INSTAGRAM USER ANALYTICS

Description:

This project centres around user analysis, focusing on tracking and understanding how users interact with a digital product—in this case, the Instagram app. The analysis aims to provide valuable insights that can be leveraged by different teams within the business to make informed decisions such as Marketing Team uses insights to launch new campaigns, Product Team utilizes data to decide on new features to develop, Development Team enhances the overall user experience based on the findings.

In this project, SQL and MySQL Workbench are the primary tools for analysing Instagram user data. The goal is to answer specific questions posed by the management team, thereby aiding the product manager and the rest of the team in making strategic decisions about the app's future development. The insights you derive could significantly impact the evolution of one of the world's most popular social media platforms.

Problem Statements:

Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

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

- Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.

  Your Task: Find the users who have never posted a single photo on Instagram.
- 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.

  Your Task: Identify the winner of the contest and provide their details to the team.
- Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

  Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform.
- Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.

  Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign.

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

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

  Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Design:

Steps taken to load the data into the database:

- "CREATE DATABASE" - Used to create the database.
- Adding tables and Columns names on the database.
- "INSERT INTO" – Used to add values to the tables.
- "SELECT" – Used for the Querying the data to get the desired result.

❖ Software used – MySQL Workbench 8.0 CE

1. To find the most loyal users i.e. the top 5 oldest users of Instagram:

- We will use the data from the users table by selecting the "username" and "created_at" columns.
- Then using the "ORDER BY" function we will order the desired output by sorting with the "created_at" column in ascending order "ASC".
- Then using the "LIMIT" function, the output will be displayed for top 5 oldest Instagram users.

Query:

```
SELECT username,
        created_at
FROM ig_clone.users
ORDER BY created_at ASC
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. To Find the most inactive users i.e. the users who have never posted a single photo on Instagram:

- We will first select username column from the users table.
- Simplifying the users and photos table name by giving alias u and p respectively.
- Then we will "LEFT JOIN" photos table on the users table, on u.id = p.user_id because, both the u.id and p.user_id have common values in them.
- Then we will find rows from the users table where the p.id "IS NULL".

Query:

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

Output:

username	user_id
Aniya_Hackett	5
Kasandra_Homenick	7
Jadyn81	14
Rocio33	21
Maxwell.Halvorson	24
Tierra.Trantow	25
Pearl7	34
Ollie_Ledner37	36
Mckenna17	41
David.Osinski47	45
Morgan.Kassulke	49
Linnea59	53
Duane60	54
Julien_Schmidt	57
Mike.Auer39	66
Franco_Keebler64	68
Nia_Haag	71
Hulda.Macejkovic	74
Leslie67	75
Janelle.Nikolaus81	76
Darby_Herzog	80
Esther.Zulauf61	81
Bartholome.Bernhard	83
Jessyca_West	89
Esmeralda.Mraz57	90
Bethany20	91

3. To find out the Winner of the contest by sorting out the image got most number of likes:

- First we will select the u.username, p.id, p.image_url and "COUNT(*)" as Total.
- Then, we will "INNER JOIN" the three tables : photos, likes and users, on l.photo_id = p.id and p.user_id = u.id
- Then, by using "GROUP BY" function we will group the output on the basis of p.id
- Then, using "ORDER BY" function we will sorting the data on the basis of the total in descending order "DESC".
- Then, to find the most liked photo we will using "LIMIT" function to view only the top liked photo's information.

Query:

```
SELECT u.id AS user_id,
       u.username,
       p.id AS photo_id,
       p.image_url,
       COUNT(*) AS Total
FROM ig_clone.users u
INNER JOIN photos p
    ON p.user_id = u.id
INNER JOIN likes l
    ON l.photo_id = p.id
GROUP BY p.id
ORDER BY Total DESC
LIMIT 1;
```

Output:

user_id	username	photo_id	image_url	Total
52	Zack_Kemmer93	145	https://jarret.name	48

4. To find the top 5 most commonly used hashtags on Instagram:

- We need to select the tag_name column from the tag table and the "COUNT(*)" as total function so as to count the number of tags used individually.
- Then, we need to join tags table and photo_tags table, on t.id = pt.tag_id cause they contain the same content in them i.e. tag_id
- Then using the "GROUP BY" function we need to group the desired output on the basis of t.tag_name
- Then using the "ORDER BY" function we need to sort the output on the basis of total(total number of tags per tag_name) in descending order "DESC"
- Then, to find the top 5 most used tag names we will use the "LIMIT" 5 function.

Query:

