PROJECT- INSTAGRAM USER ANALYTICS

-SQL Fundamentals

Project Description:

This project is about Instagram user analytics. This is to understand and track how users engage and interact with our digital product (software or mobile application). And from analyzing that data, business insights can be derived for the marketing, product and development teams across the business, which can be used for making better decisions like launching a new marketing campaign, deciding on features to build for an app, tracking the success of the app by measuring user engagement and improving the experience altogether that help in enhancing the business growth.

Through this project, I will be finding out the top 5 oldest users, users who never posted any photos, user who got the most likes on a single photo, most commonly used hashtags, weekday on which most users register on, and many other insights from the provided data, as asked by the team.

Approach:

I tried to understand the purpose of this project. I read each and every tasks provided by the team and understood what data they needed from it. SQL is used to perform the analysis. The dataset given by the team contains the commands for creating the database for doing Instagram user analysis. It contains the details of Instagram users, photos, likes, follows, comments, tags, photo tags, etc. I imported the data to the SQL database playground into the Schema SQL section and executed it. I analyzed the data completely and started querying. I performed several SQL commands and gained insights from the results got, which the marketing, product and development teams needed.

Tech-Stack Used:

I used DB Fiddle which is an online SQL database playground. It is a useful and free online SQL editor to learn and practice SQL coding for a wider range of databases. It also generates results quickly.

Insights:

While making the Instagram user analytics project, I performed several SQL queries as per the given tasks, gained many insights and knowledge, and provided a detailed report.

After getting the desired output data, I understood about the total users, user behaviour, user's age group, how much time they spend on Instagram, user engagement, active and inactive users, their searches, user interests, hashtags searched, photos they liked, suitable days to launch ad campaigns, etc. By analyzing this data, from a marketer's perspective, we can understand our target audience and implement apt social media marketing strategies like launching ad campaigns, rewarding the loyal users, sending promotional mails to inactive users, etc., to make our business product/service get noticed by the users and tempting them to purchase it.

Most of the people use Instagram for sharing their updates, photos, memories, entertainment, etc. And some use it for their business purposes. Instagram has become an effective marketing tool for promoting various business products/services. In Instagram, users usually discover new brand products through sponsored ads that pop up when they watch stories or reels. So product advertisements on Instagram help to achieve desired goals.

From an investor's perspective, I want to know if the platform is crowded with fake and dummy accounts. This is to make sure that our product's reach has been delivered to the actual people, not to bots or fake accounts. There is no use of sending ads to bots/fake accounts and it has no value. From the collected data from Instagram feed, I checked the user likes, interests, photos they posted or liked, products they liked, etc., on Instagram and based on that, I will be investing money in such products in which the chances of users buying it might be high. So it will be a good decision to make products based on user interests and to promote ads of such products for grabbing user attention. This helps in boosting the business growth and gaining profit.

Result:

While making this project, I have achieved the confidence to carry out the tasks provided by the team and got the required output. It has helped me to derive useful insights for the team, from the output data. This project has helped me to gain a clear understanding about SQL and helped me to use my imagination to improve my practical knowledge in it. I learned to apply real time SQL knowledge in solving the tasks of this project.

Below are each of the tasks given by the teams, the SQL queries used to carry out the tasks and their respective outputs.

A) Marketing:-

1. **Rewarding Most Loyal Users:** People who have been using the platform for the longest time.
   Task: Find the 5 oldest users of the Instagram from the database provided.

Query:

```
SELECT
        username, created_at
FROM
        ig_clone.users
ORDER BY
        created_at
LIMIT 5
```

Output:

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. **Remind Inactive Users to Start Posting:** By sending them promotional emails to post their 1st photo.
   Task: Find the users who have never posted a single photo on Instagram.

Query:

```
SELECT
        u.username
FROM
        ig_clone.users u
LEFT JOIN
        ig_clone.photos p
ON u.id = p.user_id
WHERE
     p.user_id is null
ORDER BY
        u.username;

```

Output:

username
Aniya_Hackett
Bartholome.Bernhard
Bethany20
Darby_Herzog
David.Osinski47
Duane60
Esmeralda.Mraz57
Esther.Zulauf61
Franco_Keebler64
Hulda.Macejkovic
Jadyn81
Janelle.Nikolaus81
Jessyca_West
Julien_Schmidt
Kasandra_Homenick
Leslie67
Linnea59
Maxwell.Halvorson
Mckenna17
Mike.Auer39
Morgan.Kassulke
Nia_Haag
Ollie_Ledner37
Pearl7
Rocio33
Tierra.Trantow

3. **Declaring Contest Winner:** The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.
   Task: Identify the winner of the contest and provide their details to the team.

Query:

```
SELECT username FROM
(
SELECT
       likes.Photo_id,
       users.username,
       COUNT(likes.User_id) AS like_user
FROM
       ig_clone.likes likes
INNER JOIN
       ig_clone.photos photos
ON
       likes.Photo_id = photos.id
INNER JOIN
       ig_clone.users users
ON
       photos.user_id = users.id
GROUP BY
       likes.Photo_id, users.username
ORDER BY
       like_user DESC
LIMIT 1
) base
```

Output:

username
Zack_Kemmer93

Zach_Kemmer93 is the winner of the contest.

4. **Hashtag Researching:** A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
   Task: Identify and suggest the top 5 most commonly used hashtags on the platform.

Query:

```
SELECT
       t.tag_name,
       COUNT(p.photo_id) AS num_tags
FROM
       ig_clone.photo_tags p
INNER JOIN
       ig_clone.tags t
ON p.tag_id = t.id
GROUP BY
       tag_name
ORDER BY
       num_tags DESC
LIMIT 5
```

Output:

tag_name	num_tags
smile	59
beach	42
party	39
fun	38
food	24

5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs.
   Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign.

Query:

```
SELECT
       WEEKDAY(created_at) AS weekday,
       COUNT(username) AS num_users
FROM
      ig_clone.users
GROUP BY 1
ORDER BY 2 DESC
```

Output:

weekday	num_users
3	16
6	16
4	15
1	14
0	14
2	13
5	12

0-Monday
1-Tuesday
2-Wednesday
3-Thursday
4-Friday
5-Saturday
6-Sunday

By analyzing the output generated, it can be seen that the more number of users register on Thursdays and Sundays of the week. So it is better to schedule ad campaigns on these days.

B) Investor Metrics:

1. **User Engagement:** Are users still as active and post on Instagram or they are making fewer posts.
   Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.

Query:

```
WITH CTE AS (SELECT u.id AS userid, COUNT(p.id) AS photoid
FROM ig_clone.users AS u
LEFT JOIN ig_clone.photos AS p ON u.id = p.user_id
GROUP BY u.id )
SELECT SUM(photoid) AS total_photos,
COUNT(userid) AS total_users,
SUM(photoid)/COUNT(userid) AS photos_per_user
FROM CTE
```

Output:

total_photos	total_users	photos_per_user
257	100	2.5700

2. **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts.
   Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

Query:

```
WITH photo_count AS (
  SELECT
         user_id,
         COUNT(photo_id) AS num_like
  FROM
       ig_clone.likes
  GROUP BY
       user_id
  ORDER BY
       num_like DESC )
SELECT * FROM photo_count
WHERE num_like =
(SELECT COUNT(*) FROM ig_clone.photos)
```

Output:

user_id	num_like
75	257
21	257
24	257
91	257
36	257
41	257
14	257
76	257
54	257
57	257
66	257
5	257
71	257