**Project Description:** In this project we are trying to analyse instagram dataset . We will try to draw some key insights which can be used to grow the business (like finding bot users)

**Approach:** At first created the database ig_clone, Then created tables like *users,photos,likes,follows,comments,tags,photo_tags*. Then inserted the data into the respective tables**.** Then executed some SQL queries as per need to get insight from the data.

**Tech-Stack Used:** MySQL Server as the database, MySQL WORKBENCH as a query editor
**Insights:** Below insights we could make like :

- 5 oldest users
- Users who have not yet uploaded a single photo
- The user whose photo got the maximum like and the photo details
- Top 5 commonly used hashtags
- On which day maximum users were registered on Instagram
- Total users , Total photos, avg photo per user
- Photos per User
- Bot Users on instagram

**SQL Queries & Outputs:**

```
use ig_clone;
```

**/* Rewarding Most Loyal Users: People who have been using the platform for the longest time.
Your Task: Find the 5 oldest users of the Instagram from the database provided*/**

```
select * from users
order by created_at asc
limit 5;
```

```
2   /* Rewarding Most Loyal Users: People who have been using the platform for the longest time.
3   Your Task: Find the 5 oldest users of the Instagram from the database provided*/
4
5   select * from users
6   order by created_at asc
7   limit 5;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows:

| 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 |

**/*Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.
Your Task: Find the users who have never posted a single photo on Instagram*/**

```
(select * from users where id not in
(select distinct user_id from photos))
```

```
9   /*Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.
10  Your Task: Find the users who have never posted a single photo on Instagram*/
11
12  (select * from users where id not in
13  (select distinct user_id from photos))
14
```

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

| id | username | created_at |
|---|---|---|
| 5 | Aniya_Hackett | 2016-12-07 01:04:39 |
| 7 | Kasandra_Homenick | 2016-12-12 06:50:08 |
| 14 | Jaclyn81 | 2017-02-06 23:29:16 |
| 21 | Rocio33 | 2017-01-23 11:51:15 |
| 24 | Maxwell.Halvorson | 2017-04-18 02:32:44 |
| 25 | Tierra.Trantow | 2016-10-03 12:49:21 |
| 34 | Pearl7 | 2016-07-08 21:42:01 |
| 36 | Ollie_Ledner37 | 2016-08-04 15:42:20 |
| 41 | Mckenna17 | 2016-07-17 17:25:45 |
| 45 | David.Osinski47 | 2017-02-05 21:23:37 |
| 49 | Morgan.Kassulke | 2016-10-30 12:42:31 |
| 53 | Linnea59 | 2017-02-07 07:49:34 |
| 54 | Duane60 | 2016-12-21 04:43:38 |

Users who have not posted single photo.c

**/*Declaring Contest Winner: The team started a contest and the user who gets the most likes on a single photo will win**
**the contest now they wish to declare the winner.**
**Your Task: Identify the winner of the contest and provide their details to the team */**

#with cte approach

```
with cte1 as
(select photo_id,count(*) as total_likes
from likes
group by photo_id
order by 2 desc
limit 1),
cte2 as
(select *
from photos p
join cte1 c
on p.id=c.photo_id )
select user_id,photo_id,image_url,total_likes from cte2
```

```
with cte1 as
(select photo_id,count(*) as total_likes
from likes
group by photo_id
order by 2 desc
limit 1),
cte2 as
(select *
from photos p
join cte1 c
on p.id=c.photo_id )
select u.id,u.username,c2.photo_id,c2.image_url,c2.total_likes
from users u join cte2 c2
on u.id=c2.user_id
```

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

| id | username | photo_id | image_url | total_likes |
|---|---|---|---|---|
| 52 | Zack_Kemmer93 | 145 | https://jarret.name | 48 |

**#Another Way**

```
select u.username,p.id,p.image_url,COUNT(*) AS total_likes
from photos p
join likes l
on l.photo_id = p.id
join users u
on p.user_id = u.id
group by p.id
order by 4 desc
limit 1;
```

**/* Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.**
**Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform */**

```
select t.id,t.tag_name,count(*) as tag_count
from tags t
inner join photo_tags pt
on t.id=pt.tag_id
group by t.id
order by 3 desc
limit 5
```

**#Another Way**

```
with cte1 as(
select tag_id,count(*) as tag_count
from photo_tags
group by tag_id
order by 2 desc
limit 5)
```

select tag_id,tag_name,tag_count
from tags t inner join cte1 c
on t.id=c.tag_id

```
/* Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform */
select t.id,t.tag_name,count(*) as tag_count
from tags t
inner join photo_tags pt
on t.id=pt.tag_id
group by t.id
order by 3 desc
limit 5
```

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

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

**/*Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.**
**Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign**
**/***
select dayname(created_at),count(*) from users
group by dayname(created_at)
order by 2 desc

**#more generic approach**

with cte1 as(select dayname(created_at) as day_name,count(*) as tc
from users group by dayname(created_at) order by 2 desc),
cte2 as (select day_name,tc,dense_rank() over(order by tc desc) as 'rn1' from cte1 )
select * from cte2 where rn1=1

