Trainity

Sql Project

Instagram User Analytics

2023

Prepared by :

Pawan Rana

+918445578888

LinkedIn

work.pawanrana@gmail.com

TABLE CONTENT

OBJECTIVE

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 your tool to analyze Instagram user data and answer questions posed by the management team. This project aims to use your SQL skills to extract meaningful insights from the data.

SQL TASKS

A) MARKETING ANALYSIS:

- Loyal User Reward:
- Inactive User Engagement:
- Contest Winner Declaration:
- Hashtag Research:
- Ad Campaign Launch:

B) INVESTOR METRICS

- User Engagement:
- Bots & Fake Accounts:

DATA ANALYSIS

STEP 1

Created and inserted the values in the database using DDL and DML commands. The raw data was provided by the product manager.

STEP 2

Understanding data means - How many tables are imported, their name and content in them?

There are *7* tables in total

TABLE NAME	COLUMN NAMES
USERS	*'id', 'username', 'created_at'*
TAGS	*'id', 'tag_name', 'created_at'*
PHOTOS	*'id', 'image_url', 'user_id', 'created_dat'*
PHOTO_TAGS	*'photo_id', 'tag_id'*
LIKES	*'user_id', 'photo_id', 'created_at'*
FOLLOWS	*'follower_id', 'followee_id', 'created_at'*
COMMENTS	*'id', 'comment_text', 'user_id', 'photo_id', 'created_at'*

SQL QUERIES

A) Marketing Analysis

Loyal User Reward: identifying the five Instagram users with the highest seniority based on the provided database. The marketing team wants to reward the most loyal users.

SQL QUERY

```
# TOP 5 USER OF INSTAGRAM ALONG WITH THIER USERNAME
SELECT
    *
FROM
    users
ORDER BY created_at
LIMIT 5;
```

OUTPUT

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

SQL QUERIES

A) Marketing Analysis

Inactive User Engagement: Instagram users who have not posted a single photo on the platform. The team aims to motivate inactive users to begin posting by sending them promotional emails.

SQL QUERY

```
# INSTAGRAM USERS WHO HAVE NEVER POSTED ON THE PLATFORM
SELECT
    id, username
FROM
    users
WHERE
    id NOT IN (SELECT
            user_id
        FROM
            photos);
```

OUTPUT

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

SQL QUERIES

A) Marketing Analysis

Contest Winner Declaration: The individual whose single photo gains the highest number of likes will be declared the contest winner, and their details will be furnished to the team.

SQL QUERY

```
# DECLARING A CONTEST WINNER WHO GETS THE MOST LIKES ON A SINGLE PHOTO.
WITH MOST_LIKE As
(
    SELECT
            PHOTO_ID,
        COUNT(PHOTO_ID) OVER ( PARTITION BY PHOTO_ID ) AS MAX_LIKES
    FROM
            LIKES
    ORDER BY
            MAX_LIKES DESC
    LIMIT 1),

PHOTOS_LIKE AS
(
    SELECT
        ML.MAX_LIKES,P.ID,P.IMAGE_URL,P.USER_ID
    FROM
        MOST_LIKE AS ML
    LEFT JOIN
        PHOTOS AS P
    ON P.ID = ML.PHOTO_ID)


SELECT
        U.ID, U.USERNAME,PL.MAX_LIKES,PL.IMAGE_URL
FROM
    USERS U
INNER JOIN
    PHOTOS_LIKE PL
ON PL.USER_ID = U.ID;
```

OUTPUT

ID	USERNAME	MAX_LIKES	IMAGE_URL
52	Zack_Kemmer93	48	https://jarret.name

SQL QUERIES

A) Marketing Analysis

Hashtag Research: The partner brand is seeking information on the most popular hashtags to incorporate in their posts for maximum reach. The task is to identify and recommend the top five most frequently used hashtags on the platform.

SQL QUERY

```sql
# Identify and suggest the top 5 most commonly used hashtags on the platform
with tag_detail as
(
    select
        photo_id,tag_id,
        count(tag_id) over (partition by tag_id ) as total_tags
    from
        photo_tags)


select
    tags.tag_name,tag_detail.total_tags as tags_count
from
        tag_detail
inner join
        tags
on tag_detail.tag_id = tags.id
group by
        tag_name
order by
        total_tags desc
limit 5;
```

OUTPUT

tag_name	tags_count
smile	59
beach	42
party	39
fun	38
lol	24

SQL QUERIES

A) Marketing Analysis

Ad Campaign Launch: The team is interested in finding the optimal day of the week to commence their ad campaigns. The task is to analyze and identify the day when the highest number of users register on Instagram, offering valuable insights to schedule the ad campaign effectively.

