Instagram User Analytics with SQL

Introduction

The Instagram User Analytics SQL Project is aimed at analyzing user behavior and engagement patterns on Instagram using SQL. This project is designed to provide actionable insights to the product, marketing, and investor teams by querying a dataset that simulates Instagram user interactions such as posts, likes, comments, and account creation. By leveraging SQL, the goal is to extract meaningful data that can influence business strategies, optimize marketing efforts, and improve overall user experience on the platform.

With the rise of social media, understanding user activity has become paramount to sustaining growth. Whether it's identifying loyal users, optimizing ad placements, or detecting bots, this project applies data analytics to inform Instagram's strategic decisions.

Objectives

The main objectives of the Instagram User Analytics SQL Project are:

1. Analyze User Interactions: Understand how users are engaging with the platform by analyzing posts, likes, and comments.
2. Identify Marketing Opportunities: Provide insights to the marketing team on how to effectively engage users and optimize campaigns.
3. Track Key Investor Metrics: Supply investors with essential metrics about user engagement and activity on the platform.
4. Detect Fake Accounts: Identify potential bot or fake accounts to ensure data integrity and platform authenticity.
5. Inform Product Development: Help the product team by highlighting user trends and activity patterns that can influence future feature developments.

Business-Related Questions

The project seeks to answer several key business-related questions through SQL analysis, which will help different teams make data-driven decisions:

1. Loyalty and User Retention:
    - Who are the most loyal users who have been using the platform for the longest time?
2. Inactive User Engagement:
    - Which users have never posted on Instagram, and how can we re-engage them to increase platform activity?
3. Content Engagement:
    - Who is the winner of a contest based on the number of likes received on a single post?
4. Hashtag Optimization:
    - What are the most popular hashtags used on the platform that brands can leverage to improve their reach?
5. Ad Campaign Timing:
    - What is the best day of the week to launch an ad campaign based on user registration trends?
6. User Activity:
    - What is the average number of posts per user, and how does this reflect overall platform engagement?
7. Bot and Fake Account Detection:
    - Which users exhibit bot-like behaviour by liking every post on the platform?

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

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

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

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

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

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson

3. Determine the winner of a contest by finding the user with the most likes on a single photo.

```sql
SELECT p.user_id, u.username, p.id AS photo_id, COUNT(l.user_id) AS total_likes
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, p.user_id, u.username
ORDER BY total_likes DESC
LIMIT 1;
```

user_id	username	photo_id	total_likes
52	Zack_Kemmer93	145	48

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

```sql
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.tag_name
ORDER BY usage_count DESC
LIMIT 5;
```

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

5. Determine the best day of the week to launch ads based on user registration trends.

```sql
SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS total_registrations
FROM users
GROUP BY registration_day
ORDER BY total_registrations DESC
LIMIT 1;
```

registration_day	total_registrations
Thursday	32

6. Calculate the average number of posts per user and the total number of photos divided by the total number of users.

```sql
SELECT AVG(post_count) AS avg_posts_per_user
FROM (
  SELECT user_id, COUNT(*) AS post_count
  FROM photos
  GROUP BY user_id
) AS user_post_counts;
```

avg_posts_per_user
6.9459

```sql
SELECT COUNT(*) / (SELECT COUNT(*) FROM users) AS total_photos_per_user
FROM photos;
```

total_photos_per_user
2.5700

7. Identify users who have liked every single photo on the platform, which may indicate bot or fake activity.

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

id	username
5	Aniya_Hackett
14	Jadyn81
21	Rocio33
24	Maxwell.Halvorson
36	Ollie_Ledner37

Key Takeaways

The key takeaways from the project's analysis are:

- Loyal Users Matter: Identifying the most loyal users can help Instagram create tailored reward programs, improving user retention and fostering brand loyalty.
- Re-engagement Potential: Many users have never posted on the platform. Targeting these inactive users with personalized marketing campaigns can boost engagement and drive more content creation.
- Influencer Insights: The contest analysis reveals the most influential users, whose content resonates well with the audience. These users could be prime candidates for influencer marketing campaigns or partnerships.
- Hashtags Drive Reach: Popular hashtags are a critical tool for brands to optimize their posts and reach more users. By using the top five most frequently used hashtags, brands can improve their marketing ROI.
- Ad Campaign Optimization: Knowing the most common day for user registrations allows marketers to time their ads for maximum exposure, leading to better performance and conversion rates.
- Engagement Metrics for Investors: The average posts per user and the total photos per user metrics give investors a clear understanding of platform engagement, demonstrating that users are actively contributing content to the platform.
- Fake Account Detection: Identifying users who behave like bots (e.g., liking every post) ensures that Instagram maintains a trusted and genuine user base, which is crucial for advertisers and investors alike.

Call to Actions

Based on the findings from this project, Instagram can take several actions to improve user engagement, optimize marketing efforts, and ensure investor confidence:

- Reward Long-Term Users: Implement loyalty programs for the oldest and most active users to increase retention and incentivize long-term engagement.
- Target Inactive Users: Create re-engagement campaigns that encourage users who have never posted to start participating. Offering incentives such as exclusive features or content may increase their activity.
- Leverage Influencers: Partner with the most engaged users and contest winners to promote user-generated content and run marketing campaigns. Highlight their success stories to inspire more engagement from the community.
- Optimize Hashtag Use: Provide guidance to brands on using the top-performing hashtags to improve their campaign reach and engagement. Instagram can also promote hashtag challenges around these popular tags to boost platform interaction.
- Improve Ad Timing: Use insights about user registration trends to time ad campaigns effectively, maximizing ad visibility during peak user activity days.
- Enhance Platform Integrity: Continue monitoring and removing fake or bot accounts to maintain a high level of trust among users and advertisers. This is essential for ensuring the accuracy of engagement data, which influences investor confidence and advertising revenue.
- Future Product Development: Use engagement trends and hashtag popularity to inform future feature development, ensuring that Instagram continues to evolve with user preferences and behavior.

Thank you!