Project Description

We have here a vast data of 100 users generated over the period of 06/05/2016 to 04/05/2017 from one of the most famous social media platforms available to the public to use. Apart from users data additionally we have here photos, likes, hashtags, comments, follows, tags etc. this report is made in an effort to analyse user interaction so that we can derive some helpful insights to improve user interaction and grow business.
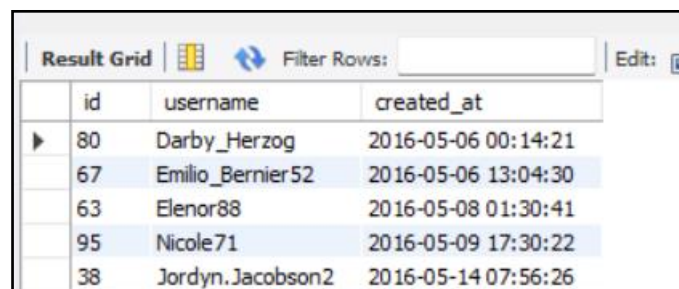
Approach

I started with understanding the characteristic of available data to find insightful information within metadata, like finding out how many users are there in the available data and over what period these users were registered in the Instagram application, and started looking for answers of some meaningful questions mentioned below;

A) Marketing Analysis:

1. **Loyal User Reward:** The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time. Your Task: Identify the five oldest users on Instagram from the provided database.

```
select *
from users
order by created_at
limit 5
```
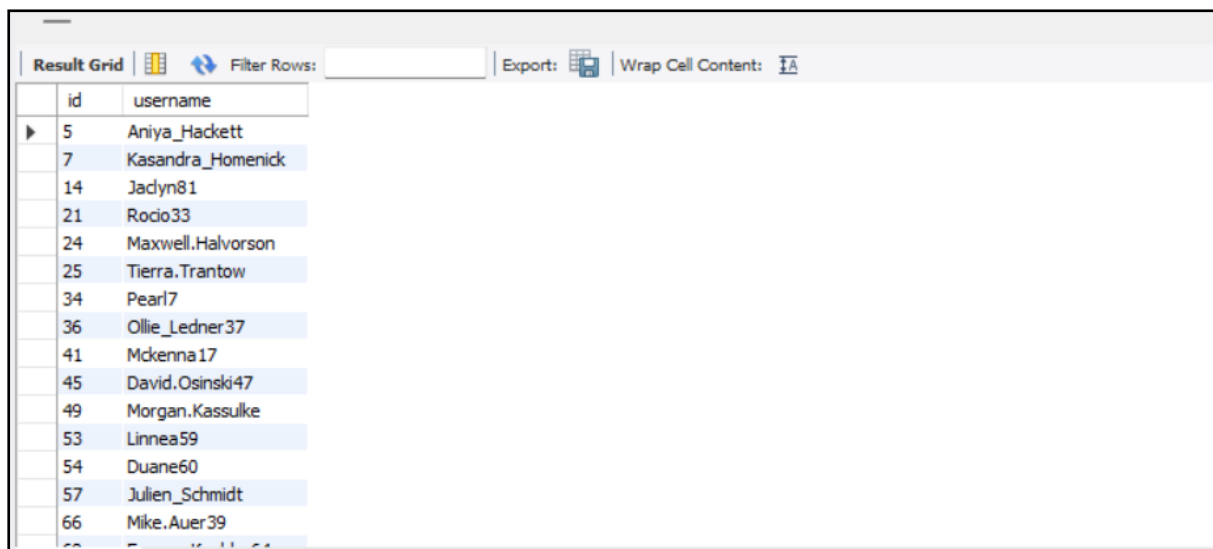
	id	username	created_at
▶	80	Darby_Herzog	2016-05-06 00:14:21
	67	Emilio_Bernier52	2016-05-06 13:04:30
	63	Elenor88	2016-05-08 01:30:41
	95	Nicole71	2016-05-09 17:30:22
	38	Jordyn.Jacobson2	2016-05-14 07:56:26

Rewarding these loyal users will create a wonderful environment for all the users of Instagram and the platform will be positioned to be a user-friendly platform.

2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
    Your Task: Identify users who have never posted a single photo on Instagram.

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


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

With the available list of inactive users, we can send them promotional emails to attract them to the application additionally we can ping some of the most famous content available in app and send it to the inactive users in turn post enjoying the famous content they would want to be more active with the app.

3. Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.
Your Task: Determine the winner of the contest and provide their details to the team.

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

username	No_of_likes	photo_id
Zack_Kemmer93	48	145

A post or photo with highest likes is most famous post across platform, in order to promote people to post such more content over time, we can award them winner of the contest to promote the incredible content and create an attractive environment for users to be inspired for creation and sharing such content which in turn attracts users to stay more active in the application.

4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
   Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

```sql
select tag_name, count(photo_id) as No_of_hashtags, tag_id
from tags as t inner join photo_tags as pt on t.id=pt.tag_id
group by tag_name
order by No_of_hashtags desc
limit 5
```
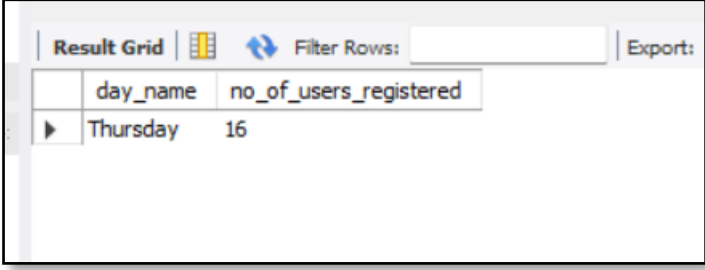
tag_name	No_of_hashtags	tag_id
smile	59	21
beach	42	20
party	39	17
fun	38	13
concert	24	18

Hashtag is a keyword that associates any post with user's interest, the most used hashtag would end up meeting more users. This keyword is extremely useful for partner brand to use and get attention of more users as the post shared by partner brand will land to the user's home screen due to the interest in the said hashtag.

