Project 2: Instagram User Analytics - SQL Fundamentals

Description:

The Aim of this project is to analyze Instagram user data and obtain business insights for marketing, product & development teams. The Objectives are,

A) The marketing team wants to launch some campaigns so they want to know the insights on

1. Rewarding the Most Loyal Users: People who have been using the platform for the longest time.
2. Reminding the Inactive Users to Start Posting: Users who have never posted a single photo on Instagram.
3. Declaring the Contest Winner: User who gets the most likes on a single photo.
4. which hashtags to use in the post to reach the most people on the platform: Top 5 most commonly used hashtags on the platform.
5. When to Launch an AD Campaign: What day of the week do most users register on.

B) The Investors want to know whether Instagram is performing well or not unlike Facebook, and so they want to know the insights on

1. The User Engagement: Users are still active and posting, or they are posting less. Also, to check the average post per user.
2. The Bots & Fake Accounts: Whether the platform is crowded with fake and dummy accounts, or not.

Approach:

Using MySQL, the database has been created which contains information on multiple tables about various fields. SQL queries are the used to extract the required data from the tables in order to acquire the corresponding insights.

Tech-Stack Used:

Software & Version used: MySQL Workbench 8.0 CE
Purpose: Used to store the data, for managing the databases and to support the SQL commands to analyze the results.

Insights:

A) The marketing team wants to launch some campaigns so they want to know the insights on

1. Rewarding the Most Loyal Users:

```
select * from users
order by timestamp(created_at)
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

Inference: These 5 are the oldest users of the Instagram from the database provided. Hence, they are the most loyal users.

2. Reminding the Inactive Users to Start Posting:

```
select users.id, users.username, count(photos.user_id) from users
left outer join photos on users.id = photos.user_id
where photos.image_url is null
group by users.id, users.username
order by users.id
```

id	username	count(photos.user_id)
5	Aniya_Hackett	0
7	Kasandra_Homenick	0
14	Jadyn81	0
21	Rocio33	0
24	Maxwell.Halvorson	0
25	Tierra.Trantow	0
34	Pearl7	0
36	Ollie_Ledner37	0
41	Mckenna17	0
45	David.Osinski47	0
49	Morgan.Kassulke	0
53	Linnea59	0
54	Duane60	0
57	Julien_Schmidt	0
66	Mike.Auer39	0
68	Franco_Keebler64	0
71	Nia_Haag	0
74	Hulda.Macejkovic	0
75	Leslie67	0
76	Janelle.Nikolaus81	0
80	Darby_Herzog	0
81	Esther.Zulauf61	0
83	Bartholome.Bernhard	0
89	Jessyca_West	0
90	Esmeralda.Mraz57	0
91	Bethany20	0

Inference: These are the users who have never posted a single photo on Instagram. Hence, promotional emails should be sent to them to post their 1st photo.

3. Declaring the Contest Winner:

```
select users.id, users.username, count(likes.photo_id) from users
left join photos on users.id = photos.user_id
left join likes on photos.id = likes.photo_id
group by users.username, likes.photo_id, users.id
order by count(likes.photo_id) desc
limit 1;
```

id	username	count(likes.photo_id)
52	Zack_Kemmer93	48

Inference: The contest winner with the most likes on a single photo is Zack_Kemmer93.

4. which hashtags to use in the post to reach the most people on the platform:

```
select tag_name, count(tags.id) from photos
left join photo_tags on photos.id = photo_tags.photo_id
left join tags on photo_tags.tag_id = tags.id
Group by tags.id
order by count(id) desc
limit 5;
```

tag_name	count(tags.id)
smile	59
beach	42
party	39
fun	38
concert	24

Inference: These are the top 5 most commonly used hashtags on the platform. Hence it is recommended that the partner brand include these hashtags in their post to reach the most people.

5. When to Launch an AD Campaign:

```
select DATE_FORMAT((created_at), '%W') as day_of_week, count(*) as most_registered
from users
group by day_of_week
order by most_registered desc
```

day_of_week	most_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Inference: Thursday and Sunday would be the best days to launch ADs.

B) The Investors want to know whether Instagram is performing well or not unlike Facebook, and so they want to know the insights on

1. The User Engagement:

```
with cte as
(select users.id, count(photos.id) as times_avg_user_posts from users
left join photos on users.id = photos.user_id
group by users.id
order by times_avg_user_posts desc)
select times_avg_user_posts, count(times_avg_user_posts) from cte
group by times_avg_user_posts
```

times_avg_user_posts	count(times_avg_user_posts)
12	1
11	1
10	1
9	1
8	2
6	1
5	14
4	13
3	9
2	13
1	18
0	26

Inference: From the extracted table, it is shown that 2 users has posted 8 times, 9 users has posted 3 times, 26 users has never posted and so on. so, the average user posts on Instagram are less than 12 times. Hence, it is concluded that the users are still as active and post on Instagram, except very few users.

```
select ((select count(*) from photos)/(select count(*) from users)) as 'avg posts'
```

avg posts
2.5700

Inference: The total number of photos on Instagram/The Total number of users = 2.5700

2. The Bots & Fake Accounts:

```
with cte as
(select users.username, count(likes.photo_id) as total_likes from likes
inner join users on users.id=likes.user_id
group by users.username)
select username, total_likes from cte
where total_likes=(select count(*) from photos)
order by username;
```

username	total_likes
Aniya_Hackett	257
Bethany20	257
Duane60	257
Jaclyn81	257
Janelle.Nikolaus81	257
Julien_Schmidt	257
Leslie67	257
Maxwell.Halvorson	257
Mckenna17	257
Mike.Auer39	257
Nia_Haag	257
Ollie_Ledner37	257
Rocio33	257

Inference: These are the users who have liked every single photo on the site. Hence, they are considered to be bots.

Results:

The Instagram user data is Analyzed and the insights are obtained, that can be used by the business team to launch the new marketing campaign. Based on the findings, further app features can be built or add in order to improve the platform. It is also suggested that measuring user engagement would help track the app success. Finally, contributing and improving the experience altogether can assist the firm expand.