Instagram User Analytics

Project Description:

In this project, we use our SQL knowledge to find Instagram user analytics. User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) to derive business insights for marketing, product & development teams.

Here are the tasks we need to perform for the different teams:

- **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.

    Our Task: Find the 5 oldest users of Instagram from the database provided.
  - ***Remind Inactive Users to Start Posting:*** By sending them promotional emails to post their 1st photo.

    Our 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.

    Our 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.

    Our Task: Identify and suggest the top 5 most used hashtags on the platform.

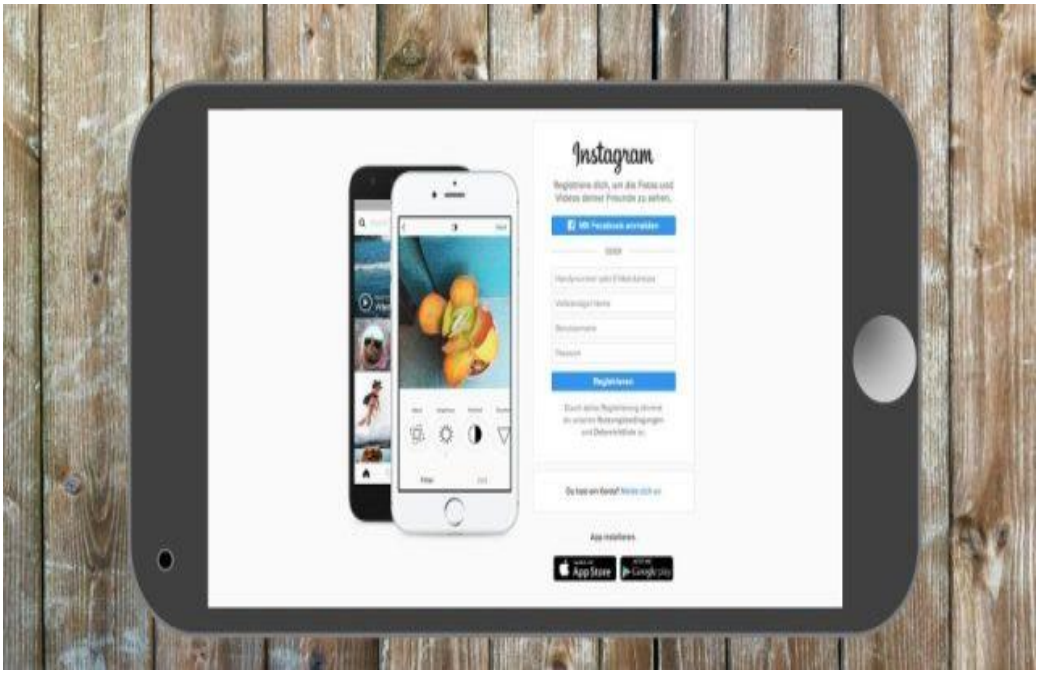
- ***Launch AD Campaign:*** The team wants to know, which day would be the best day to launch ADs.

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

- **Investor Metrics:** Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds.
  - ***User Engagement:*** Are users still as active and post on Instagram or they are making fewer posts?

    Our Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users.
  - ***Bots & Fake Accounts:*** The investors want to know if the platform is crowded with fake and dummy accounts.

    Our 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).

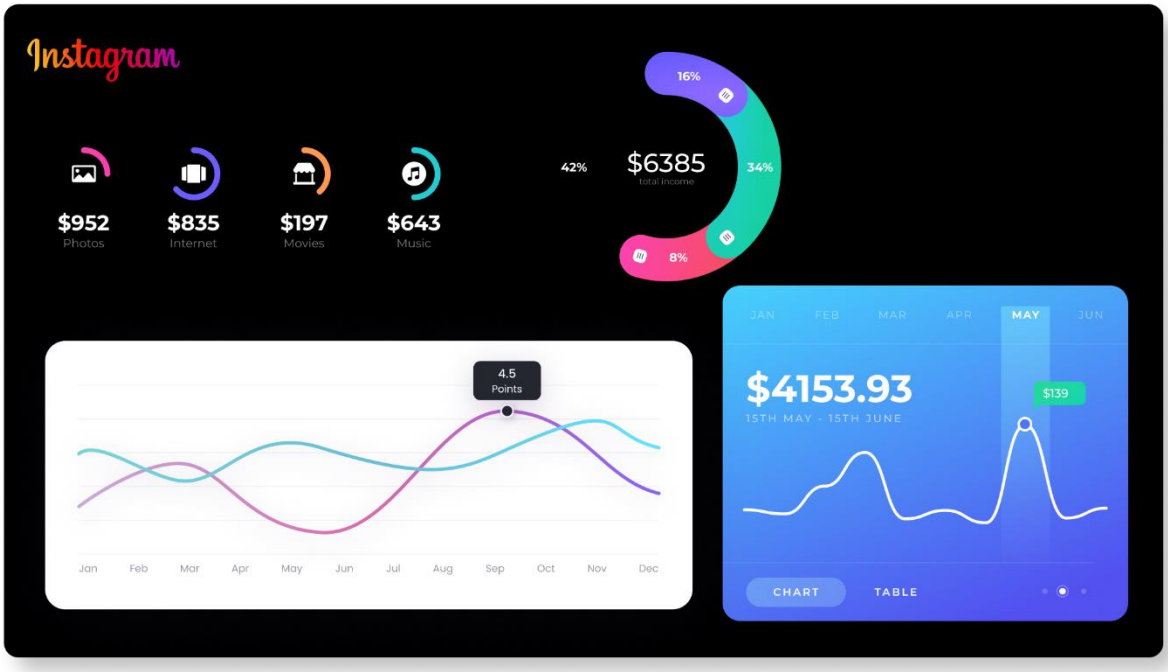


Approach:

