INSTAGRAM USER ANALYTICS

Contents

List of figures

List of tables

PROJECT DESCRIPTION

The focus of this project is on user analytics, which aims to analyse user behaviour. The process involves monitoring how users engage and interact with our digital products, which may include software or mobile applications. The primary goal is to gain valuable business insights that can be utilized by the marketing, product, and development teams.

Hypothetically a situation is given where I am working with the product team of Instagram, and the product manager has asked me to provide insights on the questions asked by the marketing team and management team. The project needs to be submitted in a week. Firstly, the marketing team needs information for their campaign. They are planning to initiate a marketing campaign, and they require diverse data from their database to execute it effectively and enhance customer engagement. They want to identify the target audience for their campaign and send them customized marketing and promotional mail. Moreover, they need other relevant information for branding partners so they can assist them in formulating a well-targeted, effective campaign. On the other side, the management team needs to gather data to present to their investors, specifically regarding the performance of the Instagram application and the user base. They require information on various aspects of user engagement, such as the number of active and inactive users, the average number of posts made by each user, and the total number of photos uploaded. Furthermore, they need information on the presence of fake and bot accounts to ensure that the app's user base consists of authentic customers.

After obtaining the necessary information required for the project or task, Oracle MySQL 8.0.33, a popular and widely used relational database management system (RDBMS), was used to perform structured query language (SQL) operations. In addition to using Oracle MySQL 8.0.33, an online database editor called "DB Fiddle" was also used.

In a nutshell, this project primarily caters to the marketing and management teams. The marketing team intends to initiate their marketing campaign, while the management team requires statistical data to effectively demonstrate the platform's efficiency for their potential investors.

APPROACH

This project primarily caters to the marketing and management teams. The marketing team intends to initiate their marketing campaign, while the management team requires statistical data to effectively demonstrate the platform's efficiency for their potential investors.

Upon comprehending the requirements of both teams, the initial course of action involved accessing the data to check its consistency and acquiring familiarity with the diverse tables within the database. Understanding how these tables can be effectively utilized to address the queries posed by each team was a pivotal aspect of this step. It is of paramount importance to ensure that I am working with appropriate and accurate data sets. Failing to undertake this critical preliminary phase may yield undesirable outcomes and result in project delays. By ensuring that I am working with the correct database and have access to comprehensive data which results into completion of the project.

Tech-Stack Used

To obtain the necessary information required for the project or task, Oracle MySQL 8.0.33, a popular and widely used Relational Database Management System (RDBMS) was used to perform Structured Query Language (SQL) operations. Oracle MySQL 8.0.33 is a powerful and flexible RDBMS that provides users with various tools and functionalities to work with databases efficiently. In addition to using Oracle MySQL 8.0.33, an online database editor called "DB Fiddle" was also used. DB Fiddle is an online platform that allows users to execute SQL queries on various databases, including MySQL. The editor provides a user-friendly interface that simplifies the process of executing SQL queries and retrieving data from databases.

FINDINGS

The initial five findings are primarily relevant to the marketing team, particularly for promoting their campaign.

Finding 1: Loyal users

The first finding suggests rewarding the most loyal users of the Instagram platform who have been using it for an extended period. This can be achieved by identifying the five oldest users of Instagram from the provided database and offering them incentives or special promotions to encourage them to continue using the platform. The SQL query for finding the most loyal users is below

```
SELECT
    users.id AS 'User_Id',
    users.username AS 'User_name',
    users.created_at AS 'Account_creation_date'
FROM
    ig_clone.users
ORDER BY created_at ASC
LIMIT 5;
```

The table called "users" was used to find the most loyal users from the "ig_clone" database. The table "users" contain the information of the Instagram users such their user_id, their users_name and creation date of their account. Now, in order to get oldest users, data was sorted by the creation date in ascending manner (ASC). So, this will produce the most loyal and the oldest users of the platform. As a result, the five oldest users eligible for the rewards are the Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71 and Jordyn.Jacobson2.

User_Id	User_name	Account_creation_date
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

Table 1: Five oldest Users on the Platform.

Finding 2: Inactive users

The marketing team wants to encourage inactive users who have not yet posted anything on Instagram to start using the platform. They plan to achieve this by sending promotional emails to these users and encouraging them to begin posting. Essentially, the goal is to re-engage with users who have not yet fully utilized the platform and encourage them to start using it actively. Now the task is finding the users who have never posted a single photo on Instagram. The image illustrates the SQL query for finding inactive users on the platform.

```
SELECT
    users.id AS 'User_ID',
    username AS 'User_name',
    created_at AS 'Account_creation_date'
FROM
    ig_clone.users
        LEFT JOIN
    ig_clone.photos ON users.id = photos.user_id
WHERE
    photos.user_id IS NULL
ORDER BY users.id ASC;
```