SQL QUERY

```
# What day of the week do most users register on?
SELECT
    DAYNAME(CREATED_AT) AS WEEKDAY,
    COUNT(DAYNAME(CREATED_AT)) AS USER_REGISTERED
FROM
    USERS
GROUP BY WEEKDAY
ORDER BY USER_REGISTERED DESC
LIMIT 2;
```

OUTPUT

WEEKDAY	USER_REGISTERED
Thursday	16
Sunday	16

SQL QUERIES

B) Investor Metrics

User Engagement: The investors are interested in determining the current level of user activity and posting frequency on Instagram, including any potential decline in the number of posts. The task is to calculate the average number of posts per user on the platform. Additionally, provide the ratio of the total number of photos on Instagram to the total number of users.

SQL QUERY

```
# 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 user
SELECT
    (SELECT
            COUNT(id)
        FROM
            photos) / (SELECT
            COUNT(DISTINCT (user_id))
        FROM
            photos) AS avg_number_of_post,
    (SELECT
            COUNT(id)
        FROM
            photos) / (SELECT
            COUNT(id)
        FROM
            users) AS photo_divide_by_users;
```

OUTPUT

avg_number_of_post	photo_divide_by_users
3.4730	2.5700

SQL QUERIES

B) Investor Metrics

Bots & Fake Accounts: The investors are seeking information about the presence of fake and inactive accounts on the platform. The task is to identify users, potentially bots, who have liked every single photo on the site, as this behaviour is not typical for a regular user.

SQL QUERY

```
# 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 AS Bot_Username,
    l.user_id AS ID,
    COUNT(user_id) AS Number_of_post_liked
FROM
    users AS u
        JOIN
    likes AS l ON u.id = l.user_id
GROUP BY user_id
HAVING Number_of_post_liked = 257;
```

OUTPUT

Bot_Username	ID	Number_of_post_liked
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

CONCLUSION

A) Marketing Analysis:

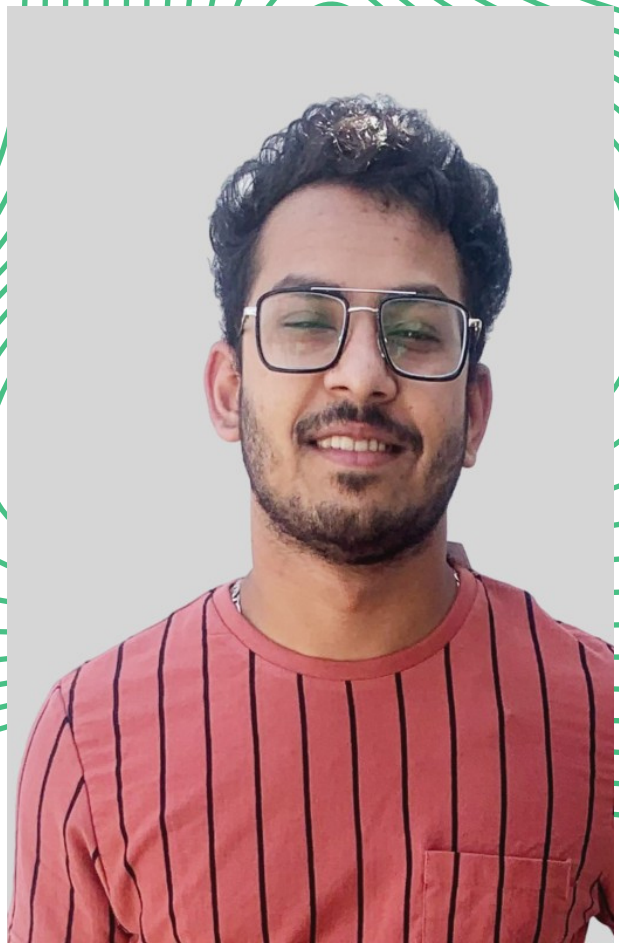
By rewarding loyal users, engaging inactive ones, leveraging user-generated content through contests, utilizing popular hashtags, and carefully planning ad campaign launches, the marketing team can enhance user engagement, expand reach, and foster a thriving community on Instagram.

B) Investor Metrics

To attract and retain investors, it is crucial for the platform to demonstrate a healthy level of user engagement. Identifying and addressing issues related to bots and fake accounts is essential to maintain the platform's integrity and ensure accurate engagement metrics. By actively monitoring and removing suspicious accounts, the platform can instill confidence in investors and create a more genuine and appealing user experience.

PAWAN RANA

DATA ANALYST



LINKEDIN - LINK

PHONE - 8445578888

MAIL - work.pawanrana@gmail.com

ABOUT ME

"As an accomplished working professional, I am eagerly seeking a dynamic role that harnesses my exceptional analytical prowess and astute business acumen to propel data-driven decision-making, critical in driving the success of the organization through effective communication, collaboration, and data-driven decision-making. I am highly motivated to collaborate with a cohort of accomplished professionals, where I can both broaden my knowledge and make meaningful contributions to the team's success."