```
SELECT tag_name,
       COUNT(*) AS Total_no_of_tags_used_by_individually
FROM ig_clone.tags t
JOIN photo_tags pt
    ON t.id = pt.tag_id
GROUP BY t.tag_name
ORDER BY Total_no_of_tags_used_by_individually DESC
LIMIT 5;
```

Output:

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

5. To find the day of week on which most users register on Instagram:

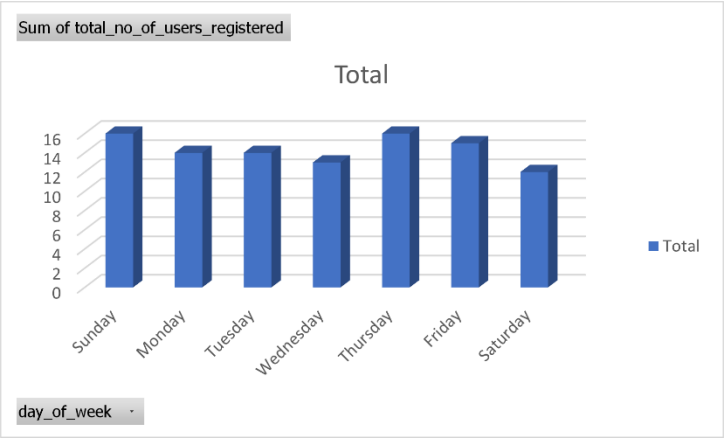
- First we define the columns of the desired output table using select "dayname(created_at)" as day_of_week and COUNT(*) as total_number_of_users_registered from the users table
- Then using the "GROUP BY" function we group the output table on the basis of day_of_week
- Then using the "ORDER BY" function we order/sort the output table on the basis of total_number_of_users_registered in descending order "DESC"

Query:

```
SELECT
    dayname(created_at) AS day_of_week,
    COUNT(*) AS total_no_of_users_registered
FROM ig_clone.users
GROUP BY day_of_week
ORDER BY total_no_of_users_registered DESC;
```

Output:

day_of_week	total_no_of_users_registered
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12



6. To find the how many times does average posts on Instagram:

- First, we need to find first the count number of photos(posts) that are present in the photos.id column of the photos table i.e. "COUNT(*)" from photos
- Similarly, we need to find the number of users that are present in the users.id column of the users table i.e. "COUNT(*)" from users
- Next, we need to divide both the values i.e. "COUNT(*)" from photos/"COUNT(*)" from users and hence we would get the total number of photos / total number of users
- To find how many times the users posts on Instagram we need to find the total occurrences of each user_id in photos table

Query:

```
SELECT COUNT(*) AS total_post_count
FROM ig_clone.photos
GROUP BY photos.id;

SELECT COUNT(*) AS total_users
FROM ig_clone.users
GROUP BY users.id;

SELECT
(SELECT COUNT(*) FROM photos) / (SELECT COUNT(*) FROM users) AS average_photos_per_user;

