Trainity Assignment Two Instagram User Analytics

Aayush Damani

Table of Contents

1. PROJECT OVERVIEW
2. APPROACH
3. TECH STACK USED
4. RESULTS AND INSIGHTS
5. GOOGLE DRIVE LINK

Project Description

- The goal of this project is to leverage SQL skills to analyze Instagram user data and provide valuable insights that can help the business grow.

- The analysis will focus on various aspects such as user engagement, potential bots, popular hashtags, and optimal times for marketing campaigns.

- The insights derived from this analysis will assist different teams within the organization, such as marketing, product, and development, in making informed decisions.

Project Overview

Marketing Analytics

Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

Investor Metrics

User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

Approach

Data Preparation:

- Set up the database using provided commands and files.
- Ensure data integrity for accurate analysis.

Analysis Execution:

- **Oldest Users:** Identify the five longest-registered users.
- **Inactive Users:** Find users who have never posted a photo.
- **Contest Winner:** Determine the user with the most likes on a single photo.
- **Popular Hashtags:** Identify the top five most used hashtags.
- **Ad Campaign Timing:** Determine the best day for launching ads based on user registration patterns.
- **User Engagement:** Calculate the average number of posts per user and the total number of photos per user ratio.
- **Bots Detection:** Identify users who have liked every single photo on the platform.

Query Execution:

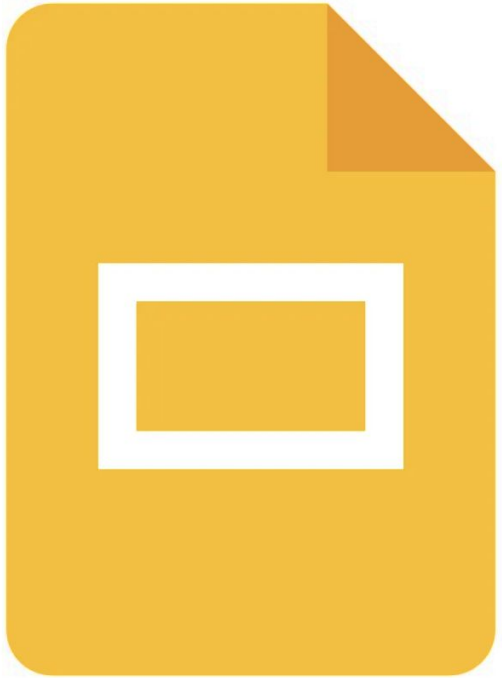
- Write and validate SQL queries in MySQL Workbench.

Reporting:

- Create a PowerPoint report with SQL queries, outputs, and insights.
- Include screenshots and summarize findings for leadership presentation.

Tech Stack Used

Google Slides



Loyal User Reward

Identifying the 5 oldest users on Instagram by identifying the date created.

SQL Code

```sql
## Loyal User Rewards

SELECT *
FROM users
ORDER BY created_at ASC
LIMIT 5;
```

Results

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
NULL	NULL	NULL

RESULTS

From the results provided in the previous slide, it is clear to see that the 5 oldest users on Instagram created their accounts in 2016.

INSIGHTS

These 5 oldest users should be rewarded with some type of loyalty prize, such as an ad-free experience for a limited period of time.

Inactive User Engagement

Identifying users that have never liked a photo.

SQL Code

```sql
## Inactive User Engagement: Identifying users that have never posted on Instagram.

SELECT * FROM users
LEFT JOIN photos
ON users.id=photos.user_id
WHERE photos.user_id IS NULL;

## Checking to see whether these user_id's exist in the photos table.

SELECT COUNT(*)
FROM users
WHERE user_id NOT IN (SELECT DISTINCT user_id FROM photos);
```

Results

id	username	created_at	id	image_url	user_id	created_dat
5	Aniya_Hackett	2016-12-07 01:04:39	NULL	NULL	NULL	NULL
7	Kasandra_Homenick	2016-12-12 06:50:08	NULL	NULL	NULL	NULL
14	Jaclyn81	2017-02-06 23:29:16	NULL	NULL	NULL	NULL
21	Rocio33	2017-01-23 11:51:15	NULL	NULL	NULL	NULL
24	Maxwell.Halvorson	2017-04-18 02:32:44	NULL	NULL	NULL	NULL
25	Tierra.Trantow	2016-10-03 12:49:21	NULL	NULL	NULL	NULL
34	Pearl7	2016-07-08 21:42:01	NULL	NULL	NULL	NULL
36	Ollie_Ledner37	2016-08-04 15:42:20	NULL	NULL	NULL	NULL
41	Mckenna17	2016-07-17 17:25:45	NULL	NULL	NULL	NULL
45	David.Osinski47	2017-02-05 21:23:37	NULL	NULL	NULL	NULL
49	Morgan.Kassulke	2016-10-30 12:42:31	NULL	NULL	NULL	NULL
53	Linnea59	2017-02-07 07:49:34	NULL	NULL	NULL	NULL
54	Duane60	2016-12-21 04:43:38	NULL	NULL	NULL	NULL
57	Julien_Schmidt	2017-02-02 23:12:48	NULL	NULL	NULL	NULL
66	Mike.Auer39	2016-07-01 17:36:15	NULL	NULL	NULL	NULL
68	Franco_Keebler64	2016-11-13 20:09:27	NULL	NULL	NULL	NULL
71	Nia_Haag	2016-05-14 15:38:50	NULL	NULL	NULL	NULL
74	Hulda.Macejkovic	2017-01-25 17:17:28	NULL	NULL	NULL	NULL
75	Leslie67	2016-09-21 05:14:01	NULL	NULL	NULL	NULL
76	Janelle.Nikolaus81	2016-07-21 09:26:09	NULL	NULL	NULL	NULL
80	Darby_Herzog	2016-05-06 00:14:21	NULL	NULL	NULL	NULL
81	Esther.Zulauf61	2017-01-14 17:02:34	NULL	NULL	NULL	NULL
83	Bartholome.Bernhard	2016-11-06 02:31:23	NULL	NULL	NULL	NULL
89	Jessyca_West	2016-09-14 23:47:05	NULL	NULL	NULL	NULL
90	Esmeralda.Mraz57	2017-03-03 11:52:27	NULL	NULL	NULL	NULL
91	Bethany20	2016-06-03 23:31:53	NULL	NULL	NULL	NULL