Inactive users are individuals who have not contributed any content on the platform. To identify these users, two tables, namely "users" and "photos," were utilized. The "users" table was used to gather information about inactive users by setting a condition on the "photos" table. This was accomplished by performing a LEFT JOIN operation between the "users" table and the "photos" table. By doing so, it became possible to identify users who have not posted any content and classify them as inactive users. To distinguish between active and inactive users, a condition was applied based on their photo upload activity. In this analysis, users who have uploaded at least one picture are considered active, while those who have not uploaded any photos are classified as inactive. This distinction was made by evaluating the user ID associated with each photo, where a null value indicates an inactive user. This approach allows for a clear differentiation between users who actively contribute content to the platform and those who have not yet done so. This approach helped in obtaining a comprehensive list of users who have not actively engaged on the platform. As a result, I found in total 26 inactive users who have been inactive on the platform, they did not even upload a single photo.

User_ID	User_name	Account_creation_date
5	Aniya_Hackett	2016-12-07 01:04:39
7	Kasandra_Homenick	2016-12-12 06:50:08
14	Jaclyn81	2017-02-06 23:29:16
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
25	Tierra.Trantow	2016-10-03 12:49:21
34	Pearl7	2016-07-08 21:42:01
36	Ollie_Ledner37	2016-08-04 15:42:20
41	Mckenna17	2016-07-17 17:25:45
45	David.Osinski47	2017-02-05 21:23:37
49	Morgan.Kassulke	2016-10-30 12:42:31
53	Linnea59	2017-02-07 07:49:34
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
68	Franco_Keebler64	2016-11-13 20:09:27
71	Nia_Haag	2016-05-14 15:38:50
74	Hulda.Macejkovic	2017-01-25 17:17:28
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
80	Darby_Herzog	2016-05-06 00:14:21
81	Esther.Zulauf61	2017-01-14 17:02:34
83	Bartholome.Bernhard	2016-11-06 02:31:23
89	Jessyca_West	2016-09-14 23:47:05
90	Esmeralda.Mraz57	2017-03-03 11:52:27
91	Bethany20	3:31:53

Table 2: Inactive Users on the Platform.

Finding 3: Declaring Contest Winner

The marketing team organised a contest where participants submitted their photos, and the winner would be determined by the number of likes their photo received. Now, the team is interested in finding out information about the user who obtained the highest number of likes on a single photo from the database and declare that user to be the winner of the contest. Below is the SQL query for finding the winner of the contest.

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

In order to determine the winner of the contest, three tables from the database were utilized. The reason for this is that there is no direct linkage between the "users" table, which contains information about the participants, and the "likes" table, which stores the number of likes for each photo. To bridge this gap, the "photos" table serves as a mediator, connecting the "users" and "likes" tables. To establish the connection between these three tables and identify the contest winner, an "INNER JOIN" was employed. This query combines the relevant data from the "users," "photos," and "likes" tables based on foreigner keys. By using this "INNER JOIN" operation, the system is able to find the user who received the maximum number of likes on a single photo and declare them as the winner of the contest. This approach ensures that the

marketing team can accurately determine the winner by leveraging the data stored in multiple tables and establishing the necessary connections through the "photos" table. The result show that Zack_Kemmer93 is the winner contest with impressive 48 likes on a single photo.

Winner_username	Photo_id	Max_likes
Zack_Kemmer93	145	48

Table 3: Declare Winner of the contest.

Finding 4: Hashtag Researching

The brand partner is interested in identifying the most popular hashtags on the platform. Users on the platform attach keywords known as "tags" to their posts, and these tags play a crucial role in helping the brand partner promote more content to users with similar interests. The task at hand is to discover and present the most frequently used hashtags on the platform to the marketing team and brand partner. This information will assist them in effectively targeting their content and engaging with the platform's user base. The following query is leveraged to fetch the top five hashtags.

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

To determine the top five hashtags, two tables were employed. The first table, named "tags," stores information about the tags that users attach to their posts during upload. Each tag is associated with a unique identification code, such as "21" for the "smile" hashtag. However, to calculate the frequency of each tag, another table called "photo_tags" is needed. This table establishes the connection between tags and the specific photos or posts they are used on.

By utilizing the INNER JOIN clause, the "tags" and "photo_tags" tables are combined. This operation allows for matching the tags used on each post and determining their count. The joined tables enable the system to identify which tags appear on which posts and count the occurrences of each tag. This process helps identify the most frequently used hashtags on the platform.

The resulting information, such as the counts of each tag, can then be used to determine the top five hashtags. This data provides valuable insights to the marketing team and brand partner, allowing them to optimize their content strategy and target their audience effectively.

Popular_Tags	Tag_Count
smile	**59**
beach	**42**
party	**39**
fun	**38**
concert	**24**

Table 4: Mostly used hashtags.