SELECT user_id,
       COUNT(*) AS total_posts
FROM ig_clone.photos
GROUP BY user_id;
```

Outputs:

average_photos_per_user
2.5700

user_id	total_posts
1	5
2	4
3	4
4	3
6	5
8	4
9	4
10	3
11	5
12	4
13	5
15	4
16	4
17	3
18	1
19	2
20	1
22	1
23	12
26	5
27	1
28	4

user_id	total_posts
29	8
30	2
31	1
32	4
33	5
35	2
37	1
38	2
39	1
40	1
42	3
43	5
44	4
46	4
47	5
48	1
50	3
51	5
52	5
55	1
56	1
58	8

user_id	total_posts
59	10
60	2
61	1
62	2
63	4
64	5
65	5
67	3
69	1
70	1
72	5
73	1
77	6
78	5
79	1
82	2
84	2
85	2
86	9
87	4
88	11
92	3

93	2
94	1
95	2
96	3
97	2
98	1
99	3
100	2

7. To find the fake and dummy accounts:

- First, we select the user_id column from the photos table Then we select the username column from the a table
- Then, we select the "COUNT(*)" function to count total number of likes from the likes table
- Then we "INNER JOIN" users and likes table on the basis of u.id and l.user_id, using the "ON" function
- Then by using the "GROUP BY" function we group the desired output table on the basis of l.user_id
- Then, we search for the values from the "COUNT(*)" from photos having equal values with the total_likes_per_user

Query:

```
SELECT
    u.id AS user_id,
    u.username,
    COUNT(l.user_id) AS total_likes_per_user
FROM ig_clone.users u
INNER JOIN ig_clone.likes l
    ON u.id = l.user_id
GROUP BY u.id, u.username
HAVING  COUNT(l.user_id) = (SELECT COUNT(*) FROM ig_clone.photos)
ORDER BY u.id;
```

Output:

user_id	username	total_likes_per_user
5	Aniya_Hackett	257
14	Jadyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

Analysis:

- The most loyal users i.e. the top 5 oldest users are Darby_Herzog, Emilo_Bernier52, Elenor88, Nicole71 and Jordyn.Jacobson2.
- Out of the 100 total users there are 26 users who are inactive and they have never posted any kind of stuff of Instagram may it be any photo, video or any type of text. So, the Marketing team of Instagram needs to remind such inactive users
- The user named Zack_Kemmer93 with user_id 52 is the winner of the contest because his photo with photo_id 145 has the highest number of likes - 48
- The top 5 most commonly used #hashtags along with the total count are smile(59), beach(42), party(39), fun(38) and concert(24)
- Most of the users registered on Thursday and Sunday i.e. 16 and hence it would prove beneficial to start AD Campaign on these two days
- There are in total 257 rows i.e. 257 photos in the photos table and 100 rows i.e. 100 ids in the users table which makes the desired output to be 257/100 = 2.5700 (average users posts on Instagram)
- Out of the total user id's there are 13 such user id's who have liked each and every post on Instagram (which is not practically possible by a normal user) and so such user id's are considered as Dummy and Fake Accounts.

Root Cause Analysis – 5 Whys :

1. Why did the Marketing team want to know the most inactive users?

➔ So, they can reach out to those users via email and ask them what's keeping them away from using Instagram, with the aim of re-engaging them and improving user retention.

2. Why did the Marketing team want to know the top 5 #hashtags used?

➔ They wanted to identify the most popular hashtags, possibly to enable the tech team to add filter features for photos and videos posted using these top 5 hashtags, enhancing user experience and engagement.

3. Why did the Marketing team want to know on which day of the week the platform had the most new users registered?

➔ To optimize ad placement by running more advertisements on days with the highest new user registrations, thereby maximizing ad reach and profit.

4. Why did the Investors want to know about the average number of posts per user on Instagram?

➔ To assess user engagement levels on the platform, which is crucial for determining the platform's value and growth potential. It also helps the tech team plan for infrastructure needs to handle user traffic efficiently.

5. Why did the Investors want to know the count of BOTS and Fake accounts, if any?

➔ Ensure that their investment is going into a platform with a genuine and active user base, not inflated by bots or fake accounts, thereby protecting their investment from future liabilities.

CONCLUSION

In conclusion, it's evident that not only Instagram but many other social media platforms and commercial firms utilize such analysis to extract valuable insights from their customer data. This practice helps firms identify customers who will be assets in the future rather than liabilities.

Companies perform this analysis and customer segmentation on a weekly, monthly, quarterly, or yearly basis, depending on their business needs. This strategic approach aims to maximize future profits while minimizing costs.