5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads.
   Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

select DAYNAME(created_at) as day_name, COUNT(id) as no_of_users_registered
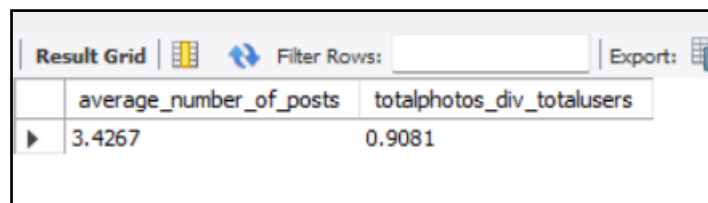from users
group by DAYNAME(created_at)
order by no_of_users_registered desc
limit 1

With the available data it is understood that most of users are registered on Thursday and Sunday. this information is vital in order to help customers do efficient marketing and reach maximum number of people in shortest period of time.

B) Investor Metrics:

1. User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
   Your Task: Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

select avg(countposts_per_user) as average_number_of_posts ,(sum(total_photos)/ sum(total_users)) as totalphotos_div_totalusers
from(
select count(p.id) as countposts_per_user, count(p.id) total_photos, count(u.id) total_users
from users u left join photos p on u.id=p.user_id
group by user_id) t

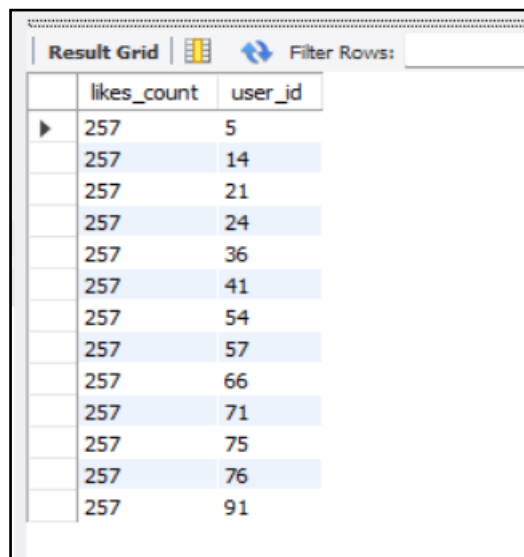| average_number_of_posts	totalphotos_div_totalusers
3.4267	0.9081

Keeping a track of total users to active users ratio is very important for the platform in order to understand pain points and areas that requires more attention. Here average number of posts per user is 3.4267~ +ve of 3 but that number alone is not enough to understand the stats of users. Toal of posts by users to total users ratio is what determines the activeness of overall users. With this information we can send more attractive content as recommendation especially to the inactive users and attract them to create such content in search of fame.

2. Botanical **Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.
   Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

select count(photo_id) likes_count, l.user_id
from users u join photos p on u.id =p.user_id join likes l on p.id=l.photo_id
group by l.user_id
having COUNT(photo_id) = (SELECT MAX(likes_count) FROM (SELECT
COUNT(photo_id) AS likes_count FROM likes GROUP BY user_id) t)
order by 1 desc

	likes_count	user_id
▶	257	5
	257	14
	257	21
	257	24
	257	36
	257	41
	257	54
	257	57
	257	66
	257	71
	257	75
	257	76
	257	91

Fake and dummy accounts are a threat not only to the platform but to the users and unethical marketing too, these accounts disrupt available information by creating wrong data leaving void and unreliable information, with the help of this list we can warn the respected users to test their characteristic and eventually if unchanged ban them from the platform making the platform free of such unwanted bots!

Tech-Stack Used

I used mysql workbench to analyse the data, availability of data was in line with requisite format of Mysql workbench hence it was best efficient platform to perform queries against the metadata

Insights

The business model of Instagram is quite interesting with two major metrics no. of users and no. of customers, more users attract more customers (i.e advertisers and subscription buyers). The data allowed me to attain a lot of information like most loyal users of Instagram and by awarding them overall experience of users is enhanced and creates a wonderful environment to stay active and socialise more. This data provided with information about users and from the data I could figure out inactive users and by attracting them to stay active the platform can benefit from it. And awarding those who provide the most engaging content will benefit users and enhance overall experience of all users in turn creating more and more opportunities for our customers.

Analysing all of the data helps understand a lot of patterns behind the behaviour of users, like we found out that majorly people sign up on Thursday and Sunday, with that information we can deploy correct efficient marketing strategy by pulling different ads on Thursday and Sunday but that is not all of it, with help of most trending hashtag information found from the data can aid marketing and reach furthermore by mentioning the same hashtag in the ad or post by customers.

Instagram is a wonderful platform to socialise and majority of population actively using it is Genz public.in order to keep the flow of platform smooth metrics like average posts per user and overall post to total users ratio is very helpful low numbers would be a sign of pain point and we can deploy more efforts to that part by encouraging people to create and share incredible content to attract more and more people, here in this data we saw that nearly average posts are 3 per user but content to user ratio is 0.9 which is 90% and will significantly rise overtime as platform thrives.

Result

1. Information of top 5 loyal users
2. List of inactive users (26)
3. No.1 contest winner with most likes on a photo (48 likes)
4. List of top 5 hashtags
5. Finding most active day of the week
6. Average no. of posts per user – 3.42 and total content to total user = 0.9
7. List of Bots/dummy account (13 No.)