

Instagram User Analytics

▪ Project Description:

The purpose of this project is to analyse how users interact and engage with the Instagram application in order to derive actionable insights. The database has been provided by the team to work on the data to collect useful insights for marketing, product and development teams of Instagram. The project will answer the following problem statements:

A) **Marketing:** The marketing team wants to launch some campaigns, and they require help with following:

1. **Rewarding most loyal users.**
2. **Remind Inactive users to start posting via promotional emails.**
3. **Declare contest winner with the most likes on a single photo.**
4. **Research the most popular hashtags for a partner brand.**
5. **Best day to launch AD campaign.**

B) **Investor Metrics:** The investors of the company seek an assessment of Instagram's performance and its differentiation from other social media platforms, ensuring it remains relevant and impactful in the market. The assessment will focus on the following criteria:

1. **User Engagement.**
2. **Bots and fake accounts.**

▪ Approach:

To accomplish the necessary tasks and finalize the project, SQL queries were employed using the MySQL Command Line client. Following provided instructions, the database and corresponding tables were created, data was imported into MySQL, and relevant queries were executed to derive the required insights.

1. **Creating database:** The database file has been provided by the team. The following analyzation can be done from the provided database:

The database has been created using DML and DDL SQL queries which is called ig_clone. SQL codes have been used to create tables. There is total 7 tables created which are:

Users: To store information about the users.

Photos: To store information about photos uploaded by users.

Comments: To store information about comments made by user on the photos.

Like: To store information about the likes the user gives on the photos.

Follows: To store information about relationships between users who follow each other on Instagram.

Tags: To store information about tags which are associated with photos or comments.

Photo tags: To store information about relationships between tags and photos.

Each table consists some relevant column for analysis like user id, post id, timestamp, etc.

After creating the tables, the values are inserted into all the tables.

2. **Derive insights:** After creating the database, key insights were derived from the tables using SQL queries.

- **Tech-Stack used:**

MySQL 8.0.37 community version has been used to create database and to derive key insights. The reason behind using MySQL is that it is the most popular relational database management system (RDBMS) which is easy to use and understand and also the community version is free and open-source RDBMS.

- **Insights:**

A) Marketing:

1. **Rewarding most loyal users:** The task is to identify 5 oldest users of the Instagram from the provided database.


```
1 • USE ig_clone;
2
3   # 5 oldest users of the Instagram
4 • Select *
5   from users
6   order by created_at
7   limit 5;
```

Result Grid	Filter Rows:	Edit:
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

Insights: From the above information it can be observed that the oldest users of Instagram have registered on the platform in May 2016. The information also points towards Instagram's growth over time. Also, these old users must have adopted all the new updates and features over the years which might help the development team for future developments that align with user behaviour and preferences.

2. **Remind Inactive users to start posting via promotional emails:** The task is to identify users who have never posted a single photo on Instagram.

```
# Users who have never posted on Instagram
Select *
from users
left join photos
on users.id=photos.user_id
where photos.id is null
order by created_at;
```

Result Grid   Filter Rows: <input type="text"/>			
	id	username	created_at
▶	80	Darby_Herzog	2016-05-06 00:14:21
	71	Nia_Haag	2016-05-14 15:38:50
	91	Bethany20	2016-06-03 23:31:53
	66	Mike.Auer39	2016-07-01 17:36:15
	34	Pearl7	2016-07-08 21:42:01
	41	Mckenna17	2016-07-17 17:25:45
	76	Janelle.Nikolaus81	2016-07-21 09:26:09
	36	Ollie_Ledner37	2016-08-04 15:42:20
	89	Jessyca_West	2016-09-14 23:47:05
	75	Leslie67	2016-09-21 05:14:01
	25	Tierra.Trantow	2016-10-03 12:49:21
	49	Morgan.Kassulke	2016-10-30 12:42:31
	83	Bartholome.Bern...	2016-11-06 02:31:23
	68	Franco_Keebler64	2016-11-13 20:09:27
	5	Aniya_Hackett	2016-12-07 01:04:39
	7	Kasandra_Homenick	2016-12-12 06:50:08
	54	Duane60	2016-12-21 04:43:38
	81	Esther.Zulauf61	2017-01-14 17:02:34
	21	Rocio33	2017-01-23 11:51:15
	74	Hulda.Macejkovic	2017-01-25 17:17:28
	57	Julien_Schmidt	2017-02-02 23:12:48
	45	David.Osinski47	2017-02-05 21:23:37
	14	Jadyn81	2017-02-06 23:29:16
	53	Linnea59	2017-02-07 07:49:34
	90	Esmeralda.Mraz57	2017-03-03 11:52:27
	24	Maxwell.Halvorson	2017-04-18 02:32:44

