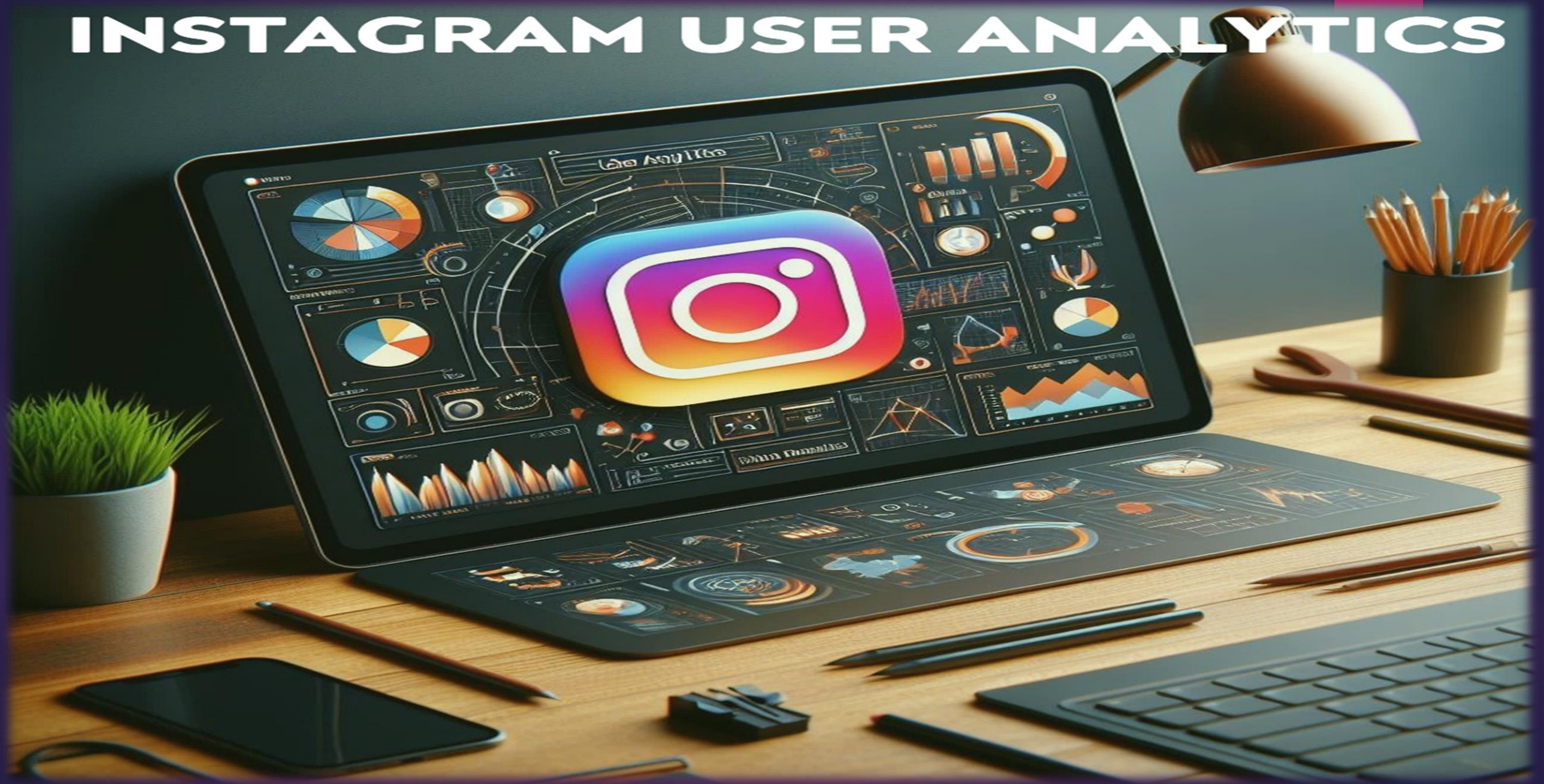
INSTAGRAM USER ANALYTICS



Created by Mainak Mukherjee | Email- subha.mainak@gmail.com

Project Description:-

- The project involves extracting and analyzing data from an Instagram database.
- Its aim is to understand Instagram user behavior and preferences.
- It seeks insights into user interests, demographics, and engagement patterns.
- The objective is to enhance content strategies, boost engagement, and optimize marketing efforts.
- The findings could influence future developments on one of the world's leading social media platforms.

-| Approach |-

Data Exploration: Exploratory Data Analysis (EDA) is used to better understand user demographics, post frequency and interaction trends.

SQL Queries and Analysis: Utilizing SQL queries, I extracted relevant subsets of data from the database to address specific analytical questions.