```
/*Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.
Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign */

  #more generic approach

with cte1 as(select dayname(created_at) as day_name,count(*) as tc
from users group by dayname(created_at) order by 2 desc),
cte2 as (select day_name,tc,dense_rank() over(order by tc desc) as 'rn1' from cte1 )
select * from cte2 where rn1=1
```

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

| day_name | tc | rn1 |
|---|---|---|
| Thursday | 16 | 1 |
| Sunday | 16 | 1 |

**/*User Engagement: Are users still as active and post on Instagram or they are making fewer posts Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users*/**

# photos per user
select  p.user_id,u.username,count(p.id)
from users u
inner join photos p
on u.id=p.user_id
group by p.user_id

```
85  /*User Engagement: Are users still as active and post on Instagram or they are making fewer posts
86  Your Task: Provide how many times does average user posts on Instagram. Also, provide the total
87  number of photos on Instagram/total number of users*/
88
89  # photos per user
90  select  p.user_id,u.username,count(p.id)
91  from users u
92  inner join photos p
93  on u.id=p.user_id
94  group by p.user_id
95
```

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

| user_id | username | count(p.id) |
|---|---|---|
| 1 | Kenton_Kirlin | 5 |
| 2 | Andre_Purdy85 | 4 |
| 3 | Harley_Lind18 | 4 |
| 4 | Arely_Bogan63 | 3 |
| 6 | Travon.Waters | 5 |
| 8 | Tabitha_Schamberger11 | 4 |
| 9 | Gus93 | 4 |
| 10 | Presley_McClure | 3 |
| 11 | Justina.Gaylord27 | 5 |
| 12 | Dereck65 | 4 |

#cte1 gives total photos,cte2gives total users,total query gives avg post per user
with cte1 as (SELECT count(id) as total_photos FROM   photos),
cte2 as (SELECT count(id) as total_users FROM   users)
select round(cte1.total_photos/cte2.total_users,2) as avg_post_per_user from cte1 inner join cte2

```
96   #cte1 gives total photos,cte2gives total users,total query gives avg post per user
97   with cte1 as (SELECT count(id) as total_photos FROM   photos),
98   cte2 as (SELECT count(id) as total_users FROM   users)
99   select round(cte1.total_photos/cte2.total_users,2) as avg_post_per_user from cte1 inner join cte2
100
```

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

| total_photos |
|---|
| 257 |

```
96    #cte1 gives total photos,cte2gives total users,total query gives avg post per user
97    with cte1 as (SELECT count(id) as total_photos FROM   photos),
98    cte2 as (SELECT count(id) as total_users FROM   users)
99    select round(cte1.total_photos/cte2.total_users,2) as avg_post_per_user from cte1 inner join cte2
100
```

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

| total_users |
|---|
| 100 |

```
96    #cte1 gives total photos,cte2gives total users,total query gives avg post per user
97    with cte1 as (SELECT count(id) as total_photos FROM   photos),
98    cte2 as (SELECT count(id) as total_users FROM   users)
99    select round(cte1.total_photos/cte2.total_users,2) as avg_post_per_user from cte1 inner join cte2
100
```

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

| avg_post_per_user |
|---|
| 2.57 |

**/*Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts**
**Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).*/**

```
with cte1 as
(
select count(*) as max_possible_like_by_bot from photos),
cte2 as
(select u.id,u.username, count(*) AS num_likes
from users u
inner join likes l
on u.id =l.user_id
group by l.user_id)
select * from cte2
join cte1
where cte2.num_likes=cte1.max_possible_like_by_bot
```

```sql
/*Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts
Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).*/

with cte1 as
(
select count(*) as max_possible_like_by_bot from photos),
cte2 as
(select u.id,u.username, count(*) AS num_likes
from users u
inner join likes l
on u.id =l.user_id
group by l.user_id)
select * from cte2
join cte1
where cte2.num_likes=cte1.max_possible_like_by_bot
```

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

| id | username | num_likes | max_possible_like_by_bot |
|---|---|---|---|
| 5 | Aniya_Hackett | 257 | 257 |
| 14 | Jadyn81 | 257 | 257 |
| 21 | Rocio33 | 257 | 257 |
| 24 | Maxwell.Halvorson | 257 | 257 |
| 36 | Ollie_Ledner37 | 257 | 257 |
| 41 | Mckenna17 | 257 | 257 |
| 54 | Duane60 | 257 | 257 |

bot Users.csv

**Result:** I was able to create meaningful insights from the dataset which can be crucial for business. I was able to sharpen my SQL skills from here. Successfully completing this project boosted my confidence.