Insights: From the above information it can be observed that there is total 26 users who have never posted any photo on Instagram. It can also be observed that all the users who have never posted any photo have all created their account either in the year 2016 or 2017. It can mean they are old users who are either not active on Instagram frequently or the account is no more in use. The marketing team can send promotional emails to these inactive users to remind them about posting on Instagram. The marketing team can also understand the audience behaviour through inactive users as these inactive users may indicate that the contents on the platform may not be relevant to them.

3. **Declare contest winner with the most likes on a single photo:** The task is to determine the winner of the contest and provide their details to the team.

```
#Determine the contest winner with most likes
select
  likes.photo_id,
  photos.image_url,
  photos.created_dat as `Date of Photo upload`,
  users.username,
  users.created_at as `Date of Account Creation`,
  count(likes.user_id) as `Total likes`
from likes
inner join photos on likes.photo_id = photos.id
inner join users on photos.user_id = users.id
group by likes.photo_id, photos.image_url, photos.created_dat, users.username, users.created_at
order by `Total likes` desc
limit 5;
```



	photo_id	image_url	Date of Photo upload	username	Date of Account Creation	Total likes
▶	145	https://jarret.name	2024-06-20 18:52:41	Zack_Kemmer93	2017-01-01 05:58:22	48
	127	https://celestine.name	2024-06-20 18:52:41	Malinda_Streich	2016-07-09 21:37:08	43
	182	https://dorcias.biz	2024-06-20 18:52:41	Adelle96	2016-10-01 00:37:57	43
	123	http://shannon.org	2024-06-20 18:52:41	Seth46	2016-07-07 11:40:27	42
	61	https://dejon.name	2024-06-20 18:52:41	Delpha.Kihn	2016-08-31 02:42:30	41


Insights: From the above data it can be observed that user with username Zack_Kemmer93 has posted the photo with maximum number of likes and hence is declared the winner of the contest. The data also shows that the top 5 users with most liked photo has more than 40 likes and are all old users since the account creation varies between the year 2016 and 2017. The photos which have received most likes are all posted recently in the year 2024. The main information which can be inferred from the above data is that the users with most liked photos are either influencers or highly engaged users. The marketing team can collaborate with these users for promotional purposes and marketing campaigns.

4. **Research the most popular hashtags for a partner brand:** The task is to identify and suggest the top 5 most commonly used hashtags on Instagram.

```
#Top 5 hashtags on Instagram
select
    tags.id as `Tag ID`,
    tags.created_at as `Tag creation date`,
    tags.tag_name,
    count(photo_tags.photo_id) as `Total occurrences`
from photo_tags
inner join
    tags on tags.id = photo_tags.tag_id
group by tags.tag_name
order by `Total occurrences` desc
limit 5;
```

Result Grid



Filter Rows:

Export:


	Tag ID	Tag creation date	tag_name	Total occurrences
▶	21	2024-06-20 18:52:42	smile	59
	20	2024-06-20 18:52:42	beach	42
	17	2024-06-20 18:52:42	party	39
	13	2024-06-20 18:52:42	fun	38
	18	2024-06-20 18:52:42	concert	24

Insights: From the above data it can be observed that top 5 most commonly used hashtags are smile, beach, party, fun, concert. The data also shows total occurrences of each tag. The most popular hashtag is smile which is used considerably more than other hashtags which means it is more popular and trending among users. The partner brand can benefit from these hashtags for marketing campaign, promotional activities and maximize their visibility and engagement.

