



Instagram User Analytics

Project Description:

This project aims to analyze Instagram dataset using SQL and MySQL Workbench to provide valuable insights to the management team and aid business growth. The insights gained from this analysis will help different teams within the organization make informed decisions about the future of the Instagram app. User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app.

Approach:

step 1: Understanding the database schema and contents.

step 2: Prepare the data for analysis. Check for data quality, completeness and accuracy.

step 2: Identifying the business tasks and check if the data you have is sufficient or not.

step 3: Write the queries and pull the data you need to answer the questions.

step 4: Using the data, analyze and answer the key questions

step 5: Provide insights to the stakeholders, supported by data and visualization if needed.

Tech-Stack Used:

MySQL Workbench 8.0.34: It's a capable and user-friendly tool for writing and executing SQL queries

Notion: It is a versatile tool that can be used for a variety of tasks, including project management, data analysis, and report generation as pdf.

Schema and insert statements:

```
ig_clone_data.sql
```

Insights:

Business tasks:

- Identify the five oldest users on Instagram from the provided database for loyal user rewards.

```
-- 1 which are the oldest 5 users?
```

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

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

- Identify users who have never posted a single photo on Instagram for inactive user engagement.

```
-- 2 find the users who are inactive (those who never posted a photo)
```

```
SELECT
    users.id, username
FROM
    users
    LEFT JOIN
    photos ON users.id = photos.user_id
WHERE
    image_url IS NULL
ORDER BY users.id;
```

users who never posted a single photo:

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
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

-
- Determine the winner of the contest(most likes on a single photo) and provide their details to the team.

```
-- 3 who got the most likes on a single photo

SELECT
    photo_id, COUNT(*) AS num_likes, photos.user_id, username
FROM
    likes
        JOIN
    photos ON likes.photo_id = photos.id
        JOIN
    users ON users.id = photos.user_id
GROUP BY photo_id
ORDER BY num_likes DESC
LIMIT 1;
```

photo_id	num_likes	user_id	username
145	48	52	Zack_Kemmer93

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

```
-- 4. top 5 most commonly used hashtags
```

```
SELECT
    tag_name, COUNT(*) AS count_tags
FROM
    photo_tags
        JOIN
    tags ON tags.id = photo_tags.tag_id
GROUP BY tag_name
ORDER BY count_tags DESC
limit 7;
```

Top 5 most commonly used hashtags:

tag_name	count_tags
smile	59
beach	42
party	39
fun	38
lol	24
concert	24
food	24

- Determine the day of the week when most users register on Instagram to decide when to launch an ad campaign.

SELECT

```
DAYNAME(created_at) AS day_of_week, COUNT(*) AS users_reg
FROM
users
GROUP BY day_of_week
ORDER BY users_reg DESC;
```

Days with most user registrations are Thursdays and Sundays.

day_of_week	users_reg
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

- Calculate the average number of posts per user on Instagram.

```
SELECT
    COUNT(*) / (SELECT
        COUNT(*) AS y
    FROM
        users) AS x
FROM
    photos;
```

On an average, a user uploads 2.57 posts.

X
2.5700

- Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

```
SELECT
    id, username, COUNT(id) AS likes_count
FROM
    users
JOIN
    likes ON users.id = likes.user_id
GROUP BY id
    HAVING likes_count = (SELECT
        COUNT(*)
    FROM
        photos);
```

These users are bot accounts as they liked every single pictures which is an unlikely event:

id	username	likes_count
5	Aniya_Hackett	257
14	Jaclyn81	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
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

Result:

- The project generated some key insights that can help the Instagram management team make data-driven decisions. Some of the key findings were:
- lists of oldest users and inactive users were generated for marketing purpose.
- Winner of the contest for most like on a photo was found out and his details like id were passed to the respective team.
- Most common hashtags were found out.
- Most active days of user registrations were Thursday and Sunday. Ad campaigns can be scheduled accordingly.
- On an average, user uploads 3.473 posts
- List of bot accounts were segregated and action will be taken by the concerned team.

The insights gained from this analysis can be used for making data-driven decisions regarding new products and features, user experience improvement etc.

Marketing team can make use of these insights to target audience more effectively and launch different campaigns.
