**Project Description :**

The aim of this project is to perform analysis on Instagram user data and gain insights into the behavior of Instagram users. SQL was used as a primary tool for data analysis and retrieval. The findings from the analysis provide valuable insights into user activity, preferences and demographics on the popular social media platform. This report presents the methodology and results of the analysis and discusses the implications of the findings for businesses and marketers."

**Approach :**

* Database Creation: Use SQL to create a database that can store the data and support the queries you need to perform the analysis.
* Data Loading: Load the data into the database.
* Data Analysis: Use SQL queries to perform the analysis and retrieve the insights you want to extract.

**Tech-Stack Used :**

1. MySQL

**Insights :**

Below I have provided SQL queries and their respective outputs and with insights/outcomes for each of the 7 questions.

**A) Marketing:**

1 - Rewarding Most Loyal Users:

select username, created\_at

from users

order by created\_at asc

|  |  |
| --- | --- |
| **username** | **created\_at** |
| Darby\_Herzog | 06-05-2016 00:14 |
| Emilio\_Bernier52 | 06-05-2016 13:04 |
| Elenor88 | 08-05-2016 01:30 |
| Nicole71 | 09-05-2016 17:30 |
| Jordyn.Jacobson2 | 14-05-2016 07:56 |

limit 5;

These 5 users are the oldest user in the database.

2 - Remind Inactive Users to Start Posting:

select u.username, count(p.user\_id)

from users as u

left join photos as p

on u.id = p.user\_id

|  |  |
| --- | --- |
| **username** | **count(p.user\_id)** |
| Aniya\_Hackett | 0 |
| Kasandra\_Homenick | 0 |
| Jaclyn81 | 0 |
| Rocio33 | 0 |
| Maxwell.Halvorson | 0 |
| Tierra.Trantow | 0 |
| Pearl7 | 0 |
| Ollie\_Ledner37 | 0 |
| Mckenna17 | 0 |
| David.Osinski47 | 0 |
| Morgan.Kassulke | 0 |
| Linnea59 | 0 |
| Duane60 | 0 |
| Julien\_Schmidt | 0 |
| Mike.Auer39 | 0 |
| Franco\_Keebler64 | 0 |
| Nia\_Haag | 0 |
| Hulda.Macejkovic | 0 |
| Leslie67 | 0 |
| Janelle.Nikolaus81 | 0 |
| Darby\_Herzog | 0 |
| Esther.Zulauf61 | 0 |
| Bartholome.Bernhard | 0 |
| Jessyca\_West | 0 |
| Esmeralda.Mraz57 | 0 |
| Bethany20 | 0 |

group by u.id

having count(p.user\_id) = 0

Here is a list of users who have not posted anything on Instagram.

3 - Declaring Contest Winner:

select u.username, u.id, count(l.photo\_id)

from users as u

left join photos as p

on u.id = p.user\_id

left join likes as l

on p.id = l.photo\_id

group by u.username, l.photo\_id, u.id

order by count(l.photo\_id) desc

limit 1;

|  |  |  |
| --- | --- | --- |
| **username** | **id** | **count(l.photo\_id)** |
| Zack\_Kemmer93 | 52 | 48 |

Zack\_Kemmer93 is the winner with most number of likes on a single photo.

4 - Hashtag Researching

select tag\_name, count(t.id)

from photos as p

left join photo\_tags as pt

on p.id = pt.photo\_id

left join tags as t

on pt.tag\_id = t.id

Group by t.id

order by count(id) desc

limit 5;

|  |  |
| --- | --- |
| **tag\_name** | **count(t.id)** |
| smile | 59 |
| beach | 42 |
| party | 39 |
| fun | 38 |
| concert | 24 |

Here is a list of all the hashtags used along with the number of times they have been used.

5 - Launch AD Campaign

with cte as

( select

case

when weekday(created\_at) = '0' then 'Monday'

when weekday(created\_at) = '1' then 'Tueday'

when weekday(created\_at) = '2' then 'Wedday'

when weekday(created\_at) = '3' then 'Thursday'

when weekday(created\_at) = '4' then 'Friday'

when weekday(created\_at) = '5' then 'Saturday'

when weekday(created\_at) = '6' then 'Sunday'

END as days\_of\_week

from users )

SELECT days\_of\_week, count(days\_of\_week)

from cte

|  |  |
| --- | --- |
| **days\_of\_week** | **count(days\_of\_week)** |
| Thursday | 16 |
| Sunday | 16 |
| Tueday | 14 |
| Saturday | 12 |
| Wedday | 13 |
| Monday | 14 |
| Friday | 15 |

group by days\_of\_week

Extracted a list of all the weekdays on which respective number of accounts were created.

“weekday()” function was used to extract weekday from DATETIME data type column.

**B) Investor Metrics -**

6 - User Engagement:

with cte as

( select u.id, count(p.id) as post\_per\_user

from users as u

left join photos as p

on u.id = p.user\_id

group by u.id

order by post\_per\_user desc )

select post\_per\_user, count(post\_per\_user)

from cte

group by post\_per\_user

|  |  |
| --- | --- |
| **post\_per\_user** | **count(post\_per\_user)** |
| 12 | 1 |
| 11 | 1 |
| 10 | 1 |
| 9 | 1 |
| 8 | 2 |
| 6 | 1 |
| 5 | 14 |
| 4 | 13 |
| 3 | 9 |
| 2 | 13 |
| 1 | 18 |
| 0 | 26 |

The above table show number of posts per user and how many users have posted each number of time. So there is only one user who has posted 12 times and two users who have posted 8 times and so on. From the extracted information we can say that an average user posts **less than 6 times**.

with cte as

( select u.id, count(p.id) as post\_per\_user

from users as u

left join photos as p

on u.id = p.user\_id

group by u.id

order by post\_per\_user desc )

select avg(post\_per\_user)

from cte

|  |
| --- |
| **avg(post\_per\_user)** |
| 2.57 |

The total number of photos on Instagram/total number of users = **2.57**

7 - Bots & Fake Accounts:

with cte as

( select l.user\_id, count(l.user\_id) as number\_of\_likes

from photos as p

left join likes as l

on p.id = l.photo\_id

left join users as u

on p.user\_id = u.id

group by l.user\_id )

select user\_id, u.username,number\_of\_likes

from cte

left join users as u

|  |  |  |
| --- | --- | --- |
| **user\_id** | **username** | **number\_of\_likes** |
| 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 |

on cte.user\_id = u.id

where number\_of\_likes = 257

These are the accounts that have liked all the photos that are available, So we can conclude that these are bots.

**Result :**

Extracted many useful insights that could help in making data driven decisions for the business.

Used MySQL, had to study few advanced topics like CTEs and datetime functions.