INSTAGRAM USER ANALYTICS

PREPARED BY
PANKAJ KUMAR
DATA ANALYST
TRAINEE @ TRAINITY

PROJECT DESCRIPTION

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) 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 engagement and improve the experience altogether while helping the business grow.

TYPES OF ANALYTICS

Marketing Analytics

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

Investor Metrics

1. User Engagement
2. Bots & Fake Accounts

MARKETING ANALYTICS

REWARDING MOST LOYAL USERS

TASK

People who have been using the platform for the longest time.Find the 5 oldest users of the Instagram from the database provided .

QUERY

```
SELECT  TOP 5 * FROM users
ORDER BY created_at
```

RESULT

	id	username	created_at
1	80	Darby_Herzog	2016-05-06 00:14:21.190
2	67	Emilio_Bernier52	2016-05-06 13:04:29.960
3	63	Elenor88	2016-05-08 01:30:40.677
4	95	Nicole71	2016-05-09 17:30:22.370
5	38	Jordyn.Jacobson2	2016-05-14 07:56:25.837

These are the five users who have been using Instagram for the longest period, and as a result, they deserve to be rewarded.

MARKETING ANALYTICS

APPROACH

This query is used to find the 5 oldest users of Instagram from the provided database.

We start by selecting the top 5 records from the "users" table. This is done using the "SELECT TOP 5" part of the query. It means we only want to see the first 5 users in our results.

Then, we want to order these records based on the "created_at" column, which represents the date when each user created their Instagram account. We use the "ORDER BY created_at" part to sort the records in ascending order. This means the users with the earliest account creation dates will appear first.

So, by executing this query, we can get the information of the 5 oldest users of Instagram from the database, as we are selecting the top 5 records sorted by their account creation dates.

MARKETING ANALYTICS

REMIND INACTIVE USERS TO START POSTING

TASK

Q. By sending them promotional emails to post their 1st photo.
Find the users who have never posted a single photo on Instagram

QUERY

```
SELECT id,username FROM users WHERE id
NOT IN (SELECT user_id FROM photos)
```

RESULT

These are the users who have never posted on Instagram. Let's remind these inactive users to start posting and engage with the platform.

	id	username
1	5	Aniya_Hackett
2	7	Kasandra_Homenick
3	14	Jaclyn81
4	21	Rocio33
5	24	Maxwell.Halvorson
6	25	Tierra.Trantow
7	34	Pearl7
8	36	Ollie_Ledner37
9	41	Mckenna17
10	45	David.Osinski47
11	49	Morgan.Kassulke
12	53	Linnea59
13	54	Duane60
14	57	Julien_Schmidt
15	66	Mike.Auer39
16	68	Franco_Keebler64
17	71	Nia_Haag
18	74	Hulda.Macejkovic
19	75	Leslie67
20	76	Janelle.Nikolaus81
21	80	Darby_Herzog
22	81	Esther.Zulauf61
23	83	Bartholome.Bernh...
24	89	Jessyca_West
25	90	Esmeralda.Mraz57
26	91	Bethany20

MARKETING ANALYTICS

APPROACH

This query helps us find the users who have never posted a single photo on Instagram.

We start by selecting the "id" and "username" columns from the "users" table. This is done using the "SELECT id, username" part of the query.

Then, we use the "WHERE" clause to filter the results. We check if the "id" of each user is not present in the subquery result.

The subquery "(SELECT user_id FROM photos)" retrieves the "user_id" values from the "photos" table, which represents users who have posted photos on Instagram.

By using the "NOT IN" operator, we exclude the users whose "id" is found in the subquery result. This means we only get the users who are not included in the list of users who have posted photos.

