

PROJECT: Instagram User Analytics**PROJECT DESCRIPTION:**

This project is basically used to derive business insights for marketing, product & development teams. In this project, the Initial goal is in the vision of the Marketing Domain which will summarize the needs of statistical data required for the campaign. In this project, the tasks are subjected to deriving the required output based on user activities, initiating a contest & Declaration of Winner of the contest. Further on providing insightful data to business holders based on the data collected.

The Second Stage of this project is to evaluate the Security & consistency of the users and their account for presenting a report of User metrics to Investors. In this stage, the focus is on the vision of evaluating user Regularity & Detecting dummy or Bot accounts present on the Instagram platform.

APPROACH:

To work on a project consisting of Statistical data, initially, it is important to understand the requirement of the project and Evaluate the nature of the Data which we are going to work with. After creating the database named 'ig_clone' initially, number of tables created, the Schema of the tables, functions used in the creation, and Determining Primary & Foreign Key constraints are taken into consideration to work with the queries smoothly.

TECH-STACK USED:

Used Mysql Workbench v8.0.31 for creating a new Database.

Used command prompt console for computation of data and deriving meaningful results.

INSIGHTS:

In this project, the requirement requested in the queries doesn't seem to be very declarative. Evaluating in terms of user's activities, photos posted, likes, and tags used are utterly repetitive which fails in carrying out a focused business decision. As it is in trend nowadays, Instagram can include 2 steps for verification which might reflect in cutting down on Bot or Dummy accounts. Further on for declaring a business decision, an accurate output should be brought up with a lesser number of alternatives present for the same requirements which will result in achieving a comprehensive Goal.

RESULTS:

While working on this project, I came to know that the activities of a common person and cases of their interests are mostly identical even though the discrimination of users based on age, sex, race, and locality fails to differentiate the case of interest. All the insight of this data is generated by the immensely distinct user though the statistics generated are identical because the mentality of people is mostly mob driven and Trend centered which doesn't reflect the true soul interest of a person.

Mentioned below are the Questions along with there Queries performed & Outputs received;

A. MARKETING CAMPAIGN: -

Q1. Rewarding Most Loyal Users: People who have been using the platform for the longest time.

OBJECTIVE: Find the 5 oldest users of Instagram from the database provided

ANS:-

```
select * from users
order by created_at asc
limit 5;
```

OUTPUT:-

```
+---+-----+-----+
| 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 |
+---+-----+-----+
```

Q2. Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.

OBJECTIVE: Find the users who have never posted a single photo on Instagram

ANS:-

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

OUTPUT:-

+-----+		
username		
+-----+		
Aniya_Hackett		
Kasandra_Homenick		
Jaclyn81		
Rocio33		
Maxwell.Halvorson		
Tierra.Trantow		
Pearl7		
Ollie_Ledner37		
Mckenna17		
David.Osinski47		
Morgan.Kassulke		
Linnea59		
Duane60		
Julien_Schmidt		
Mike.Auer39		
Franco_Keebler64		
Nia_Haag		
Hulda.Macejkovic		
Leslie67		
Janelle.Nikolaus81		
Darby_Herzog		
Esther.Zulauf61		
Bartholome.Bernhard		
Jessyca_West		
Esmeralda.Mraz57		
Bethany20		
+-----+		

Q3. 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.

OBJECTIVE: Identify the winner of the contest and provide their details to the team

ANS:-

(NOTE:178 rows displayed without using limits in query. The range of top_likes in 178 records is in between 257-244.)

```
select u.username,
p.id,
p.user_id,
max(l.photo_id) as top_likes
from users u
join photos p
on u.id = p.user_id
join likes l
on p.user_id=l.user_id
group by u.username,
p.id,
p.user_id
order by top_likes desc
limit 1;
```

OUTPUT:-

```
+-----+-----+-----+
| username | id | user_id | top_likes |
+-----+-----+-----+
| Harley_Lind18 | 10 | 3 | 257 |
+-----+-----+-----+
```

Q4. Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

OBJECTIVE: Identify and suggest the top 5 most commonly used hashtags on the platform

ANS:-

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

OUTPUT:-

common_Hastags	most_used_tag_id
sunset	5
beauty	4
delicious	4
landscape	4
food	4

Q5. Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.

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

ANS:-

```
select dayofweek(created_at) as day_num,
count(*) as most_online
from users
group by day_num
order by most_online desc
limit 10;
```

OUTPUT:-

day_num	most_online
5	16
1	16
6	15
3	14
2	14
4	13
7	12

B. INVESTOR METRICS:-

Q1. User Engagement: Are users still as active and post on Instagram or they are making fewer posts

OBJECTIVE: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

ANS:-

(NOTE:- average_user_posts value 47.3307 rounded off by using ceil function)

```
select
ceil(avg(u.id)) as avg_user_posts,
count(*) as num_of_photos,
count(distinct(u.id)) as sum_users
from users u
join photos p
on u.id = p.user_id;
```

OUTPUT:-

avg_user_posts	num_of_photos	sum_users	
48	257	74	

Q2. Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

OBJECTIVE: 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).

-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x---END-OF-DOCUMENT---x-