INSTAGRAM USER ANALYTICS

Project Description:

The project is about to analyze the user interaction and engagement in a Instagram platform and provide some valuable insights to the product team that help business to grow by taking some decisions like giving rewards to top loyal users, organizing a contest to increase the user, etc.

Approach:

We first perform the Marketing Analysis such as: Loyal user reward, Inactive user engagement, Contest winner declaration, Hashtag Research, Ad campaign Launch that helps to attract more users and then find out Investor Metrics like User engagement, Bot and fake accounts to help the investor for better decisions.

Marketing Analysis:

Loyal User Reward: Reward the most loyal user

Task: find the top 5 oldest users in Instagram?

Query:

```
use ig_clone;
SELECT * FROM users;
SELECT username FROM users ORDER BY created_at ASC LIMIT 5;
```

First, I find all users, then search for top user by sorting the column 'created_at' in ascending order and limit of 5 to get 5 oldest user.

Output:

username
Darby_Herzog
Emilio_Bernier52
Elenor88
Nicole71
Jordyn.Jacobson2

Inactive User Engagement: To encourage the inactive users

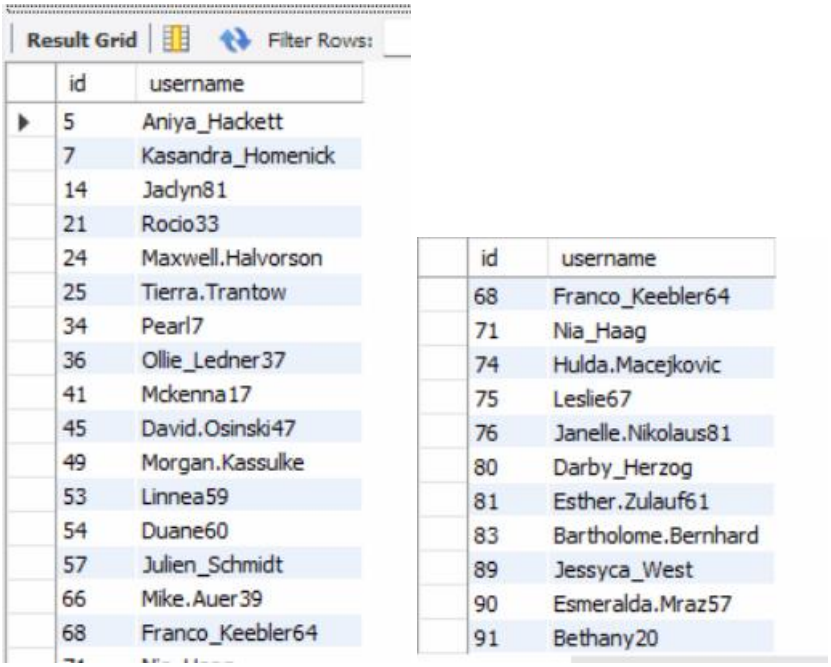
Task: Find users who never posted a photo single time?

Query:

```
4
5 •  SELECT users.id, users.username FROM users LEFT JOIN photos ON users.id = photos.user_id where photos.id is NULL;
```

First we JOIN the users table with photos table, then find the users whose photos.id is NULL , i.e users who does not having a photo.id has not posted a single pic.

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

id	username
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

Contest Winner Declaration: The team wants to declare a user with most likes on a single photo in a contest as a Contest Winner.

Task: Find the user details having maximum likes in a single photo to award him a winner?

```
12 •  SELECT photo_id FROM likes GROUP BY photo_id
13    ORDER BY COUNT(*) DESC LIMIT 1;
```

First, I want find the ‘photo_id’ which get the maximum likes using COUNT(*) functions in a likes table .

photo_id
145

```
19 •  SELECT * FROM users WHERE id = (SELECT user_id from photos WHERE photos.id = (SELECT photo_id FROM likes
20    GROUP BY photo_id ORDER BY COUNT(*) DESC LIMIT 1));
```

id	username	created_at
52	Zack_Kemmer93	2017-01-01 05:58:22
NULL	NULL	NULL

Then I find user details of user who posted that photo by using WHERE condition which helps to find out the user_id of user.

