Sai Arjun Reddy Kumbam

Project 2 : Instagram User Analytics

Description

User Analysis is a process by which we track how users engage and interact with our digital product (software or mobile applications) in an attempt to derive business insights for marketing, product & development teams.

These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user management and Improve the experience altogether while helping the business grow.

You are working with the product team of Instagram and the product manager has asked you to provide insights on the questions asked by the management team.

The Problem

A) **Marketing:** The marketing team wants to launch some campaigns, and they require our help in assisting 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: Identify 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 out to most audience on 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

B) **Investor Metrics:** Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to access the app on the following grounds:

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

- **Bots & Fake Accounts:** The investors want to know if the platform is crowded with fake and dummy accounts.
  Your 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).

Design

Steps taken to load the data into the database,

- Using the 'create db' function of MySQL to create a database
- Then add tables and column names
- Then add the values into them using the 'insert into' function of MySQL
- By using the 'select' command we can query the desired output.

Software used for querying the results:

- MySQL Workbench 8.0 CE

Findings – I

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

1. We will use the data from the users table by selecting the username and created_at columns.

2. Then using the order by function we will order the desired output by sorting with the created_at column in ascending order.

3. Then using the limit function, the output will be displayed for top 5 oldest Instagram users.

Output/Result:

username	created_at
Darby_Herzog	08-07-2023 00:14
Emilio_Bernier52	08-07-2023 13:04
Elenor88	09-07-2023 01:30
Nicole71	09-07-2023 17:30
Jordyn.Jacobson2	10-07-2023 07:56

Findings – II

To find the most inactive users i.e., the users who have never provided a single photo on Instagram.

1. We will first select username column from the users table.

2. Then we will left join photos table on the users table, on users,id = photos.user_id because, both the users.id and photos.user_id have common contents in them.

3. Then we will find rows from the users table where the photos.id IS NULL.

username	user_id
Aniya_Hackett	5
Kasandr_Homenick	7
Jaclyn81	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	82
Jessyca_West	89
Esmeralda.Mraz57	90
Bethany20	91

Findings – III

To find the most the username, photo_id, image_url and total_number_of_likes of that image:

1. First, we will select the users.username, photo.id, photos. Image_url and count(*) as total.

2. Then, we'll inner join the three tables wiz: photos, likes and users, on likes.photo_id = photos.id and photos.user_id = user.id

3. Then, by using group by function we will group the output on the basis of photos.id

4. Then, using order by function we will be sorting the data on the basis of the total in descending order.

5. Then, to find the most liked photo we will be using limit function to view only the top liked photo's information.

Output/Result:

userd_id	username	photo_id	image_url	total
52	Zack_Kemmer93	145	https://jarret.name	48

Findings – IV

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

1. We need to select he tag_name column from the tag table and the count(*) as total function so as to count the number of tags used individually.

2. Then, we need to join tags table and photo_tags table, on tags.id = photo_tags.tag_id because they contain the same content in them i.e., tag_id.

3. Then using the group by function, we need to group the desired output on the basis of tags.tag_name

4. Then using the order by function we need to sort the output on the basis of total (total no.of tags per tag_name) in descending order.

5. Then, to find the top 5 most used tag names we will use the limit 5 function.

Output/Result:

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

Findings – V

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

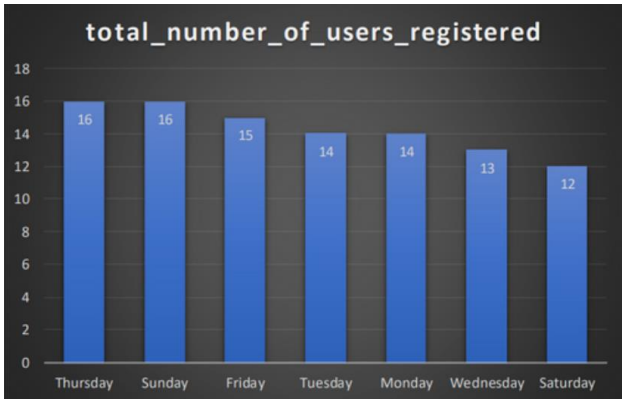
1. First we define the columns of the desired output table using the select dayname(created_at) as day_of_week and count(*) as total_number_of_users_registered from the users table .