Content Analysis: Analyzed the content of popular posts, including active users, inactive users, best days for engagement, loyal users, and prominent hashtags/captions.

Insights Generation: Through iterative analysis and interpretation of query results, I derived actionable insights regarding user behavior, content performance, and audience engagement.

Validation and Iteration: Iterative improvements were made to the analysis methodology based on feedback and emerging trends in the Instagram ecosystem.

-| Tech-Stack Used |-

- For this project, I utilized the following software and versions:

  - **MySQL Workbench 8.0 CE**

    - MySQL Workbench was chosen as the primary tool for database development and management due to its user-friendly interface, robust query editor, and comprehensive features for database design and administration.

    - Version 8.0 CE was selected to leverage the latest enhancements, bug fixes, and performance optimizations available in the MySQL ecosystem.

-: Project Agenda :-

- **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.
2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
3. Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.
4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

- **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.
2. Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

Task 1:- Identify the five oldest users on Instagram.

Query-

```
select *, rank() over(order by created_at)
as Rank_No
from users limit 5;
```

Result-

id	username	created_at	Rank_No
80	Darby_Herzog	2016-05-06 00:14:21	1
67	Emilio_Bernier52	2016-05-06 13:04:30	2
63	Elenor88	2016-05-08 01:30:41	3
95	Nicole71	2016-05-09 17:30:22	4
38	Jordyn.Jacobson2	2016-05-14 07:56:26	5

- **Insights:-** The results clearly show that **Darby Herzog**, **Emilio Bernier**, **Elenor**, **Nicole** and **Jordyn Jacobson** are the oldest Instagram users, and so are eligible for the rewards.

Task 2:- Identify users who have never posted a single photo on Instagram.

Query-

```
select users.id, users.username from users
left join photos
on users.id = photos.user_id
where photos.id is null;
```

Result-

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jadyn81
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

- **Insights:-** The results clearly demonstrate that these **26 users** have never posted a single photo on Instagram. Hence, they might be classified as inactive or less active users.

Task 3:- Determine the winner of the contest and provide details.

Query-

```
select users.id as user_id, users.username, photos.id as photo_id, photos.image_url,
count(likes.user_id) as Total_Likes
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_Likes
desc limit 1;
```

Result-

user_id	username	photo_id	image_url	Total_Likes
52	Zack_Kemmer93	145	https://jarret.name	48

- **Insights:-** From the output of this query it is clearly evident that **Zack Kemmer** is the **winner** of the contest since he has 48 likes on his Instagram post.

Task 4:- Identify and suggest the top five most commonly used hashtags on the platform.

Query-

```
select tag_name, count(tag_name) as Total_used
from tags join photo_tags
on tags.id = photo_tags.tag_id
group by id order by Total_used
desc limit 5;
```

Result-

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

- **Insights:-** The result table indicates that top 5 most commonly used hashtags are- '#**smile**' '#**beach**' '#**party**' '#**fun**' & '#**concert**'

Task 5:- Determine the day of the week when most users register on Instagram.

Query-

```
select dayname(created_at) as Day,
count(*) as Total from users
group by Day order by Total desc limit 2;
```

Result-

	Day	Total
▶	Thursday	16
	Sunday	16

- **Insights:-** Based on the query output, **Thursday** and **Sunday** are the optimal days to run Ad Campaigns.

Task 6:- 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.

Query 1-

```
select user_id, avg(id) as avg_post from photos group by user_id;
```

Result 1-

user_id	avg_post
1	3.0000
2	7.5000
3	11.5000
4	15.0000
6	19.0000
8	23.5000
9	27.5000
10	31.0000

Result 32

Query 2-

```
select round
((select count(*) from photos) / (select count(*) from users),2)
as Average_post;
```

Result 2-

Average_post
2.57

- **Insights:-** On average, each user generates **2–3** posts on Instagram.

Task 7:- Identify users (potential bots) who have liked every single photo on the site.

Query-

```
select users.username,
count(*) as Total_likes
from users join likes on
users.id = likes.user_id
group by users.id
having Total_likes = (select count(*) from photos);
```

Result-

username	Total_likes
Aniya_Hackett	257
Jadyn81	257
Rocio33	257
Maxwell.Halvorson	257
Ollie_Ledner37	257
Mckenna17	257
Duane60	257
Julien_Schmidt	257
Mike.Auer39	257
Nia_Haag	257
Leslie67	257
Janelle.Nikolaus81	257
Bethany20	257

- **Insights:-** There are 13 fake accounts that like all Instagram posts, which might be categorized as potential bots.

THANK YOU