Instagram user analytics

Report by - Siva Sankari H

1.Introduction

The aim of the project is to analyse Instagram – a popular free social media platform which allows its users to share photos, videos and also interact with others' content. The main objective here, is to analyse the interaction of user and their engagements to various contents posted in the platform. The valuable insights gained from this analysis is then used to derive meaningful decisions that can help business grow.

1.1 Project Description

Instagram is an American photo and video sharing social networking service owned by Meta Platforms. It allows its users to upload media that can be edited with filters, be organised by hashtags, and be associated with a location via geographical tagging. Posts can be shared publicly or with preapproved followers. The users can interact to the content shared by liking it, commenting on it and/or save it. They can also browse different contents with the help of hashtags and locations, and also view trending content. It allows a follow and following feature – that allows the users to add their content to followers' personal feed.

The main of this project will be to analyse the different users using Instagram. Instagram allows a business account feature along with the private and public account setting. Business account is a professional account designed for businesses and brands to use on Instagram. These are good options for local businesses, retailers, brands, organisations, and service providers.

As of January 2024, India has the largest number of Instagram users in the world with 36.4 crore users. It is to be noted that an average Indian spends about 141 mins/per day in the application while on a global level it is around 145 mins.

Analysing Instagram user data is important because it helps us in understanding the audiences that are using the platform, identify trends and their success and failure rate. The insights gained from this project can actually be used by local business holders to other important business players in advancing the product design, user experience, marketing and advertising etc.

2. Database Design

2.1 Entity Relationship Diagram (ER) –

An ER diagram is a conceptual (graphical) representation of the database – it shows the structure and the relationships in the DB using symbols and lines. It helps in organising and understanding data and plays a crucial role in the database design. ER diagram uses Entities,

Attributes and Relationships to describe a DB. An Entity may be an object with a physical existence (a person, a car, house etc.) or conceptual (company, organisation). Attributes define the entity type. the following diagram shows the ER diagram for clone Instagram DB.

2.2 Database Schema –

Schema is a blueprint that defines the structure, organization and relationships within the database. It is a logical and visual representation of how data is stored and organised in the database.