- Best day to launch AD campaign:** The task is to determine the day of the week when most users register on Instagram and also to provide insights on when to schedule the ad campaign.

```
#Best day to launch ad campaign
select
    dayname(created_at) as `Day of week`,
    count(username) as `Number of registered users`
from users
group by `Day of week`
order by `Number of registered users` desc;
```

Result Grid			Filter Rows:
	Day of week	Number of registered users	
▶	Thursday	16	
	Sunday	16	
	Friday	15	
	Tuesday	14	
	Monday	14	
	Wednesday	13	
	Saturday	12	

Insights: From the above data it is observed that most users register on Thursday and Sunday. Hence, the best day to launch AD campaign is Thursday and Sunday as it will make ad campaign successful. Furthermore, this can help Instagram to improve how they welcome new users and make decisions to improve their experience and attract more users in the future.

B) Investor Metrics:

- User Engagement:** The task is to calculate the average number of posts per user and also provide the total number of photos on Instagram divided by the total number of users.

```
#User Engagement
with userengagement as (
  select
    users.id as userid,
    count(photos.id) as photoid
  from users
  left join
    photos on photos.user_id = users.id
  group by users.id
)
select
  sum(photoid) as `Total Number of Photos`,
  count(userengagement.userid) as `Total Number of Users`,
  sum(photoid) / count(userengagement.userid) as `Total Photos / Total Users`
from userengagement;
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	Total Number of Photos	Total Number of Users	Total Photos / Total Users		
▶	257	100	2.5700		

Insights: From the above data it is observed that average number of posts per user is 2.57. This means that an average user posts between 2 to 3 times. The investors can

access this data and conclude that the user engagement is low and the team needs to plan content and engagement strategies to maximise user engagement.

2. **Bots and fake accounts:** The task is identifying potential bots who have liked every single photo on Instagram.

```
#Bots and fake accounts
with bots as(
select users.username,
       users.id as `user ID`,
       count(likes.photo_id) as `likes on all photos`
from likes
inner join users
on users.id=likes.user_id
group by users.username, users.id)
select username,`user ID`,`likes on all photos`
from bots
where `likes on all photos`=(select count(*) from photos)
order by username;
```

Result Grid			
		Filter Rows:	Export
	username	user ID	likes on all photos
▶	Aniya_Hackett	5	257
	Bethany20	91	257
	Duane60	54	257
	Jacyn81	14	257
	Janelle.Nikolaus81	76	257
	Julien_Schmidt	57	257
	Leslie67	75	257
	Maxwell.Halvorson	24	257
	Mckenna17	41	257
	Mike.Auer39	66	257
	Nia_Haag	71	257
	Ollie_Ledner37	36	257
	Rocio33	21	257

Insights: The above data shows the list of all users who have liked every photo on Instagram. Since the database consists of total 257 users, these accounts have liked all 257 photos which is not possible for a normal user. Therefore, these accounts are bots. There is total 13 bot accounts. The investors can assess these data and ask the team to remove these potential bots so that they can review the performance of the platform based on actual data. Furthermore, the above data can also help in implementing strategies to maintain trustworthy user environment.

- **Results:**

Through this project as a data analyst in Instagram, I gained significant insights that proved to be valuable. It deepened my knowledge of data analysis in organisations and process of making data driven decisions. This experience helped in improving my analytical skills and SQL skills. I became comfortable in performing queries and analysing datasets. By using MySQL Workbench, I extracted meaningful insights from datasets of users on Instagram and performed various tasks. This helped me in understanding how users interact within the application. The insights I provided will be used by different teams within Instagram like, the marketing team used data on popular hashtags to plan successful campaign, while the investors used user engagement and bot accounts to access the performance of the platform.

In conclusion this project was not just about analysing the data but also using insights to make Instagram a better platform for its users. This experience will further help in boosting my career as data analyst in shaping successful products and services in various organizations.