Hashtag Research: Most popular hashtags to reach to the most people

Task: Identify the top 5 hashtags of the platform.

Query

```
4
5 •  SELECT tag_id FROM photo_tags GROUP BY tag_id ORDER BY COUNT(*) DESC LIMIT 5;
```

Output:

Result Grid

tag_id
21
20
17
13
18

Query:

```
24 •  SELECT tag_name, tag_id, COUNT(*) AS tag_count FROM tags JOIN photo_tags ON tags.id = photo_tags.tag_id
25    GROUP BY tag_id ORDER BY tag_count DESC LIMIT 5;
```

Output:

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

First, I find out top 5 tag_id from photo_tags table and then find out the tag_name corresponding to 'tag.id' by using JOINS on tags and photo_tags table and identify the top 5 tags with also a tags count.

Ad campaign Launch: Best day to launch any ad by a team.

Task: Find the day of week when most user registered on Instagram and that day will be best for launching any Ad.

```
26
27 •  SELECT DAYNAME(created_at) as registration_day FROM users;
28
```

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

registration_day
Thursday
Sunday
Tuesday
Saturday
Wednesday
Sunday
Monday

First, we find the days at which users registers on Instagram.

```
27 •  SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS user_count FROM users
28    GROUP BY registration_day ORDER BY user_count DESC;
29
30
```

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

registration_day	user_count
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Then add a column ‘user_count’ to count the days when most user registers on Instagram.

Brand should launch any Ads On Thursday and Sunday.

Investor Metrics:

User engagement: Investor wants users who are still active and their posting behavior.

Task: Avg number of posts posted by active users i.e. , avg post per user and total number of posts divide by total number of users.

Query:

```
31 •  SELECT DISTINCT(user_id) FROM photos;
32 •  SELECT COUNT(DISTINCT(user_id)) FROM photos; # 74 user posted some post
```

Result Grid | Filter Rows

COUNT(DISTINCT(user_id))
74

First, I find the total number of active users.

Query:

```
SELECT COUNT(*) AS total_posts FROM photos; # 257 total post in instagram
```

total_posts
257

Then I find the total number of posts in Instagram.

Query:

```
SELECT COUNT(*) / COUNT(DISTINCT(user_id)) AS avg_post_per_user FROM photos; # AVG post per user
```

avg_post_per_user
3.4730

Avg photos posted by user in Instagram.

Query:

```
SELECT (SELECT COUNT(*)  FROM photos) AS total_post , (SELECT COUNT(*) FROM users) AS total_user,
 (SELECT COUNT(*)  FROM photos) / (SELECT COUNT(*) FROM users)  AS post_per_user; # post per user
```

total_post	total_user	post_per_user
257	100	2.5700

Total number of post in Instagram / Total number of users as post_per_user.

Bots & fake account: Investor wants to know platform has occupied by bots and fake accounts.

Task: User who likes every photo on a Instagram considered as a Suspicious user.

```
SELECT user_id FROM likes GROUP BY user_id
HAVING COUNT(DISTINCT photo_id) = (SELECT COUNT(*) FROM photos);
```

Find the users (user_id) from 'likes' table who likes every photo (photo_id) by equalizing count of likes done with count of photos posted on Instagram.

Tech_stack Used:

For this project we used **SQL** language for data analysis.

We use **MYSQL** as our RDBMS which used to store and manipulate data.

Insights:

- Instagram has mostly active users but still 25 percent users are still unactive on which team can focus to encourage them.
- Less than 50 % users liking the photos posted on Instagram Contest.
- 80 % of photos are posted with Hashtags.
- Thursday and Sunday are important days to organize any contest, launch Ads.
- On an average 3 photos are posted by active users in Instagram.
- 10 % users can consider to be suspicious users on Instagram.

Results:

We successfully find out the fake accounts to address the unusual behavior.

We find out best days to check the Instagram account to get latest updates.

Drive link:

https://drive.google.com/drive/folders/1f_DL85t8Xqn_-093R2HUVLLgYlIeBhgi?usp=drive_link