INSTAGRAM USER ANALYTICS

Name :- Yash Shinde

Email :- yashpradeepshinde@gmailcom

TABLE OF CONTENT

- □ Project Description
- □ Approach
- □ Tech Stack Used
- □ Analysis
    - □ (A) Marketing
    - □ (B) Investor Metrics
- □ Insights
- □ Conclusion & Result

PROJECT DESCRIPTION

PROJECT DESCRIPTION

- User analysis involves monitoring and examining user interactions and behaviours within our digital products, such as software or mobile applications.
- The primary objective is to extract valuable business intelligence for marketing, product, and development teams.
- By leveraging these insights, various departments throughout the organization can make informed decisions. Such as,
- launching effective marketing campaigns, determining which features to implement in the app, measuring the app's performance through user engagement metrics, and enhancing the overall user experience.
- In this project I have analyzed the Instagram dataset to find valuable insights from the data.

APPROACH

APPROACH

- My approach through this project would be to solve the questions through SQL queries using the MySQL Workbench.
- I will provide a detailed explanation of the query along with the results of each query.
- I will perform my analysis using the following list of points.

(A) Marketing:-

1. Rewarding Most Loyal Users
2. Remind Inactive Users to Start Posting
3. Declaring the contest winners
4. Hashtag Researching
5. Launch AD Campaign

(B) Investor Metrics:-

1. User Engagement
2. Bots and Fake Accounts

TECH STACK USED

- MySQL Workbench
- Microsoft PowerPoint

ANALYSIS

(A) Marketing
(B) Investor Metrics

MARKETING

1. Rewarding Most Loyal Users
2. Remind Inactive Users to Start Posting
3. Declaring the contest winners
4. Hashtag Researching
5. Launch AD Campaign

Marketing

REWARDING MOST LOYAL USERS : PEOPLE WHO HAVE BEEN USING THE PLATFORM FOR THE LONGEST TIME. THAT MEANS, FINDING 5 OLDEST USERS OF THE INSTAGRAM FROM THE DATABASE.

- Steps for finding the most loyal users of Instagram (i.e. 5 oldest users) :

  Step 1) I have selected the ig_clone database for accessing data (use ig_clone).

  Step 2) I have used the data from the users table by selecting the username and created_at columns.

  Step 3) Then used "IS NOT NULL" command to filter empty usernames.

  Step 4) Then used "ORDER BY" command to order the result in ascending order.

  Step 5) Then used "LIMIT" command to select top 5 results.

Query (to find 5 oldest Instagram users) :

```
use ig_clone;
SELECT
   username, created_at
FROM
   users
WHERE
   username IS NOT NULL
ORDER BY created_at ASC
LIMIT 5;
```

Marketing

REWARDING MOST LOYAL USERS : PEOPLE WHO HAVE BEEN USING THE PLATFORM FOR THE LONGEST TIME. THAT MEANS, FINDING 5 OLDEST USERS OF THE INSTAGRAM FROM THE DATABASE.

Output/Result Table :

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

Marketing

REMIND INACTIVE USERS TO START POSTING : REMINDING USERS WHO HAS NOT POSTED ANY PHOTO ON INSTAGRAM. THAT MEANS, FINDING USERS WHO HAVE NEVER POSTED A SINGLE PHOTO ON INSTAGRAM.

- Steps for finding the inactive users (i.e. users who have never posted a single photo on Instagram.) :

  Step 1) I have used data from users table by selecting columns username and id.

  Step 2) Then I have used WHERE statement along with NOT IN statement to filter those columns.

  Step 3) Inside the IN statement I have selected DISTINCT user_id from photos column. (Thus we can get those users who have posted the photos.)

  Step 4) Then it is filtered by WHERE statement. (Which gives only user_ids who have not posted any photo yet

  Step 5) Then I used ORDER BY command to order the result by user ids.

Query/Program :

```
SELECT username, id AS user_id
FROM users
WHERE id NOT IN (
   SELECT DISTINCT user_id
   FROM photos
)
ORDER BY id;
```

Marketing

Output/Result Table :

username	user_id
Aniya_Hackett	5
Kasandra_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	83
Jessyca_West	89
Esmeralda.Mraz57	90
Bethany20	91

Marketing

DECLARING CONTEST WINNER : THE USER WHO GETS THE MOST LIKES ON A SINGLE PHOTO WILL WIN THE CONTEST. IDENTIFYING THE WINNER OF THE CONTEST BY FINDING THAT USER WHO HAS GOT MOST LIKES ON A SINGLE PHOTO.

- Steps for finding the contest winner (i.e. users who have got most likes on a single photo on Instagram.) :

  Step 1) I have selected 'id' & 'username' columns from the 'users' table & 'id' column from photos table.