2. Then using the group by function we group the output table on the basis of day_of_week.

3. Then using the order by function we order/sort the output table on the basis of total_number_of_users_registered in descending order.

Output/Result

day_of_week	total_number_of_users_registered
Thusrday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12



Findings – VI

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

1. First, we need to find the count of number of photos(posts) that are present in the photos.id column of the photos table i.e. count(*) from photos.

2. 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.

3. 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.

4. To find how many times the users posts on Instagram we need to find the total occurrences of each user_id in photos table.

Output/Result:

total_photos_divide_total_photos
2.57

Findings – VI (Continues.)

user_id	user_post_count
1	5
2	4
3	4
4	3
5	3
6	5
7	4
8	4
9	4
10	3
11	5
12	4
13	5
14	3
15	4
16	4
17	3
18	1
19	2
20	1
21	1
22	1
23	12
24	2
25	6
26	5
27	1
28	4
29	8
30	2
31	1
32	4
33	5
34	4
35	2
36	5
37	1
38	2
39	1
40	1
41	1
42	3
43	5
44	4
45	3
46	3
47	5
48	1
49	1
50	3
50	3
51	5
52	5
53	3
54	1
55	1
56	1
57	8
58	8
59	10
60	2
61	1
62	2
63	4
64	5
65	5
66	5
67	3
68	1
69	1
70	1
71	2
72	5
73	1
74	1
75	2
76	1
77	6
78	5
79	1
80	3
81	4
82	2
83	1
84	2
85	2
86	9
87	4
88	11
89	6
90	2
91	1
92	3
93	2
94	1
95	2
96	3
97	2
98	1
99	3
100	2

Findings – VII

To find the bots and fake accounts:

1. First, we select the user_id column from the photos table

2. Then we select the username column from the users table

3. Then, we select the count(*) function to count total number of likes from the likes table.

4. Then, we inner join users and likes table on the bases of users.id and likes.user_id, using the on function/clause.

5. Then, by using he group by function we group the desired output table on the basis of likes.user_id.

6. Then, we search for the values from the count(*) from photos having equal values with the total_likes_per_user.

Output/Result:

user_id	user_name	total_likes_per_user
5	Aniya_Hackett	257
14	Jaclyn81	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

After performing the analysis I have the following points:-

- The most loyal users i.e, the top 5 oldest users are:

username	created_at
Darby_Herzog	08-07-2023 00:14
Emilio_Bernier52	08-07-2023 13:04
Elenor88	09-07-2023 01:30
Nicole71	09-07-2023 17:30
Jordyn.Jacobson2	10-07-2023 07:56

- 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.

- So, the user named Zack_Kemmer93 with user_id 53 is the winner of the contest cause his photo with photo_id 145 has the highest number of likes i.e. 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 Saturday i.e. 16 and hence it would prove beneficial to start AD campaign on these two days.

- So, 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 = 2.57 (avg users posts in 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) and so such user id's are considered as BOTS and fake accounts.

Using the 5 Whys approach I'm finding the root cause of the following:

- Why did the Marketing team wanted to know the most inactive users?
- → So, they can reach out to those users via mail and ask them What's keeping them away from using the Instagram.
- Why did the Marketing team wanted to know the top 5 #hashtags used?
- → May be the tech team wanted to add some filter features for photos and videos posted using the top 5 mentioned #hashtags
- Why did the Marketing team wanted to know on which day of the week the platform had the most new users registered?
- → So, that they can run more Ads of various brands during such days and also get profit from it.
- Why did the investors wanted to know about the average posts per user on Instagram ?
- → It's a fact that every brand on social platform is determined by the user engagement on such platforms, also investors wanted to know whether the platform has the right and authenticated user base. It also helps the tech team determine how to handle such traffic on the platform with the latest tech without disrupting the smooth and efficient functioning of the platform.
- Why did the investors wanted to know the count of BOTS and Fake accounts if any?
- → So that the investors are assured that they are investing into an Asset and not a future liability.

Conclusion

In conclusion, I would like to conclude that not only Instagram but many other social media and commercial firms use such Analysis to find the insights from their customer data which in turn help the firms to find the customers who will be an Asset to the firm in the future and not some Liability.

Such Analysis and sorting of the customer base is done at an weekly, monthly, quarterly or yearly basis as per the needs of the business firms so as to maximize their profits in future with minimal cost to the company