RESULTS

From the results provided in the previous slide, it is clear to see that there are 26 users that have never posted a picture on Instagram before.

INSIGHTS

These users can be targeted with promotional emails in order to encourage them to post more.

These users should be investigated further in order to determine whether they are active in comments, likes, messaging etc.

Contest Winner Declaration

Identifying the photo and user with the most likes.

SQL Code

```sql
## Contest Winner Declaration: Identifying the user with the most likes on a single photo

SELECT photo_id, COUNT(*) FROM likes
GROUP BY photo_id
ORDER BY COUNT(*) DESC
LIMIT 1;

SELECT * FROM photos
INNER JOIN users
ON photos.user_id=users.id
WHERE photos.id=145;
```

Results

id	image_url	user_id	created_dat	id	username	created_at
145	https://jarret.name	52	2024-07-30 16:35:02	52	Zack_Kemmer93	2017-01-01 05:58:22

RESULTS

From the results provided in the previous slide, it is clear to see that Zach_Kemmer93 (userid - 52) received the most likes for his photo (photo id - 145).

INSIGHTS

This user should be rewarded with some prize for winning the contest.

Hashtag Research

Identifying the most popular hashtags.

SQL Code

```sql
## Hashtag Research: Identifying Top 5 tags used

SELECT tag_name, COUNT(id) FROM tags
GROUP BY tag_name;
```

Results

tag_name	COUNT(id)
beach	1
beauty	1
concert	1
delicious	1
dreamy	1

RESULTS

From the results provided in the previous slide, it is clear to see that the number of times a hashtag was used most was 1 in the given data. No hashtag was used more than once.

INSIGHTS

There is not enough data to draw any insights from the results. It would be ideal to collect more data about hashtag usage.

Ad Campaign Launch

Identifying the day of the week where most new users register

SQL Code

```sql
## Ad Campaign Launch: Identifying the day of the week most users register on Instagram

SELECT COUNT(id), DAYNAME(created_at) AS day_name
FROM users
GROUP BY day_name;
```

Results

COUNT(id)	day_name
16	Thursday
16	Sunday
14	Tuesday
12	Saturday
13	Wednesday
14	Monday
15	Friday

RESULTS

From the results provided in the previous slide, it is clear to see that Thursday and Sunday are the days where most new accounts are created.

INSIGHTS

This suggests that the ad campaigns should be focused on Thursdays and Sundays in order to capitalize on the new user joining rate on these days.

User Engagement

Calculating the Average Number of Photos per User, and the Total Photos/Total Users

SQL Code

```sql
## User Engagement: Average Photos per User, and total number of photos/number of users on Instagram.

SELECT AVG(photo_count) AS average_photos_per_user
FROM (
    SELECT COUNT(id) AS photo_count
    FROM photos
    GROUP BY user_id
) AS subquery;
```

Results

average_photos_per_user
3.4730

RESULTS

From the results provided in the previous slide, it is clear to see that the average photos per user is 3.47 (this calculation excludes users that have never posted before).

The total number of photos on Instagram is 257, and the total users is 100. Therefore the total number of photos/total number

INSIGHTS

These numbers could be used as key performance indicators to assess marketing campaigns' efficiency i.e. the marketing campaign would be deemed successful if the average photos per user goes up to 4.5.

Bots & Fake Accounts

Finding the users that have liked every single photo.

SQL Code

```sql
## Bots: Identify users that have liked every single photo

SELECT users.username, user_id, COUNT(user_id) FROM likes
INNER JOIN users
ON likes.user_id=users.id
GROUP BY user_id
ORDER BY COUNT(user_id) DESC
LIMIT 20;
```

Results

username	user_id	COUNT(user_id)
Janelle.Nikolaus81	76	257
Aniya_Hackett	5	257
Mckenna17	41	257
Mike.Auer39	66	257
Duane60	54	257
Jaclyn81	14	257
Nia_Haag	71	257
Ollie_Ledner37	36	257
Leslie67	75	257
Julien_Schmidt	57	257
Rocio33	21	257
Bethany20	91	257
Maxwell.Halvorson	24	257
Annalise.McKenz…	16	103
Keenan.Schamb…	96	98

RESULTS

From the results provided in the previous slide, it is clear to see that there are multiple users that have liked every single photo (the ones with 257 likes on distinct photos).

INSIGHTS

These accounts could be investigated further to confirm whether they are a bot or not, by looking at the date the account was created, or exploring the other usage of the Instagram app such as comments, tags etc.

DRIVE LINK

https://docs.google.com/presentation/d/1_K-e0AE1_zjRJmPH5OLUB5ho_3RBkkyxYPX7QVSaj3k/edit?usp=sharing

THANK YOU FOR YOUR TIME.