  Step 2) Then I have performed INNER JOIN of three tables (users, photos, likes) .

  Step 3) Then I have applied GROUP BY command on the id column of photos (i.e. photos.id)

  Step 4) Then I have ordered the result by using ORDER BY statement in descending order.

  Step 5) Provide limit of 1. Since, we want the photo with the highest number of likes.

Query/Program :

```
SELECT users.id AS user_id, users.username,
photos.id AS photo_id,
        count(*) AS total
FROM photos
INNER JOIN likes ON likes.photo_id = photos.id
INNER JOIN users ON photos.user_id = users.id
GROUP BY photos.id
ORDER BY total DESC
LIMIT 1;
```

Marketing

DECLARING CONTEST WINNER : THE USER WHO GETS THE MOST LIKES ON A SINGLE PHOTO WILL WIN THE CONTEST. IDENTIFYING THE WINNER OF THE CONTEST BY FINDING THAT USER WHO HAS GOT MOST LIKES ON A SINGLE PHOTO.

Output/Result Table :

user_id	username	photo_id	total
52	Zack_Kemmer93	145	48

The winner of the contest is Zack_Kemmer93 having user_id 52. Because his photo with photo_id 145 has gotten the highest likes.
Thus, providing these details to the team. So that they can declare the winner of the contest.

Marketing

HASHTAG RESEARCHING : A PARTNER BRAND WANTS TO KNOW, WHICH HASHTAGS TO USE IN THE POST TO REACH THE MOST PEOPLE ON THE PLATFORM. THUS, FINDING THE TOP 5 MOST COMMONLY USED HASHTAGS ON THE PLATFORM.

- Steps for Hashtag Researching.(i.e. finding top 5 most used hashtags on Instagram.) :

  Step 1) I have selected 'tag_id' column from 'photo_tags' table & 'tag_name' column from 'tags' table.

  Step 2) Then I used COUNT function to count the total number of tags in the table.

  Step 3) Then I have performed INNER JOIN on the both tables as "photo_tags.tag_id = tags.id".

  Step 4) I used GROUP BY command to group the tags in the table.

  Step 5)Then ordered those tags by TOTAL in descending order. Then using LIMIT command I got the top 5 most used hashtags.

Query/Program :

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

Marketing

HASHTAG RESEARCHING : A PARTNER BRAND WANTS TO KNOW, WHICH HASHTAGS TO USE IN THE POST TO REACH THE MOST PEOPLE ON THE PLATFORM. THUS, FINDING THE TOP 5 MOST COMMONLY USED HASHTAGS ON THE PLATFORM.

Output/Result Table :

tag_id	tag_name	total
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

Marketing

LAUNCH AD CAMPAIGN: THE TEAM WANTS TO KNOW, WHICH DAY WOULD BE THE BEST DAY TO LAUNCH ADS. FINDING THE DAY OF THE WEEK WHEN MOST USERS REGISTER. TO PROVIDE INSIGHTS ON WHEN TO SCHEDULE AN AD CAMPAIGN.

- Steps for finding the best day to Launch AD Campaign :

  Step 1) I have used DAYNAME function on column ‘created_at’ to create a new column ‘day_of_week’ which has extracted the day on which the account is created from the original datestamp.

  Step 2) Then I used GROUP BY command along with count to find the total number of accounts created on each day.

  Step 3) Then by using the ORDER BY method I have sorted those seven days according to their total.

Query/Program :

```
SELECT DAYNAME(created_at) AS day_of_week,
    COUNT(*) AS total
FROM users
GROUP BY day_of_week
ORDER BY total DESC;
```

Marketing

Output/Result Table :

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

- ❖ Thus, the best day of the week to launch ads is Thursday or Sunday.
- ❖ If we future want to know out of those two days. It will be better to launch a campaign on a Sunday.
- ❖ Since, Sunday is a holiday and most users will be active at that time.

INVESTOR METRICS

1. User Engagement
2. Bots and Fake Accounts

Investor Metrics

USER ENGAGEMENT: TO FIND DO USERS ARE STILL ACTIVE AND POSTING ON INSTAGRAM. FINDING HOW MANY TIMES DOES AVERAGE USER POSTS ON INSTAGRAM. ALSO, PROVIDE THE TOTAL NUMBER OF PHOTOS ON INSTAGRAM/TOTAL NUMBER OF USERS.

- Steps for finding the User Engagement (i.e. finding how many times does average user posts on Instagram.) :

  Step 1) I have selected the total number of rows in the photos table and the number of rows in the user's table. Using the COUTN(*) command.

  Step 2) Then divides both counts and stored it as "average_photos".

