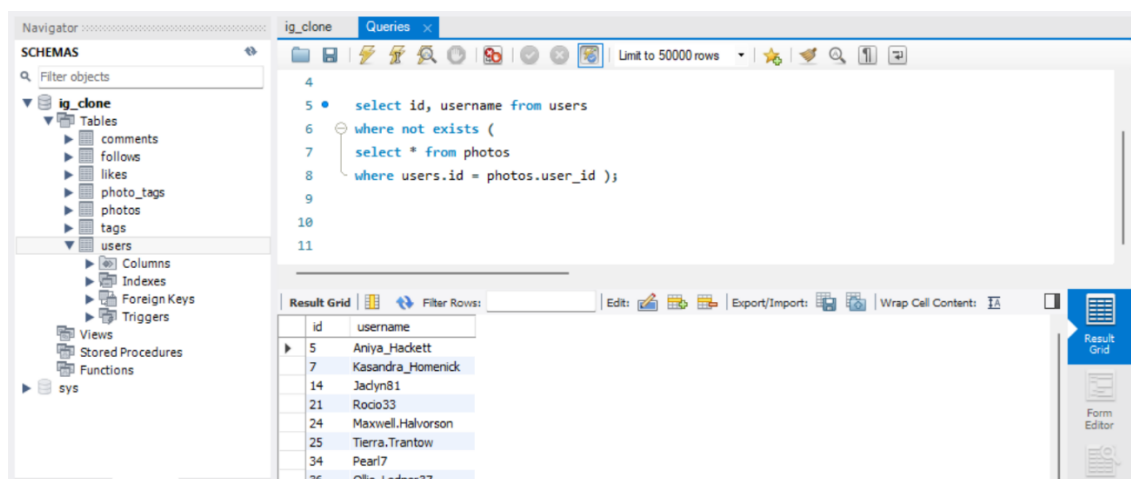
Below the query editor, the 'Result Grid' displays the results of the query:

id	username	created_at
80	Darby_Hierzog	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

Insight- we found most loyal users by finding who has been using the Instagram for the longest time and found 5 oldest users on Instagram which are given above in output table.

2. Inactive User Engagement

We found inactive user by finding who have not posted any photo. We found it from photos table where we searched which user id was not present.



The screenshot shows a database query interface with a sidebar on the left displaying a schema tree for 'ig_clone'. The main area shows a query editor with the following SQL code:

```
4
5 • select id, username from users
6   where not exists (
7     select * from photos
8     where users.id = photos.user_id );
9
10
11
```

Below the query editor, the 'Result Grid' displays the results of the query:

id	username
5	Aniya_Hackett
7	Kassandra_Homenick
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37

The image displays two screenshots of a data tool's 'Result Grid' interface. The top screenshot shows a list of users with IDs 5 through 66. The bottom screenshot shows a continuation of the list from ID 49 to 91, with a final row showing NULL values.

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

Insight - list of users who have never posted a single photo on Instagram this helps team to encourage inactive users to start engaging. The list is given above in the output table.

3. Contest Winner Declaration

We found contest winner by finding the most liked photo from likes table and joining it with photos table that gives as number of likes per photo and then we return the details of user with most liked photos from users table.

The screenshot shows a database query editor with a schema tree on the left and a query editor on the right. The schema tree shows a database named 'ig_clone' with tables: comments, follows, likes, photo_tags, photos, tags, and users. The query editor contains the following SQL query:

```

9
10 • select photo_id, photos.image_url, users.username, count(likes.user_id) as total_likes from photos
11 join likes on photos.id = likes.photo_id
12 join users on photos.user_id = users.id
13 group by photos.id, users.username
14 order by total_likes desc
15 limit 1;
16
17

```

The result grid shows the following data:

photo_id	image_url	username	total_likes
145	https://jarret.name	Zack_Kemmer93	48

Insight - found the person who has most likes on the post and identified the details about them. The person is Zack having username Zack_Kemmer93 and he has 48 likes.

4. Hashtag Research

We found five most commonly used photo by accessing photo_tags table where we counted count tags by photo_id from photo_tags table and then ordering it in descending order and finding the top 5 hashtags.

The screenshot shows a database query editor with a schema tree on the left and a query editor on the right. The schema tree shows a database named 'ig_clone' with tables: comments, follows, likes, photo_tags, photos, tags, and users. The query editor contains the following SQL query:

```

18
19 • select tag_name, count(photo_tags.photo_id) as usage_count from tags
20 join photo_tags on tags.id = photo_tags.tag_id
21 group by tags.id, tags.tag_name
22 order by usage_count desc
23 limit 5;
24
25
26

```

The result grid shows the following data:

tag_name	usage_count
smile	59
beach	42
party	39
fun	38
concert	24

Insight - identified five most used hashtags. This will help the users to maximize their reach on post by using most used hashtags. The hashtags are smile, beach, party, fun and concert.

5. Ad Campaign Launch

We found most users registration from users table where we counted the created_at which is day at which a user created account and then arrange it in descending order and got two most registered days.

Navigator

ig_clone

Queries

Limit to 50000 rows

```

23
24 • select dayname(created_at) as day_of_week, count(*) as registrations from users
25   group by dayname(created_at)
26   order by registrations desc
27   limit 5;
28
29

```

Result Grid

day_of_week	registrations
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14

Insight - found the day on which the maximum registration was received which will help the team to schedule ad campaign. The days are Thursday and Sunday.

B) Investor Metrics:

1. User Engagement

We found average number of posts per user by finding the count of total photos divided by count of total number of users.

SCHEMAS

ig_clone

Tables

comments

follows

likes

photo_tags

photos

tags

users

Columns

Indexes

Foreign Keys

Triggers

Views

Stored Procedures

Functions

sys

Limit to 50000 rows

```

29
30 • select count(*) as total_photos,
31   (select count(*) from users) as total_users,
32   round(count(*) / (select count(*) from users)) as avg_posts from photos;
33
34
35

```

Result Grid

total_photos	total_users	avg_posts
257	100	3

Insight –found average number of posts on Instagram which is 3.

2. Bots & Fake Accounts

We found dummy and fake accounts by finding out the user who has liked every single photo we did it by joining user and likes table and counted number of likes per user and compared it with total number of photos and the accounts which were equal are printed.

The screenshot shows a database interface with a 'Navigator' on the left and a 'Queries' window on the right. The 'Navigator' shows a schema named 'ig_clone' with tables like 'comments', 'follows', 'likes', 'photo_tags', 'photos', 'tags', and 'users'. The 'Queries' window shows a SQL query that joins the 'users' and 'likes' tables to count the total likes for each user. The 'Result Grid' shows the output of the query, listing 13 users with 257 likes each.

```

34 select users.id, users.username, count(likes.photo_id) as total_likes from users
35 join likes on users.id = likes.user_id
36 group by users.id, users.username
37 having count(likes.photo_id) = (select count(*) from photos);
38

```

id	username	total_likes
5	Aniya_Hackett	257
14	Jadyn81	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

Insight – we found user who have liked every post on the site this helps us to understand that the platform is crowded with 13 fake and dummy accounts.

Conclusion Insight

Found that most users who have not posted photos are dummy and fake users.

Result

Through this project we discovered user engagement patterns and behavior which helped us to understand the marketing and investor goals for future Instagram growth. It helped to gain deeper knowledge to analyze the situation and gain more knowledge in SQL querying and relational database.