

Instagram User Analytics

Project Description

This project focused on analyzing user interactions and engagement data from Instagram to provide actionable insights that can help drive business decisions. The goal was to analyze key aspects of user behavior, such as engagement patterns, inactive users, contest winners, and hashtag usage, using **SQL**. These insights can be leveraged by marketing, product, and development teams to improve platform strategies, user engagement, and overall business growth.

The project included several tasks aimed at answering critical business questions using **SQL queries** to extract meaningful data from the provided Instagram user database.

Approach

The approach to executing this project was structured and methodical:

1. **Database Setup:**

The first step involved setting up the database in **Microsoft SQL Server**. This included running the necessary commands to create tables, establish relationships, and ensure the integrity of the data.

```
CREATE DATABASE ig_clone;
USE ig_clone;
```

2. **Marketing Analysis:**

Each marketing-related task was addressed by writing SQL queries tailored to the specific requirements. Tasks included identifying loyal users, inactive users, the winner of a photo contest, top trending hashtags, and the best day of the week to launch ads. SQL queries like **ORDER BY**, **GROUP BY**, and **COUNT** were used to extract the required data.

3. **Investor Metrics:**

For investor-related tasks, SQL was used to calculate average posts per user, identify potential bots, and assess user engagement. Queries involving aggregate functions like **AVG** and **COUNT** were written to gather insights into platform activity and account authenticity.

4. **Query Execution:**

After writing each query, I ensured that they were efficient and returned the correct results. The outputs were verified, and snapshots of the SQL queries and their corresponding results were captured for the report.

Tech-Stack Used

- **Microsoft SQL Server:**

For this project, **Microsoft SQL Server** was used as the database management system. This tool was chosen because of its robust functionality for handling large datasets, performing complex queries, and supporting a wide variety of analytical tasks. It provided an efficient platform for managing the Instagram user data and performing analysis.

SQL Tasks :

A) Marketing Analysis:

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

```
select top 5 * from users  
order by created_at
```

	id	username	created_at
1	80	Darby_Herzog	2016-05-06 00:14:21.190
2	67	Emilio_Bernier52	2016-05-06 13:04:29.960
3	63	Elenor88	2016-05-08 01:30:40.677
4	95	Nicole71	2016-05-09 17:30:22.370
5	38	Jordyn.Jacobson2	2016-05-14 07:56:25.837

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

```
select * from users, photos  
select u.id, username from users u left join photos p on u.id=p.user_id where image_url is null order by id
```

	id	username
1	5	Aniya_Hackett
2	7	Kassandra_Homenick
3	14	Jaclyn81
4	21	Rocio33
5	24	Maxwell.Halvorson
6	25	Tierra.Trantow
7	34	Pearl7
8	36	Ollie_Ledner37
9	41	Mckenna17
10	45	David.Osinski47
11	49	Morgan.Kassulke
12	53	Linnea59
13	54	Duane60
14	57	Julien_Schmidt
15	66	Mike.Auer39
16	68	Franco_Keebler64
17	71	Nia_Haag
18	74	Hulda.Macejkovic
19	75	Leslie67
20	76	Janelle.Nikolaus81
21	80	Darby_Herzog
22	81	Esther.Zulauf61
23	83	Bartholome.Bernha...
24	89	Jessyca_West
25	90	Esmeralda.Mraz57
26	91	Bethany20

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

```
select top 1 u.username, l.photo_id, count(l.user_id)[No. of Likes] from likes l join photos p on l.photo_id=p.id
join users u on p.user_id=u.id
group by u.username, l.photo_id
order by count(l.user_id) desc
```

	username	photo_id	No. of Likes
1	Zack_Kemmer93	145	48

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

```
select top 5 t.tag_name, count(p.photo_id)[Times] from tags t join photo_tags p on t.id=p.tag_id
group by t.tag_name
order by count(p.photo_id) desc
```

	tag_name	Times
1	smile	59
2	beach	42
3	party	39
4	fun	38
5	lol	24

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

```
] select DATENAME(weekday,created_at)[DayName], count(username)[Times] from users
group by DATENAME(weekday,created_at)
order by count(username) desc
```