MARKETING ANALYTICS

DECLARING CONTEST WINNER

TASK

Q. 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. Identify the winner of the contest and provide their details to the team

```
WITH liked_counts_table AS
(SELECT l.user_id AS liked_by_id,l.photo_id,p.user_id AS liked_to_id,u.username,
COUNT(*) OVER (PARTITION BY photo_id ORDER BY photo_id)AS liked_counts
FROM likes l
INNER JOIN photos p
ON l.photo_id = p.id
INNER JOIN users u
ON p.user_id = u.id)

SELECT DISTINCT(username),liked_counts FROM liked_counts_table
WHERE liked_counts = (SELECT MAX(liked_counts) FROM liked_counts_table)
```

QUERY

MARKETING ANALYTICS

DECLARING CONTEST WINNER

RESULT

	username	liked_counts
1	Zack_Kemmer93	48

The winner of the contest, with the highest number of likes, is Zack_Kemmer93, who received a total of 48 likes.

APPROACH

1. Common Table Expression (CTE):The query begins with a CTE named "liked_counts_table." It allows us to create a temporary table that holds the intermediate results of our query for easier manipulation.
2. Subquery inside the CTE:The subquery inside the CTE joins three tables: "likes," "photos," and "users." It retrieves the following information:
   - "liked_by_id": The user ID of the person who liked the photo.
   - "photo_id": The ID of the photo.
   - "liked_to_id": The user ID of the person whose photo was liked.
   - "username": The username of the user whose photo was liked.
   - "liked_counts": The count of likes for each photo.

MARKETING ANALYTICS

DECLARING CONTEST WINNER

3. The "COUNT(*) OVER (PARTITION BY photo_id ORDER BY photo_id)" part calculates the number of likes for each photo by partitioning the data based on the photo ID and ordering it by the photo ID itself.
4. Main Query:The main query selects distinct usernames and their corresponding liked_counts from the "liked_counts_table." It only includes rows where the "liked_counts" value matches the maximum number of likes found in the "liked_counts_table."

MARKETING ANALYTICS

HASHTAG RESEARCHING

TASK

Q. A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform. Identify and suggest the top 5 most commonly used hashtags on the platform

QUERY

```
SELECT TOP 5 pt.tag_id,tg.tag_name,COUNT(*) AS tag_count
FROM photo_tags pt
INNER JOIN tags tg
ON pt.tag_id = tg.id
GROUP BY tag_id,tag_name
ORDER BY tag_count DESC
```

MARKETING ANALYTICS

HASHTAG RESEARCHING

RESULT

The result displays the top five most frequently used hashtags. Among them, the hashtag "smile" has been used 59 times, while the hashtag "lol" has been used 24 times.

	tag_id	tag_name	tag_count
1	21	smile	59
2	20	beach	42
3	17	party	39
4	13	fun	38
5	11	lol	24

APPROACH

1. We have two tables, "photo_tags" and "tags." The "photo_tags" table stores the relationship between photos and hashtags, while the "tags" table contains information about the hashtags themselves.
2. We want to find the most commonly used hashtags, so we need to count how many times each hashtag has been used. We achieve this by using the "COUNT(*)" function and grouping the results based on the "tag_id" and "tag_name" columns.
3. The "INNER JOIN" clause is used to combine the "photo_tags" and "tags" tables based on the "tag_id" column, ensuring that we get the relevant hashtag information.

MARKETING ANALYTICS

LAUNCH AD CAMPAIGN

TASK

Q. The team wants to know, which day would be the best day to launch ADs.What day of the week do most users register on Provide insights on when to schedule an ad campaign

QUERY

```
SELECT DATENAME(WEEKDAY, created_at) AS registration_day,COUNT(*) AS
num_of_registration
FROM users
GROUP BY DATENAME(WEEKDAY, created_at)
ORDER BY num_of_registration DESC
```

MARKETING ANALYTICS

LAUNCH AD CAMPAIGN

RESULT

Most users tend to register on Sundays and Thursdays. Therefore, the best days to launch an ad campaign would be Sunday and Thursday.

	registration_day	num_of_registration
1	Sunday	16
2	Thursday	16
3	Friday	15
4	Monday	14
5	Tuesday	14
6	Wednesday	13
7	Saturday	12

APPROACH

1. We have a table called "users" that contains information about registered users.
2. We want to determine the day of the week when most users register. To do this, we use the DATENAME(WEEKDAY, created_at) function, which retrieves the name of the weekday from the "created_at" column.
3. Next, we count the number of registrations for each weekday using the COUNT(*) function and group the results based on the weekday.
4. Finally, we sort the results in descending order based on the number of registrations using the ORDER BY clause.