We are provided with the data set. We assume the data to be clean. We explored the schema of all the provided tables in the MySQL workbench and understood how the tables relate to each other. The database was cautiously converted into SQL tables to ensure most benefit from the database provided. Then we used the database and executed some of the SQL queries of our knowledge to obtain the required solution.

Tech-Stack Used:

The software used for this project is ***MySQL Workbench 8.0 CE.***



Insights:

Rewarding Most Loyal Users: To know about the most loyal users we need to find the users whose accounts were create very long time ago.

Here is the following query we used to find the most loyal users:

```
SELECT
    username, created_at
FROM
    users
ORDER BY created_at
LIMIT 5;
```

This query resulted in the following 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

These are the 5 most loyal users of Instagram

Reminding Inactive users to start posting: The inactive users are the ones who do not post on their account in this context the photos.created_at is null.

Here is the query we used to find the inactive users:

```
SELECT
    users.id, users.username
FROM
```

```
    users
        LEFT JOIN
    photos ON users.id = photos.user_id
WHERE
    created_dat IS NULL;
```

This query resulted in the following output:

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

The above are the users who had not posted a single photo till now

Declaring Contest Winner: To know about the contest winner we need to count the no of likes on each photo of users and find the highest liked photo.

Here is the following query we used to find the highest liked photo:

```
SELECT
    photos.user_id,
    users.username,
    likes.photo_id,
    COUNT(likes.user_id) AS number_of_Likes
FROM
    likes
        JOIN
    photos
        JOIN
    users ON users.id = photos.user_id
        AND photos.id = likes.photo_id
GROUP BY likes.photo_id
ORDER BY number_of_Likes DESC
LIMIT 1;
```

This query resulted in the following output:

user_id	username	photo_id	number_of_Likes
52	Zack_Kemmer93	145	48

Hashtag Researching:

To find about the top hashtags used in the posts we need to find the no of times the hashtag is used in the photos

Here is the query to find the highest used hashtag that is used in the photos:

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

This query resulted in the following output:

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

Launch AD Campaign:

To launch a ad campaign we need to convey the message to most no of audience, so we need to find the day when most candidates register on ?

Here is the query to find the day when the most no of users register:

```
SELECT
    DAYOFWEEK(created_at) AS day_of_the_week,
    COUNT(users.id) AS number_of_Accounts_Registered
FROM
    users
GROUP BY day_of_the_week
ORDER BY number_of_Accounts_Registered DESC;
```

The query resulted in the following output:

day_of_the_week	number_of_Accounts_Registered
5	16
1	16
6	15
3	14
2	14
4	13
7	12

User Engagement: To provide how many times does average user posts on Instagram and the total number of photos on Instagram/total number of users. To do that we need the count of total photos on the Instagram and the total no of users.

Here is the query :

```
SELECT
    COUNT(DISTINCT users.id) as 'total users',
    COUNT(photos.id) as 'total no of photos',
    ROUND((COUNT(photos.id) / COUNT(DISTINCT users.id)), 2) as
'total_photos/total_users'
FROM
    users
        LEFT JOIN
    photos ON users.id = photos.user_id;
```

The query resulted in the following output:

total users	total no of photos	total_photos/total_users
100	257	2.57

We can also use data to find the avg of posts by users who have posted on their timeline:

```
SELECT
    COUNT(distinct users.id) AS 'no of users who post',
    COUNT(photos.id) AS 'total no of post',
    COUNT(photos.id)/count(distinct users.id) AS 'no of times avg users post'
FROM
    users
        JOIN
```

```
    photos ON users.id = photos.user_id;
```

The query resulted in the following output:

no of users who post	total no of post	no of times avg users post
74	257	3.4730

Bots and Fake accounts:

To 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).

Here is the query for our task:

```
SELECT
    users.username,
    likes.user_id,
    COUNT(likes.created_at) AS “number of photos liked”
FROM
    likes
        JOIN
    users ON users.id = likes.user_id
GROUP BY likes.user_id
HAVING COUNT(likes.created_at)  =  257;
```

The query resulted in the following output:

username	user_id	number of photos liked
Aniya_Hackett	5	257
Jadyn81	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

These usernames are that of

Result: By completing the project, I am feeling more confident in my SQL knowledge. It really helped me to brush up on my concepts related to Sub-queries and Aggregate functions. It also helped me to understand the table schema and how normalization can better help to understand the dataset.

BY

THOTAMCHETTY RANJITH

THANKYOU