PROJECT 2

INSTAGRAM USER ANALYTICS

SQL Fundamentals

Project description:

As a data analyst on Instagram product team, this project focuses on analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

Using SQL and MySQL workbench as a tool, the goal is to track how users engage with a digital product, to provide insights derived from this analysis that can help various teams within the business to take informed decisions.

Approach:

- The questions were carefully reviewed to understand the requirements of the provide questions.
- The tables and columns that will be used to answer each question were identified.
- Required information was extracted using MySQL queries using MySQL software.
- The insights derived from the analysis were carefully reviewed to to further take informed decisions.

Tech-Stack Used: I am using MySQL Workbench 8.0.38-winx64 CE and it is ideal for this project as it provides a user-friendly interface for running SQL queries,

managing databases, visualising trends, making it easier to analyze and derive insights efficiently.

SQL Tasks:

A) Marketing Analysis:

1.Loyal User Reward: Identify the five oldest users on Instagram from the provided database.

Run SQL query as:

```
USE ig_clone;

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

Result:

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

The table shows the five oldest users on Instagram from provided database.

2. Inactive User Engagement: Identify users who have never posted a single photo on Instagram.

Query:

```
USE ig_clone;

select username, t1.id as `user id`
from users t1
left join photos t2
on t1.id = t2.user_id
where t2.user_id is null
```

Result:

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

The table shows 26 users that have never posted a single photo on Instagram. Promotional emails can be sent to these inactive users.

3. Contest Winner Declaration: Determine the winner of the contest and provide their details to the team.

Query:

```
USE ig_clone;

select t3.id, t3.username, photo_id, t2.image_url , count(*) as number_of_likes
from likes t1
inner join photos t2 on t1.photo_id = t2.id
inner join users t3 on t2.user_id = t3.id
group by photo_id
order by number_of_likes DESC
limit 1;
```

Result:

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

The table shows that Zack_kemmer93 has most of the likes in single photo and thus is winner of the contest.

4. Hashtag Research: Identify and suggest the top five most used hashtags on the platform.

Query:

```
use ig_clone;

select tag_name, count(*) as most_used_hashtag
from photo_tags t1
inner join tags t2
on t1.tag_id = t2.id
group by tag_name
order by most_used_hashtag DESC
limit 5;
```

Result:

	tag_name	most_used_hashtag
▶	smile	59
	beach	42
	party	39
	fun	38
	concert	24

The table shows top five most used hashtags on the platform.

5. Ad Campaign Launch: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

Query:

```
use ig_clone;

select  dayname(created_at) as weeks, count(*) as most_users_registered
from users
group by weeks
order by most_users_registered DESC;
```

Result:

	weeks	most_users_registered
▶	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

The table shows most users registered on Thursdays and Sundays on Instagram, so these days are best to schedule an ad campaign.

B) Investor Metrics:

1.User Engagement: 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.

a. Query for average number of posts per user

```
use ig_clone;

select avg(number_of_post) as avg_post_per_user
from
(select user_id, count(*) as number_of_post
from photos
group by user_id ) as average_post;
```

Result:

	avg_post_per_user
▶	3.4730

The table shows that average user posts more than 3 photos.

b. Query for total number of photos on Instagram divided by the total number of users.

```
use ig_cone;

select count(id) / (select count(id) as total_users
from users) as `total photos to total users`
from photos ;
```

Result:

	total photos to total users
▶	2.5700

2. Bots & Fake Accounts: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Query:

```
use ig_clone;

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

Result:

	username	user_id	total_likes
▶	Aniya_Hackett	5	257
	Jaclyn81	14	257
	Rocio33	21	257
	Maxwell.Halvorson	24	257
	Ollie_Ledner37	36	257
	Mckenna17	41	257
	Duane60	54	257
	Julien_Schmidt	57	257
	Mike.Auer39	66	257
	Nia_Haag	71	257
	Leslie67	75	257
	Janelle.Nikolaus81	76	257
	Bethany20	91	257

The table shows the users that have liked every single photo on Instagram that Is typically not normal for very user so these users can be potential bots.

Insights:

The various insights were derived from the analysis of data:

- The most loyal users, i.e., those who have been using the platform for the longest time have been using the platform since 2016
- Some users are inactive on the platform so promotional emails can be scheduled for those users to encourage them to start posting.
- The most used hashtags on the platform are smile, beach, party, fun and concert.
- Thursdays and Sundays are the days on which most users registered on Instagram so these days can be scheduled to launch an ad campaign.
- Average user posts more than 3 photos on Instagram.
- There are potential bots of fake accounts on Instagram as they liked every post on platform which is not normal for a normal user.

All these insights can be used by various departments to take informed decisions that can help business grow.

Result:

Through this project I was able to manage a structured database for the project. The project helped in gaining understanding of managing tables, applying joins, and performing complex queries to extract relevant information through hands-on experience in performing data analysis using SQL queries, filtering data that provided valuable insights.

Also, applied MySQL in a real-world case, translating theoretical knowledge into practical use which is essential for any data-driven role.