INVESTOR METRICS

USER ENGAGEMENT

TASK

Q. Are users still as active and post on Instagram or they are making fewer posts. (A) Provide how many times does average user posts on Instagram. (B) Also, provide the total number of photos on Instagram/total number of users

(A) QUERY

```
SELECT registered_user_id,SUM(post_flag) AS num_of_post FROM(
SELECT u.id AS registered_user_id,p.user_id AS post_user_id,
   CASE WHEN u.id = p.user_id THEN 1 ELSE 0
   END AS post_flag
FROM photos P
RIGHT JOIN users u
ON p.user_id = u.id)x
GROUP BY registered_user_id
```

INVESTOR METRICS

USER ENGAGEMENT

RESULT

The following result displays the number of posts made by each user on Instagram.

	registered_user_id	num_of_post
1	1	5
2	2	4
3	3	4
4	4	3
5	5	0
6	6	5
7	7	0
8	8	4
9	9	4
10	10	3
11	11	5
12	12	4
13	13	5
14	14	0
15	15	4
16	16	4
17	17	3
18	18	1
19	19	2
20	20	1
21	21	0
22	22	1
23	23	12

	registered_user_id	num_of_post
23	23	12
24	24	0
25	25	0
26	26	5
27	27	1
28	28	4
29	29	8
30	30	2
31	31	1
32	32	4
33	33	5
34	34	0
35	35	2
36	36	0
37	37	1
38	38	2
39	39	1
40	40	1
41	41	0
42	42	3
43	43	5
44	44	4
45	45	0
46	46	4
47	47	5
48	48	1

	registered_user_id	num_of_post
49	49	0
50	50	3
51	51	5
52	52	5
53	53	0
54	54	0
55	55	1
56	56	1
57	57	0
58	58	8
59	59	10
60	60	2
61	61	1
62	62	2
63	63	4
64	64	5
65	65	5
66	66	0
67	67	3
68	68	0
69	69	1
70	70	1
71	71	0
72	72	5
73	73	1
74	74	0

	registered_user_id	num_of_post
75	75	0
76	76	0
77	77	6
78	78	5
79	79	1
80	80	0
81	81	0
82	82	2
83	83	0
84	84	2
85	85	2
86	86	9
87	87	4
88	88	11
89	89	0
90	90	0
91	91	0
92	92	3
93	93	2
94	94	1
95	95	2
96	96	3
97	97	2
98	98	1
99	99	3
100	100	2

MARKETING ANALYTICS

USER ENGAGEMENT

APPROACH

1. We have two tables, "photos" and "users," which contain information about posts and users on Instagram, respectively.
2. The query combines the "photos" and "users" tables using a RIGHT JOIN to ensure we include all users, even if they haven't made any posts.
3. We assign a "post_flag" of 1 for each user who has made a post, and 0 for users who haven't made any posts.
4. The inner query (subquery) creates a temporary table "x" with the user IDs and their corresponding post flags.
5. Finally, we group the results by the registered user ID and calculate the sum of the post flags for each user.