	DayName	Times
1	Sunday	16
2	Thursday	16
3	Friday	15
4	Monday	14
5	Tuesday	14
6	Wednesday	13
7	Saturday	12

B) Investor Metrics:

1. 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.

```
with cte as (  
  select u.id as userID, count(p.id) as PhotoID from users u left join photos p on u.id=p.user_id  
  group by u.id  
)  
  
select  
  SUM(photoID) [Total Photos],  
  COUNT(userID) [Total Users],  
  CAST(SUM(PhotoID) AS DECIMAL) / NULLIF(COUNT(userID), 0) AS photoperuser  
from cte
```

	Total Photos	Total Users	photoperuser
1	257	100	2.570000000000

2. 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.id AS userID, u.username, count(l.photo_id)[Likes] from users u join likes l on u.id = l.user_id  
group by u.id, u.username  
having COUNT(distinct l.photo_id) = (select COUNT(*) from photos)  
order by u.username
```

	userID	username	Likes
1	5	Aniya_Hackett	257
2	91	Bethany20	257
3	54	Duane60	257
4	14	Jaclyn81	257
5	76	Janelle.Nikolaus81	257
6	57	Julien_Schmidt	257
7	75	Leslie67	257
8	24	Maxwell.Halvorson	257
9	41	Mckenna17	257
10	66	Mike.Auer39	257
11	71	Nia_Haag	257
12	36	Ollie_Ledner37	257
13	21	Rocio33	257

Insights

The analysis of Instagram user data yielded several key insights:

1. **Loyal Users:**

The analysis of the oldest users helped identify a segment of loyal users who have been active on the platform for the longest period. These users can be targeted with loyalty programs or special campaigns.

2. **Inactive Users:**

By identifying users who have never posted a photo, the marketing team can focus on re-engaging these users through email campaigns or app notifications, encouraging them to become active participants.

3. **Contest Winner:**

The winner of the contest, based on the highest number of likes on a single photo, was identified as **zack_kemmer93**. This user could be a valuable asset for the marketing team to promote further engagement or collaborate in influencer campaigns, leveraging their popularity to drive more user interaction and visibility on the platform.

4. **Top Hashtags:**

The most commonly used hashtags were identified as **#smile (59 times)**, **#beach (42 times)**, **#party (39 times)**, **#fun (38 times)**, and **#lol (24 times)**. These insights can guide brands in selecting these popular hashtags for their posts to maximize visibility, engagement, and reach among the Instagram audience.

5. **Best Day for Ad Campaign:**

From the data above, the days with the highest number of registrations are **Sunday** and **Thursday**, both with 16 registrations. These two days are the **best days to launch ads**, as they reflect a higher level of user activity on Instagram.

This information is crucial for the marketing team to schedule ads for maximum impact.

6. **User Engagement:**

Calculating the average number of posts per user provided insight into the platform's activity level. This data can help assess whether user engagement has increased or decreased over time.

7. **Bots and Fake Accounts:**

By identifying users who liked every single photo on the platform, the project helped pinpoint potential fake accounts or bots. This insight can help in maintaining platform integrity and ensuring that engagement metrics are reliable.

Result

The project successfully answered key business questions and provided valuable insights into user behavior, engagement patterns, and platform activity. By leveraging **SQL** to analyze the data, the insights can be used by marketing, product, and development teams to refine their strategies, engage with users more effectively, and ensure the growth of the platform.

Key achievements of the project include:

- Identification of loyal and inactive users for targeted campaigns.
- Recognition of contest winners and popular hashtags to improve brand visibility.
- Insights into user engagement trends and potential bot accounts.

Overall, the project enhanced my skills in **SQL** and database analysis, and I gained a deeper understanding of how data analytics can influence strategic decisions for a large-scale platform like Instagram.