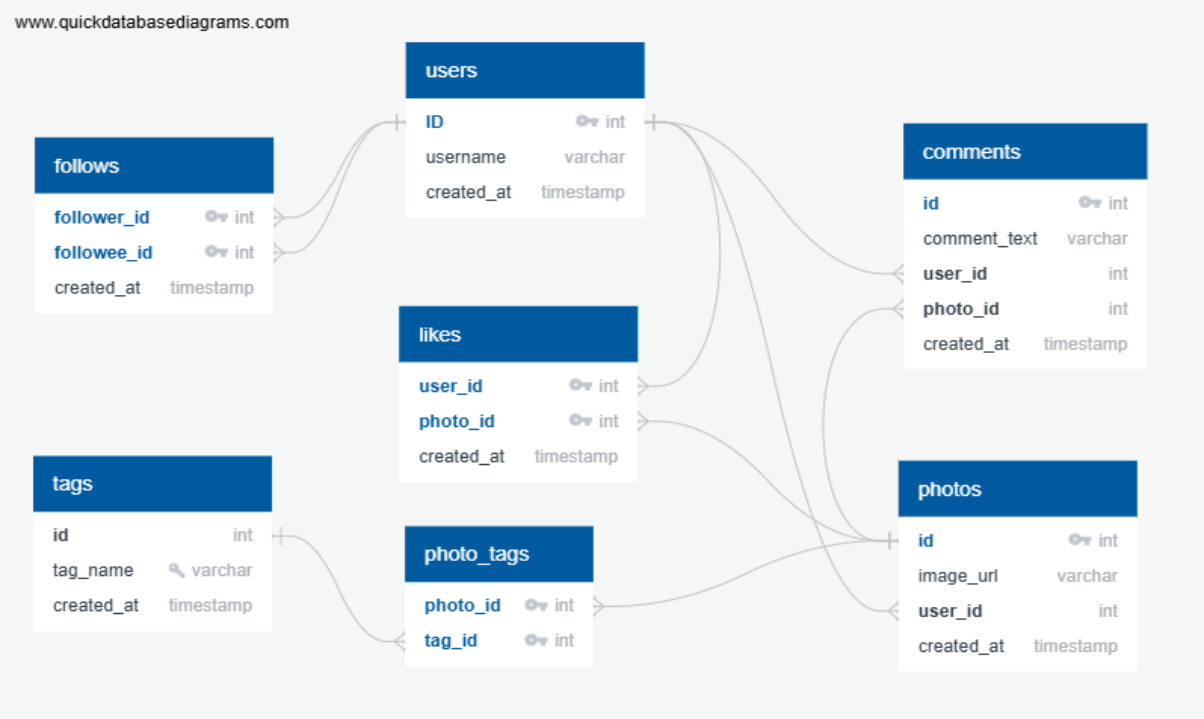

Figure 1- ER Diagram for a model Instagram DB (Image credits – QuickDBD)

The model database ig_clone created here, contains 7 tables that are related to each other. The table information are as follows –

Table 1 – Users: On installing the Instagram application, user will be prompted to create an account. The created account will be stored in the **users** table**.**

Field name	Purpose
ID – Primary Key	The unique id given to each account
username	unique display name given by the user for their account – acts an Instagram handle
Created_at	time at which user created the account – stored in timestamp format.

Table 2 – follows: Each user will be given a provision to follow another users' account by sending and accepting a follow request. They can view, comment, like and share this content to others' or save it.

Field name	Purpose
Follower_id - Primary Key	ID of the follower account.
Followee_id - Primary Key	ID of the followee account.
Created_at	time at which user got a follow request or sent by him.

Table 3 – tags: User can directly interact with the other users' by tagging them directly.

Field name	Purpose
ID	Each tag name will be assigned an id.
Tag_name - Primary Key	Name of the tag
Created_at	Time at which tag was used.

Table 4 – photos

Field name	Purpose
ID - Primary Key	Id given to the uplaoded photo.
Image_url	url of the uploaded image
User_id	Id of the account that posted the image
Created_at	Time at which photo was uplaoded.

Table 5 – comments

Field name	Purpose
ID - Primary Key	Comment id
User_id	Id the of commenter.
Photo_id	Id of the photo on which user commented
Comment_text	Message entered.
Created_at	Time at which comment received.

Table 6 – photo_tags

Field name	Purpose
Photo_id - Primary Key	Id of the photo tagged
Tag_id - Primary Key	Tag id.

Table 7 – likes

Field name	Purpose
User_id - Primary Key	The unique id given to each account
Photo_id - Primary Key	Id of the photo receiving like.
Created_at	Time at which a like was received.

Database has been created according to the sample dataset given along with the problem statement. The queries used is attached in the last for reference. The goal of the project is to analyse the market trends and perform investor metrics.

3. Problem Statement –

Analysing user interactions and engagement with Instagram app to provide valuable insights that can help business grow. User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business.

4. Approach and Analysis of the Result–

A) Marketing Analysis:

1. Loyal User Reward - The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
Task: Identify the five oldest users on Instagram from the provided database.

Approach –

The task here is to identify the oldest users or the initial users in the platform. We can fetch this by arranging all the accounts in the users table in ascending order.

A select query with an order by statement will achieve this. The limit constraint will fetch only the top most rows from the final result.

Query and Output –

```
11    # Loyal User Reward -
12 •  select * from users order by created_at limit 5;
13
```

Result Grid | Filter Rows: | Edit:

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

Output 1: Five oldest users on Instagram

Analysis – The output is sorted according to the *created_id field*. The oldest user was created at 06/05/2016 going by name Darby Herzog having user id = 80.

2. Inactive user engagement – Encourage inactive users to post by sending them provisional emails. **Task: Identify users who have never posted a single photo on Instagram.**

Approach –

Identify inactive user – Can be achieved by joining both the users and photos table through a left join. The photos table have id (photo_id) field as primary and foreign key. It is linked to the users table on *users.id = photos.id* constraint.