INVESTOR METRICS

USER ENGAGEMENT

(B) QUERY

```
SELECT SUM(post_flag)/MAX(registered_user_id) AS
total_photos_to_number_of_users_ratio
FROM(
SELECT u.id AS registered_user_id,p.user_id AS post_user_id,p.image_url,
   CASE WHEN u.id = p.user_id THEN 1 ELSE 0
   END AS post_flag
FROM photos P
RIGHT JOIN users u
ON p.user_id = u.id)x
```

RESULT

	total_photos_to_number_of_users_ratio
1	2

The above result displays the total number of photos on Instagram divided by the total number of users.

MARKETING ANALYTICS

USER ENGAGEMENT

APPROACH

1. We have two tables, "photos" and "users," which contain information about posts and users on Instagram, respectively.
2. The query combines the "photos" and "users" tables using a RIGHT JOIN to ensure we include all users, even if they haven't made any posts.
3. We assign a "post_flag" of 1 for each user who has made a post, and 0 for users who haven't made any posts.
4. The inner query (subquery) creates a temporary table "x" with the user IDs, their corresponding post flags, and other relevant information.
5. Finally, we calculate the sum of the post flags (representing the total number of posts) and divide it by the maximum registered user ID to get the ratio of total photos to the number of users.

MARKETING ANALYTICS

BOTS & FAKE ACCOUNTS

TASK

Q. The investors want to know if the platform is crowded with fake and dummy accounts .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

```
SELECT l.user_id, u.username
FROM likes l
JOIN users u ON l.user_id = u.id
GROUP BY l.user_id, u.username
HAVING COUNT(DISTINCT l.photo_id) = (SELECT COUNT(*) FROM photos)
```

INVESTOR METRICS

BOTS & FAKE ACCOUNTS

RESULT

	user_id	username
1	5	Aniya_Hackett
2	14	Jaclyn81
3	21	Rocio33
4	24	Maxwell.Halvorson
5	36	Ollie_Ledner37
6	41	Mckenna17
7	54	Duane60
8	57	Julien_Schmidt
9	66	Mike.Auer39
10	71	Nia_Haag
11	75	Leslie67
12	76	Janelle.Nikolaus...
13	91	Bethany20

These accounts have liked every single photo on the site, indicating a strong possibility that they are bots or potentially bots.

MARKETING ANALYTICS

BOTS & FAKE ACCOUNTS

APPROACH

1. We have two tables, "likes" and "users," which contain information about likes and user accounts on the platform, respectively.
2. The query joins the "likes" and "users" tables based on the user ID.
3. We group the results by the user ID and username to aggregate the data for each user.
4. In the HAVING clause, we filter the results to only include users who have liked every single photo on the site. We do this by comparing the count of distinct photo IDs liked by each user with the total count of photos on the site.

CONCLUSIONS

In conclusion, by leveraging SQL server and analyzing user data, we have provided valuable insights to address the questions raised by the management team of Instagram. Through our analysis, we have identified the following key findings:

A) Marketing Insights:

- Most Loyal Users: We have identified the 5 oldest users of Instagram, who have been using the platform for the longest time. These users can be rewarded for their loyalty and engagement.
- Inactive Users: We have identified users who have never posted a single photo on Instagram. By sending them promotional emails, we can encourage them to start posting and increase user engagement.
- Contest Winner: Based on the contest criteria of most likes on a single photo, we have identified the winner and provided their details to the team for further actions.
- Hashtag Research: We have identified the top 5 most commonly used hashtags on the platform, which can be helpful for partner brands in reaching a larger audience.
- AD Campaign Scheduling: By analyzing the registration patterns of users, we have identified the best days to launch AD campaigns, namely Sundays and Thursdays.

CONCLUSIONS

B) Investor Metrics:

- User Engagement: We have determined the average number of posts made by users on Instagram, providing insights into user activity and engagement. Additionally, we have calculated the ratio of total photos on Instagram to the total number of users, which helps assess overall user engagement.
- Bots & Fake Accounts: By identifying users who have liked every single photo on the site, we have identified potential bot accounts that may indicate the presence of fake and dummy accounts on the platform.

THANK YOU