

Instagram User Analysis

Project Description

This project aims to get useful insights from extracting raw data and also get a in-dept analysis of user engagement of Instagram which will help the product team to launch ad campaign and new features.

This project mainly focus on some important aspects.

☐ Marketing

- A. Loyal User Reward
- B. Inactive User Engagement
- C. Contest Winner Declaration
- D. Hashtag Research
- E. Ad Campaign Launch

☐ Investors

- A. User Engagement
- B. Bots and Fake Accounts

Approach:

The project was done suing MySQL, where queries were utilized to create a database from the raw data.

Tech-Stack Used

MySQL

The purpose of use this tool is to create the data base and store records. It is also used to carry out The required analysis by writing DQL queries.

1. Top 5 oldest user on Instagram

Input:

```
USE ig_clone;  
SELECT *FROM users  
ORDER BY created_at  
LIMIT 5;
```

Explain: From data, I want to get the data from users table and order this with created Date. I want 5 names so I limit this to 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
•	NULL	NULL	NULL

Highlight: According to data they are the first five user who have created their account on Instagram in 2016. They are oldest user since Instagram launched.

2. Inactive user engagement

Input:

```
SELECT p.id, username AS Inactive_username  
FROM photos p  
RIGHT JOIN users u  
ON p.user_id = u.id  
WHERE image_url IS Null;
```

Explain: I want inactive username. So get this data from photos table and join with user So that I get the output of inactive users by who haven't post any image until now by 'IS NULL' query.

Output:

Highlight: According to data I get the inactive usernames. There are total 26 users Who are inactive.

1	Inactive_username
2	Aniya_Hackett
3	Kasandra_Homenick
4	Jaclyn81
5	Rocio33
6	Maxwell.Halvorson
7	Tierra.Trantow
8	Pearl7
9	Ollie_Ledner37
10	Mckenna17
11	David.Osinski47
12	Morgan.Kassulke
13	Linnea59
14	Duane60
15	Julien_Schmidt
16	Mike.Auer39
17	Franco_Keebler64
18	Nia_Haag
19	Hulda.Macejkovic
20	Leslie67
21	Janelle.Nikolaus81
22	Darby_Herzog
23	Esther.Zulauf61
24	Bartholome.Bernhard
25	Jessyca_West
26	Esmeralda.Mraz57
27	Bethany20

3. Contest Winner Declaration

Input:

```
SELECT username,  
       p.id,  
       p.image_url,  
       count(l.user_id) AS total_likes  
FROM photos p  
JOIN likes l  
      ON l.user_id = p.user_id  
JOIN users u  
      ON p.user_id = u.id  
GROUP BY p.id  
ORDER BY total_likes DESC  
LIMIT 3;
```

Explain: I need the username of the winner of the contest. So I count user_id to get total likes from photos table join with user table. There are total 3 photo which likes are same.

Output:

	username	id	image_url	total_likes
►	Annalise.McKenzie 16	53	http://collin.com	103
	Annalise.McKenzie 16	52	https://hershel.com	103
	Annalise.McKenzie 16	51	https://abagail.com	103

Highlight: According to data, I get one username who win the contest. And I see this user get Same likes in 3 photos that is 103.

4. Hashtag Research

Input:

```
SELECT t.tag_name,  
       count(pt.photo_id) AS total_times_used  
FROM photo_tags pt  
JOIN tags t  
      ON pt.tag_id = t.id  
GROUP BY t.id  
ORDER BY total_times_used DESC  
LIMIT 5;
```

Explain: I need the hashtag name which is used most. So what I did , I count photo_id from photo tags table Join with the tag table.

Output:

	tag_name	total_times_used
▶	smile	59
	beach	42
	party	39
	fun	38
	concert	24

Highlight: According to the data, I got 5 tag name which is using most. Also I get the number how many times the tag is use.

5. Ad Campaign Launch

Input:

```
SELECT DATE_FORMAT((created_at), '%W') AS Best_Day_of_Week,  
       count(username) AS No_of_register  
FROM users  
GROUP BY Best_Day_of_Week  
ORDER BY No_of_register DESC  
LIMIT 1;
```

Explain: I need the date for ad campaign launch. So I run date format query and %W for get the day which day users are register most.

Output:

	Best_Day_of_Week	No_of_register
▶	Thursday	16

Highlight: The day is thursday and that day users are register most. Total register user is 16.

B. Investor Metrics

1. User Engagement

Input:

```
SELECT
    (SELECT count(users.username) FROM users) AS total_users,
    count(photos.id) AS total_photos,
    (SELECT count(*) FROM photos)/(SELECT count(*) FROM users) AS average_posts_per_user
FROM users, photos;
```

Explain: I need total username number and total photos so that I will get average post per user.

Output:

	total_users	total_photos	average_posts_per_user
▶	100	25700	2.5700

Highlight: I get the number of average posts per user by total photos divided by total users.

2. Bots & Fake Accounts

Input:

```
SELECT u.username, count(*) AS no_of_likes
FROM users u
JOIN likes l ON u.id = l.user_id
GROUP BY u.id
HAVING no_of_likes = (SELECT count(*) FROM photos);
```

Explain: I need username which accounts are fake. So I count number of likes form user table And group this by id from user table. If number of likes are same as the number of people than This account is fake.

Output:

	username	no_of_likes
▶	Aniya_Hackett	257
	J& Aniya_Hackett	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

Highlight: I get some usernames which no of likes are 257 same as total user number. This accounts are fake.

Result:

The project answers the questions and thus provides the required insights relevant for the product.

This project has helped me to understand how data can be of relevance to a business and has helped to hone my SQL skills and Excel skills.