Query (to find average number of posts) :

```
SELECT
   (SELECT COUNT(*) FROM photos) /
(SELECT COUNT(*) FROM users)
   AS average_photoes;
```

Output/Result Table :

average_photoes
2.5700

Investor Metrics

USER ENGAGEMENT: TO FIND DO USERS ARE STILL ACTIVE AND POSTING ON INSTAGRAM. FINDING HOW MANY TIMES DOES AVERAGE USER POSTS ON INSTAGRAM. ALSO, PROVIDE THE TOTAL NUMBER OF PHOTOS ON INSTAGRAM/TOTAL NUMBER OF USERS.

- Steps for finding the total posts done by each user.:

  Step 1) I have used ROW_NUMBER() function to arrange the final result in descending order according to the user post count.

  Step 2) Then in FROM statement I created a table using the nested SELECT command which has two columns ‘user_id’ and ‘user_post_count’.. Having alia as ‘user_post_count’.

  Step 3) Then lastly used the ORDER BY command to order the final output according to index_column. Which is indexed in the descending order of the user_post_count column so that we can find helpful insight about User Engagement.

Query (to find average number of posts) :

```
SELECT
   ROW_NUMBER() OVER (ORDER BY
user_post_count DESC) AS index_column,
   user_id, user_post_count
FROM (
   SELECT user_id, COUNT(*) AS user_post_count
   FROM photos
   GROUP BY user_id
) AS user_post_counts
ORDER BY index_column;
```