The LEFT JOIN command returns all rows from the left table i.e., users table, and the matching rows from the right table i.e., photos table. The result is NULL from the right side, if there is no match.

Query and output –

```
16 •  select u.id as users_id, username, count(distinct p.id) as photo_id from users u
17    left join photos p on p.user_id = u.id
18    where p.image_url is null
19    group by users_id;
20
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

users_id	username	photo_id
5	Aniya_Hackett	0
7	Kasandra_Homenick	0
14	Jaclyn81	0
21	Rocio33	0
24	Maxwell.Halvorson	0
25	Tierra.Trantow	0
34	Pearl7	0
36	Ollie_Ledner37	0
41	Mckenna17	0
45	David.Osinski47	0

Output 2: Accounts not posted a single photo

Analysis –

The output obtained shows a total of 26 entries having posted 0 images. We are counting the total number of images posted by each user to identify inactive account. 0 in the photo_id field indicates our target – inactive user.

3. Contest Winner Declaration –

Task: Determine the winner of the contest and provide their details to the team.

Approach –

The user with the most likes on a single photo win. We have got the photo_id with the greatest number of likes. the winner of the contest – user with maximum likes for a single photo. The query and output here show the first 5 potential winners.

Query and Output –

```
SELECT p.id, p.user_id, COUNT(l.user_id) AS total_likes, p.image_url FROM photos p
JOIN likes l ON p.id = l.photo_id
GROUP BY p.id, p.user_id ORDER BY total_likes DESC limit 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

id	user_id	total_likes	image_url
145	52	48	https://jarret.name
127	46	43	https://celestine.name
182	65	43	https://dorcas.biz
123	44	42	http://shannon.org
61	20	41	https://dejon.name

Output 3: Contest Winner Declaration Sheet

Analysis –

Winner contest is user_id = 52 who has posted image with id = 145. He has got a total like of 48. The image that got most likes is https://jarret.name.

4. Hashtag Research – know the most popular hashtags to use in their posts to reach the most people.

Task: Identify and suggest the top five most commonly used hashtags on the platform.

Approach –

We need to join table tags and photo_tags in order to find out the most commonly used hashtags. The tags table has tag_id which can be used to join with photo_tags table.

Query and Output-

```
select distinct pt.tag_id, t.tag_name, count(pt.photo_id) as no_of_uses
from tags t left join photo_tags pt on pt.tag_id = t.id
group by pt.tag_id, t.tag_name order by no_of_uses desc limit 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

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

Output 4: Most common hashtags being used

Analysis –

The output below shows the 5 most commonly used hashtags along with the tag_name. We can see that tag_id = 21 with name = smile has been used 59 times. This means 59 users has used this tag in the photos. The most commonly used tags are smile, beach, party, fun and concert.

5. Ad Campaign Launch: Identify the best day of the week to launch ads.
Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

Approach –

We need to determine the day when most of the users register in Instagram. This info is readily available in the users table. dayname(created_at) gives the day from the created_at timestamp.

Query and Output –

```
43 •  select dayname(created_at), count(id)  from users group by dayname(created_at) limit 10;
44
45
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

dayname(created_at)	count(id)
Thursday	16
Sunday	16
Tuesday	14
Saturday	12
Wednesday	13
Monday	14
Friday	15

Output 5: Most common days where users registered on Instagram

Analysis –

From the output above it can be seen that most of the users' register on Thursday's and Sunday's. The add campaigns launched on these days have probability of seeing success more.

B) Investor Metrics:

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

Approach – join users and photos table to compute the average. Use the expression count(photos.image_url) – to calulate the total number of posts by a single user. Dividing it by count(distinct user.id) will give the required average.