Finding 5: Launch ADs Campaign

The team and brand partners are interested in determining the most suitable day of the week to launch their advertisement campaign. Additionally, the marketing team wants to identify the day when the highest number of users joined their platform. This information is crucial for maximizing the reach of their promotional emails and encouraging inactive users to become active participants on the platform. The primary objective is to find the specific day of the week with the highest user registration rate. The query for finding the most appropriate day of the week is following below.

```
SELECT
    DAYNAME(users.created_at) AS Week_Dayname,
    COUNT(users.username) AS Users_registered
FROM
    ig_clone.users
GROUP BY DAYNAME(users.created_at)
ORDER BY COUNT(users.username) DESC
```

To determine the day of the week when users joined the platform, we utilized the "users" table which contains the account creation date. In order to convert the date into the corresponding day name, I employed the "DAYNAME" function. By using the "COUNT" clause, we were able to accurately calculate the number of users who joined the platform on each specific day. This approach provided us with precise measurement and analysis of user registration patterns. The findings revealed that Thursday and Sunday has the highest number of user registrations. So ideally, Thursday and Sunday would be the two most suitable days to run the advertisement campaign.

Week_Dayname	User_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Table 5: Launch Campaign Suitable Day

The further following findings represent the data for management team which they want to present to investors.

Finding 6: Investor Metrics

The management team is seeking user engagement statistics to present to the investors. The investors are interested in assessing the performance of Instagram and ensuring that it is not becoming obsolete like Facebook. They want to evaluate the app based on several factors. One of these factors is the level of user engagement on the platform, which can be measured by certain criteria such as the number of active users, the total number of posts on the platform, and the average number of posts per user. To gather this information from the database and determine the efficiency of user engagement, the following query has been employed.

```
WITH UE AS
(
SELECT
    users.id AS User_Count,
        COUNT(photos.user_id) AS Photos_Upload
FROM
    ig_clone.users
LEFT JOIN
    ig_clone.photos
ON
    users.id=photos.user_id
GROUP BY
    users.id
)
SELECT
COUNT(User_Count) AS Total_Users,
        SUM(Photos_Upload) AS Total_Photos_Uploaded,
    SUM(Photos_Upload)/COUNT(User_Count) AS Average_photos_per_User
FROM
    UE
```

Figure 1: Investor metrics

Finding 7: Bots & Fake Accounts

The investors are interested in understanding the level of user engagement on the platform. They want to determine whether the platform is being utilized by genuine users or if it contains a large number of fake and inactive accounts. To identify these accounts, a strategy is proposed: accounts that excessively like pictures on the platform are considered potential bots or fake accounts, as it is highly unlikely for a regular user to like every single picture available.

The investors seek to gauge the authenticity and active participation of users on the platform. They want to ensure that the platform's popularity and engagement are driven by genuine users rather than artificially inflated metrics. To achieve this, the focus is placed on identifying accounts that exhibit unusual behaviour, such as indiscriminately liking a large number of pictures.

```sql
WITH Likes_Count AS
  (
    SELECT
        users.username AS Username,
        likes.user_id AS Account_Id,
        COUNT(likes.photo_id) AS Photos_Likes
    From
        ig_clone.likes
    LEFT JOIN
        ig_clone.users
    ON
        likes.user_id=users.id
    GROUP BY
        user_id
    ORDER BY
        Photos_Likes DESC
  )
SELECT *
FROM
   Likes_Count
WHERE
   Photos_Likes=257
```

Username	Account_Id	Photos_Likes
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

Table 6: Fake and Bot Account

SUMMARY

The results are presented in accordance with the question inquired by the marketing and management team.

- **Loyal Users:** There are five individuals who demonstrate remarkable loyalty as users of the Instagram application. These esteemed users have maintained their association with the platform since its inception. Their respective names are **Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71, and Jordyn.Jacobson2**. It is noteworthy that these five individuals represent the earliest adopters of the platform who continue to actively engage with it.
- **Inactive users:** There is a group of 26 users who are currently inactive on the platform. It is recommended that the marketing team sends them promotional emails with the objective of re-engaging them and encouraging their return to the platform. These users' details can be found in the table 2.
- **Winner contest: Zack_Kemmer93** has emerged as the victor in the contest, securing the top position due to the exceptional number of likes received on a single photograph. Among the numerous entries, it was Zack_Kemmer93's image with the unique identifier 145 that garnered the most attention, accumulating an impressive total of 48 likes.
- **Frequent hashtag:** The hashtags most commonly utilized on the platform are organized in descending order as follows: **Smile, beach, party, fun, and concert.**
- **Campaign day**: Both **Thursday** and **Sunday** attract the highest influx of new users to the platform. Consequently, either of these days would be deemed suitable for the marketing team to dispatch promotional emails.
- **Investor metrics**: Regarding the user engagement on the platform, there is a total of **100 users** registered. Out of these users, **26 are currently inactive**. As for the content, a total of **257 photos** have been uploaded to the platform. On average, each user has uploaded approximately **2.57 photos**. These statistics provide insights into the level of user activity and content generation on the platform. Moreover, there **13 fake or bot accounts** operating on the platform.