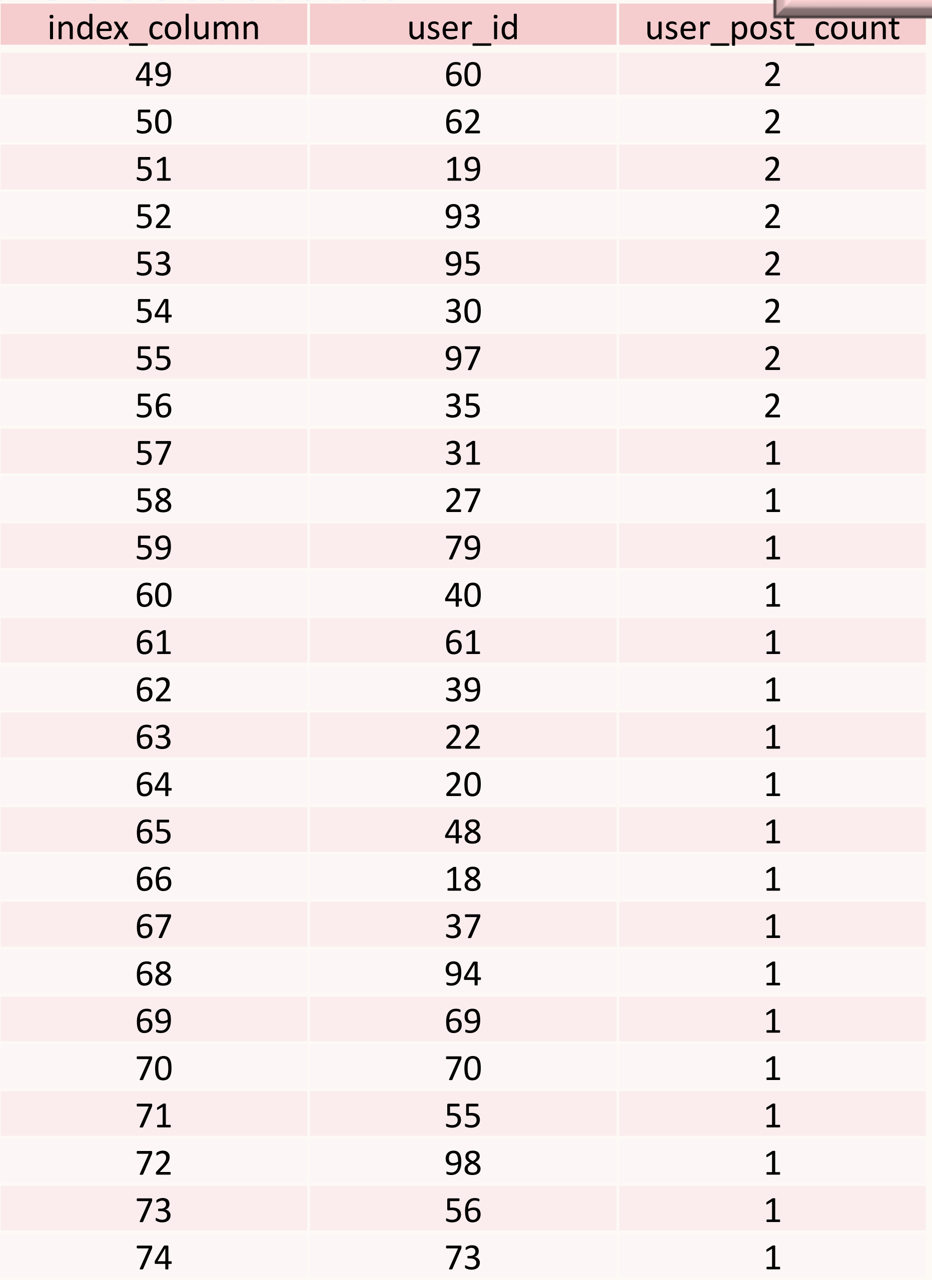
Investor Metrics

Output/Result Table :

index_column	user_id	user_post_count
1	23	12
2	88	11
3	59	10
4	86	9
5	58	8
6	29	8
7	77	6
8	33	5
9	52	5
10	47	5
11	6	5
12	13	5
13	51	5
14	78	5
15	11	5
16	1	5
17	72	5
18	65	5
19	43	5
20	64	5
21	26	5
22	2	4
23	3	4
24	9	4
25	8	4
26	46	4
27	12	4
28	44	4
29	63	4
30	32	4
31	28	4
32	16	4
33	15	4
34	87	4
35	4	3
36	10	3
37	50	3
38	67	3
39	17	3
40	42	3
41	92	3
42	96	3
43	99	3
44	38	2
45	100	2
46	82	2
47	84	2
48	85	2

Investor Metrics

Output/Result Table :



index_column	user_id	user_post_count
49	60	2
50	62	2
51	19	2
52	93	2
53	95	2
54	30	2
55	97	2
56	35	2
57	31	1
58	27	1
59	79	1
60	40	1
61	61	1
62	39	1
63	22	1
64	20	1
65	48	1
66	18	1
67	37	1
68	94	1
69	69	1
70	70	1
71	55	1
72	98	1
73	56	1
74	73	1

- From, The above result we can see that Out of 100 users, 74 users have posted photos on Instagram.
- Further, Out of those 74 users 43 users have posted above average posts on Instagram.
- This shows that users are still active and posting on Instagram.
- And after we encourage inactive users to start posting by sending them promotional emails, User Engagements are going to increase further.

Investor Metrics

BOTS & FAKE ACCOUNTS: THE INVESTORS WANT TO KNOW IF THE PLATFORM IS CROWDED WITH FAKE AND DUMMY ACCOUNTS. BY PROVIDING DATA ON THE USERS WHO HAVE LIKED EVERY SINGLE PHOTO ON THE SITE. (BECAUSE NORMAL USERS DON’T LIKE EVERY PHOTO)

- Steps for finding the inactive users (i.e. users who have never posted a single photo on Instagram.) :

  Step 1) I have selected id & username columns from users column.

  Step 2) Then I derived a table of total number of likes by each users from a Subquery as ‘total likes per user’.

  Step 3) Then after that column is derived I have applied HAVING command to filter those rows to get only those rows which have maximum number of likes. (i.e. likes all photos.)

Query (to find users who have liked all posts) :

```
SELECT
  id, username,
  (SELECT COUNT(*) FROM likes WHERE user_id
= users.id) AS total_likes_per_user
FROM users
HAVING
  total_likes_per_user = (SELECT COUNT(*) FROM
photos);
```

Investor Metrics

BOTS & FAKE ACCOUNTS: THE INVESTORS WANT TO KNOW IF THE PLATFORM IS CROWDED WITH FAKE AND DUMMY ACCOUNTS. BY PROVIDING DATA ON THE USERS WHO HAVE LIKED EVERY SINGLE PHOTO ON THE SITE. (BECAUSE NORMAL USERS DON'T LIKE EVERY PHOTO)

Output/Result Table :

id	username	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

- So, there are total of 13 accounts which may be Bots & Fake Accounts.
- Since, normal users can't like all the posts on Instagram.

CONCLUSION

CONCLUSION

- Thus, I have explored provided Instagram data and performed user analytics.
- Given all the required insights into the Marketing and Investees Matrices.
- I have learned to handle the database on the MySQL Workbench.
- I have learned to gain insights by using Queries.
- All the respective queries and their output is attached to this report.
- GitHub Repository and drive links are given as follows.

GitHub Repository:- https://github.com/ShindeYash/Instagram_User_Analytics.git

Drive Link:- https://drive.google.com/drive/folders/14QqMxwuKXOVNEjOHJ4mtFlWPdXSox71r?usp=drive_link

THANK YOU

Yash Shinde
yashpradeepshinde@gmail.com