Query and output -

```
47 •  select users.id, count(photos.image_url)/count(distinct users.id) as avg_post_per_users
48    from users left join photos on users.id = photos.user_id
49    group by users.id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	avg_post_per_users
1	5.0000
2	4.0000
3	4.0000
4	3.0000
5	0.0000
6	5.0000
7	0.0000
8	4.0000
9	4.0000

Output 6: Average number of posts per user on Instagram

Analysis –

It can be seen that user with id=1 has posted 5 posts on average. We also have 26 users with 0 average posts indicating their inactive status.

We have on total 257 posts and 100 users on Instagram clone database.

```
52 •  select count(photos.image_url) as total_photos_Instagram, count(distinct users.id) as total_users_Instagram
53    from users left join photos on users.id = photos.user_id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

total_photos_Instagram	total_users_Instagram
257	100

Output 7: total users and posts on Instagram

2.Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.
Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Approach – Identify the bots. Users who has liked every single photos can be considered as bot account. The query below gives us a list of user id, username and no_of_likes given by that user. We get 77 records – which means 23 users have not liked a single post.

```
63 •  select users.id,users.username, count(likes.photo_id) as no_likes
64    from users join likes on users.id = likes.user_id
65    group by users.id, users.username;
66
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	username	no_likes
2	Andre_Purdy85	94
3	Harley_Lind18	79
4	Arely_Bogan63	93
5	Aniya_Hackett	257
6	Travon.Waters	82
8	Tabitha_Schamberger11	79
9	Gus93	85
10	Presley_McClure	87
11	Justina.Gaylord27	89

Output 8: number of likes each user has given

We can see that user with id = 2 has liked 94 posts/photos. Another user with id = 5 has liked 257 posts. We have total of 257 photos and user has liked all the photos (group by condition on users.id makes sure the count obtained is distinct). This may be potential bot behaviour.

The query below shows total number of such bot accounts. we have 13 such potential bot accounts.

```
67    #count of bot accounts
68 •  select users.id,users.username, count(likes.photo_id) as no_likes
69    from users join likes on users.id = likes.user_id
70    group by users.id, users.username
71    having no_likes = 257;
72
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	username	no_likes
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

Output 9: Bot account counts

5.Tech Stack Used

1) MySQL Workbench –
   a) We have used latest version v 8.0.40.
   b) MySQL Workbench is a powerful visual database design tool specifically designed for MySQL databases. It offers a user-friendly interface and a comprehensive suite of tools for database administration, development, and modelling.
   c) We have used for writing, execute and debug SQL queries. It's database design allows us to create and edit ER diagrams to visually model database structures.
2) SQL Server Management Studio –
   a) SQL SMS is used to interact with the SQL server database.
   b) The key features include – Query Editor – used for write, execute and debug queries, Database Engine – for connecting and managing the server instances. Data import and Export Wizard – for importing and exporting various sources like csv, excel and other databases.

6. Insights and Result

This section focuses on the insights gathered while answering the case study questions.

1) Loyal User Reward – from the output it can be seen that the earliest user on the platform dates back to 6th May,2016. Maximum users registered on Thursday's (16) and Sunday's (16) and lowest registration on Saturday (12).
2) Inactive Users Engagement – we have a total of 26 inactive users. These people have not posted even a single photo since their joining on the platform. Potential reasons could be – the user can be a private person who doesn't like to share things online, or maybe they are uninterested in the platform itself or maybe they can be facing some technical issue. It can be seen that all of these 26 users registered on the platform on year 2016 – 2017.
3) Ad Campaign Launch – Most of the users' have registered on Thursday and Sunday. As our DB is based on social media platform most of the users register on it on recommendation. Advantage of launching a campaign on Thursday – Can create anticipation for the weekend. Most ideal for audience wanting to experience work-to-weekend transition. Sundays' launch will capitalise impulse buys as the audiences are more relaxed.
4) User engagement – we can encourage inactive or less active users' by introducing contests and challenges. Contests like maintaining posts streak for 30 days etc,. We can introduce attractive prizes to incentivize the user-generated content. We can encourage interaction among the users by implementing new features. There may be some new or elderly users who will need some assistance in using the platform. Simplifying the platform interface, giving certain suggestions or prompts for posting or introducing a chat bot to resolve common issue raising in the platform may encourage more users to download and use the application.