Instagram User Analytics

Project by: Sumit Gope

Project Description

The project is basically about to analyse the user behaviour and activities on instagram and provide insights to the marketing, product & development teams for their business.

We have the user data to analyse and answer the different question asked by different teams. These insights then used by marketing teams to create new AD campaign. We will use SQL to perform analysis.

Approach

We will collect the database and analyse it and also analyse the schema. Then we will write and run SQL queries to extract data which are need to know. Basically the data which are required for different teams. We will use Operators, Aggregate Functions, Sorting Functions, Joins and also subqueries to perform complex queries.

Tech-Stack Used

I am using MySQL Workbench 8.0 CE software for write and run the SQL queries on the database.

Insights

I have derived following insights from the data Analysis.

1.Task: Find the 5 oldest users of the Instagram from the database provided.

Query:

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

OUTPUT:

username	created_at
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

Insights

2. Task: Find the users who have never posted a single photo on Instagram.

Query:

```
SELECT id, username
FROM users
WHERE id NOT IN (
SELECT DISTINCT user_id
FROM photos
 );
```

Output:

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
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhar
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

Insights

3. Task: Identify the winner of the contest and provide their details to the team. According to photo likes.

Query:

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

Output:

id	username	photo_id	image_url	likes_count
52	Zack_Kemmer93	145	https://jarret.name	48

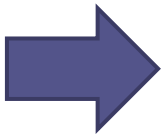
Insights

4. Task: Identify and suggest the top 5 most commonly used hashtags on the platform.

Query:

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

Output:



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

Insights

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

Query:

```
SELECT DAYNAME(created_at) AS day_of_week, COUNT(*) AS
 num_registrations
FROM users
GROUP BY DAYNAME(created_at)
ORDER BY num_registrations DESC
LIMIT 1;
```

Output:

day_of_week	num_registrations
Thursday	16

Insights

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

1st Query:

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

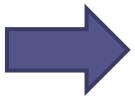
Output:

avg_posts_per_user
3.473

2nd Query:

```
SELECT COUNT(*) AS total_photos, COUNT(DISTINCT user_id) AS
total_users
FROM photos;
```

Output:



total_photos	total_users
257	74

Insights

7.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 user_id, COUNT(DISTINCT photo_id) AS photo_count
FROM likes
GROUP BY user_id
HAVING photo_count = (SELECT COUNT(DISTINCT id) FROM photos);
```

Output:

user_id	photo_count
5	257
14	257
21	257
24	257
36	257
41	257
54	257
57	257
66	257
71	257
75	257
76	257
91	257

While working on this project I get to know about the queries and its real time uses and also get to know about what kind of marketing and product teams asked question for there business purpose.

Through this project I extract the information that who are the oldest users of instagram, the users who never posted a single photo, the user who gets the most likes in his/her photo, most commonly hashtags used by users and the best day for release AD campaign.

Result

I have successfully analyzed the user behavior on Instagram and provided insights to the marketing and product teams. We have answered various questions asked by different teams using SQL queries. Our analysis has helped us understand user engagement, fake accounts, and other important metrics.