

Instagram User Analytics Project

Project Description:

This project focuses on analyzing Instagram platform's user data to derive actionable insights and address specific business queries. The aim is to identify loyal users, inactive accounts, engagement trends, and popular content, which are critical to enhancing user retention, content strategies, and operational efficiency. The approach will involve using SQL queries to extract meaningful patterns from the data stored in relational tables.

Approach:

The analysis was carried out in the following steps:

Understand the Database: Explored all relational tables about users, photos, likes, tags, to know how data is linked together.

Define Queries: Set SQL queries according to required tasks, such as extracting loyal users, inactive account users, top hashtags of photos, and engagement over time.

Data Extraction/Analysis: Ran the defined queries to fetch the required information and analyze the patterns extracted and get insights.

Validation: Cross-verified the results to obtain an accurate and reliable interpretation.

Documentation: Documented the process and insights gained for clear presentation and sharing with stakeholders.

Tech-Stack Used:

MySQL Workbench: Used for database management, executing queries, and visualizing results due to its user-friendly interface and powerful SQL capabilities.

MySQL Database: Chosen for its reliability and ability to handle relational data effectively.

Insights:

1. The five oldest users were identified providing the team with a list of the loyal users to reward potential.
2. A number of users who are inactive were detected so that areas where user engagements can be improved.
3. There was a winner for this contest, and this provided information on the photo received the most likes and represents the type of content for which users have a preference for.
4. The analysis of hashtags brought up the most popular five ones, which, therefore, are used in guiding marketing efforts for increasing reach.
5. User registration peaks at certain days. This is a clue for scheduling promotional campaigns.
6. The project brings out the possibilities of bot accounts, with unusual engagement patterns, thereby assuring better data quality and user experience.

QUERIES:

1. Identify the five oldest users on Instagram from the provided database.

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

Output:

	username	created_at
▶	Darby_Herzog	2016-05-06 00:14:21
	Darby_Herzog	2016-05-06 00:14:21
	Darby_Herzog	2016-05-06 00:14:21
	Emilio_Bernier52	2016-05-06 13:04:30
	Emilio_Bernier52	2016-05-06 13:04:30

2. Identify users who have never posted a single photo on Instagram.

```
SELECT u.username
FROM users u
LEFT JOIN photos p ON u.id = p.user_id
WHERE p.id IS NULL;
```

Output:

username	username
Ressie_Stanton46	Alysa22
Elenor88	Milford_Gleichner42
Florence99	Delfina_VonRueden...
Adelle96	Rick29
Mike.Auer39	Clint27
Emilio_Bernier52	Jessyca_West
Franco_Keebler64	Esmeralda.Mraz57
Karley_Bosco	Bethany20
Erick5	Frederik_Rice
Nia_Haag	Willie_Leuschke
Kathryn80	Damon35
Jaylan.Lakin	Nicole71
Hulda.Macejkovic	Keenan.Schamberg...
Leslie67	Tomas.Beatty93
Janelle.Nikolaus81	Imani_Nicolas17
Donald.Fritsch	Alek_Watsica
Colten.Harris76	Javonte83
Katarina.Dibbert	Kenton_Kirlin
Darby_Herzog	Andre_Purdy85
Esther.Zulauf61	Harley_Lind18
Aracely.Johnston98	Arely_Bogan63
Bartholome.Bernhard	Aniya_Hackett
Alysa22	Travon Waters

3. Determine the winner of the contest and provide their details to the team.

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

Output:

Result Grid	Filter Rows:	Export:	Wr
username	image_url	likes_count	
Zack_Kemmer93	https://jarret.name	48	

4. Identify and suggest the top five most commonly used hashtags on the platform.

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

Output:

Result Grid	Filter Rows:
tag_name	usage_count
smile	59
beach	42
party	39
fun	38
concert	24

5. Determine the day of the week when most users register on Instagram.

```
SELECT DAYNAME(created_at) AS registration_day, COUNT(id) AS user_count
FROM users
GROUP BY registration_day
ORDER BY user_count DESC
LIMIT 1;
```

Output:

Result Grid	Filter Rows:
registration_day	user_count
Thursday	48

6. 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.

```
SELECT AVG(photo_count) AS avg_posts_per_user
FROM (
    SELECT COUNT(p.id) AS photo_count
    FROM users u
    LEFT JOIN photos p ON u.id = p.user_id
    GROUP BY u.id
) AS user_photo_counts;

-- Total photos divided by total users
SELECT COUNT(*) / (SELECT COUNT(*) FROM users) AS photos_per_user
FROM photos;
```

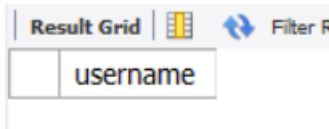
Output:

Result Grid		Filter Rows:
	photos_per_user	
▶	2.5700	

7. Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

```
• SELECT u.username
  FROM users u
 WHERE NOT EXISTS (
     SELECT 1
     FROM photos p
     WHERE NOT EXISTS (
         SELECT 1
         FROM likes l
         WHERE l.user_id = u.id AND l.photo_id = p.id
     )
 )
 );
```

Output:



The screenshot shows a database interface with a 'Result Grid' tab. Below the tab, there is a single column header labeled 'username'. The grid is currently empty, showing only the header row.

username

Result:

The project provided detailed analytics on user engagement and preferences in terms of content for the platform, as well as usage patterns. With this information, it is possible to further improve user retention strategies, marketing campaigns, and the overall effectiveness of the platform. This hands-on experience deepened my understanding of database design, SQL querying, and data-driven decision-making.