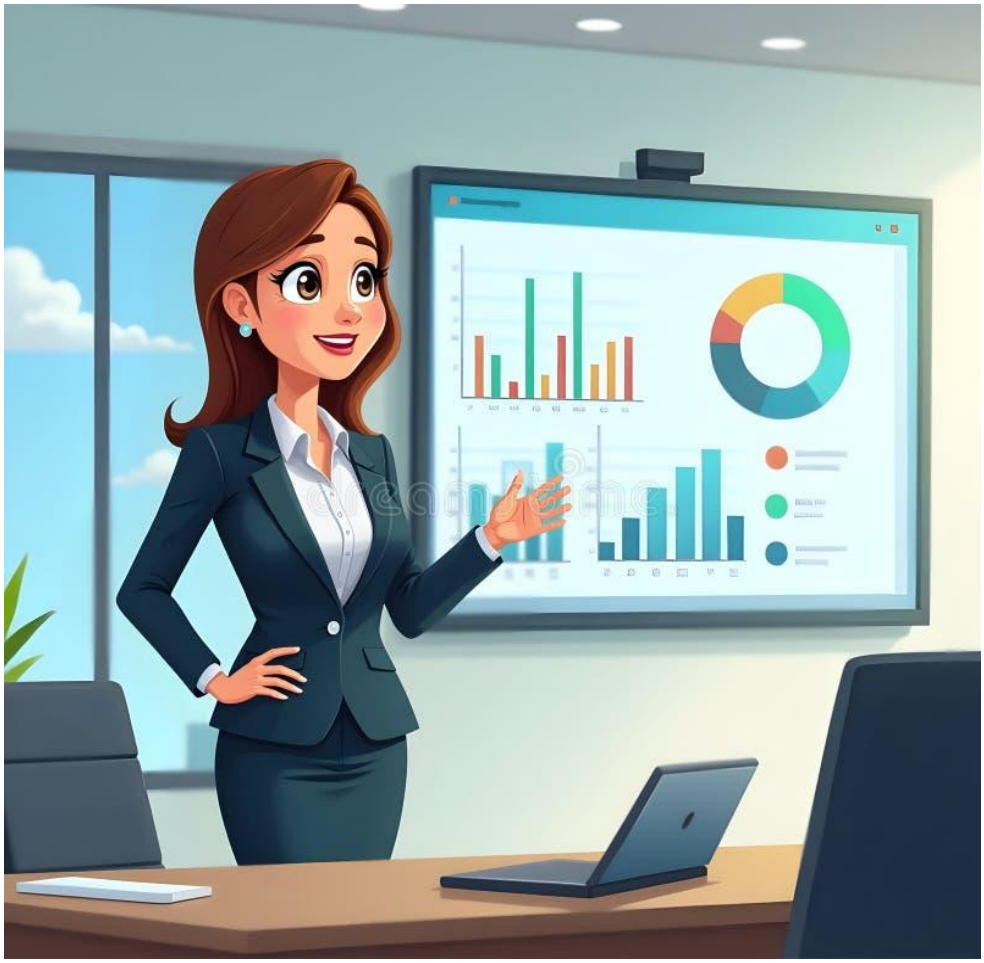
SOCIAL MEDIA HABITS AND FEELINGS PROJECT

SQL INSIGHTS



CREATED BY

YOKESH PREETHI S

DATA ANALYST

PROJECT OVERVIEW

This project explores social media usage and its perceived impact across different demographics. The findings highlight that the majority of users in this dataset are young (18-25) students who predominantly use Facebook. While Facebook has the highest number of users reporting both positive and negative impacts, negative feedback appears common across various time spent durations on social media. Interestingly, WhatsApp is the most used communication channel for those reporting a worsened experience. Overall, the data suggests a strong engagement with social media among younger individuals, with a notable presence of negative sentiment regardless of usage time, and WhatsApp being a key communication tool for those with negative feedback.

DATASET USED

Dataset: Social Media Aspects

Dataset Tool: SQL (PostgreSQL)

SQL TECHNIQUES USED

- **SELECT and WHERE:** Choosing specific information you want to see.
- **COUNT, AVG, SUM, ROUND:** Getting totals, averages, and cleaner numbers.
- **GROUP BY and ORDER BY:** Putting similar things together and sorting them.
- **CASE WHEN:** Counting things only when a certain rule is true.
- **UNION ALL:** Putting the results of different searches into one list.
- **WITH:** Making complex searches easier by breaking them down into steps.
- **Calculated Columns:** Creating new pieces of information from existing data.
- **AS:** Giving clearer names to the columns and results you see.

```
create table social_media (
Age_group varchar,
occupation varchar,
place varchar,
spending varchar,
feedback varchar,
email_communication varchar,
Linkedin_communication varchar,
Whatsapp_communication varchar,
mostly_used varchar,
Impact varchar
);
```

I created a table named **social_media** in PostgreSQL with columns to store information such as age group, occupation, place, spending habits, feedback, communication preferences (like email, LinkedIn, and WhatsApp), the most used platform, and the impact of social media. After creating the table, I imported the dataset from a CSV file using the import options available in PostgreSQL (through tools like pgAdmin)

BASIC LEVEL

1. How many users are from each age group?

   QUERY :

   ```
   select age_group, count (*) as num_of_users from social_media
   group by age_group;
   ```

   OUTPUT :

	age_group character varying	num_of_users bigint
1	18-25	48
2	26-40	1

INSIGHTS :

The data shows that the majority of users fall in the 18–25 age group, with 48 users, while only 1 user belongs to the 26–40 age group. This indicates that younger individuals are much more active or represented in this dataset. It suggests that social media usage or engagement is significantly higher among the younger age group compared to the older one.

2. What is the most common occupation in the dataset?

QUERY :

```
select occupation, count(*) as num_of_job from social_media
group by occupation
order by num_of_job desc ;
```

OUTPUT :

	occupation character varying	num_of_job bigint
1	Student	44
2	Freelancer	2
3	Other	1
4	Employee	1
5	Unemployed	1

QUERY :

```
select occupation, count(*) as num_of_job from social_media
group by occupation
order by num_of_job desc
 limit 1 ;
```

OUTPUT :

	occupation character varying	num_of_job bigint
1	Student	44

INSIGHTS :

The most common occupation in the dataset is Student, with a total of 44 entries. This indicates that the majority of individuals represented in the data are currently pursuing education. It highlights that students form the core of the surveyed or analyzed group, making them the primary demographic for any social media trends, usage patterns, or marketing strategies derived from this dataset.

3. Count of users by their most used social media platform.

QUERY :

```
select mostly_used, count(*) as count_of_users from social_media
 group by mostly_used;
```

OUTPUT :

	mostly_used character varying	count_of_users bigint
1	Facebook	36
2	Others	4
3	LinkedIn	3
4	Instagram	6

INSIGHTS :

The data reveals that Facebook is the most commonly used social media platform among the users, with 36 out of 49 individuals (approximately 73%) preferring it. Instagram comes next with 6 users, followed by the "Others" category with 4 users, which likely includes less popular platforms. LinkedIn is the least used, with only 3 users indicating it as their most used platform. This shows a strong preference for Facebook, while other platforms have relatively low engagement among the users.

Intermediate Level

4. How does feedback (Improved, No Change, Worsened) vary across different age groups?

   QUERY :

   ```
   select distinct(feedback), age_group, count(*) as num_of_feedback from
   social_media

   group by age_group, feedback;
   ```

   OUTPUT :

	feedback character varying	age_group character varying	num_of_feedback bigint
1	No Change	18-25	17
2	Improved	18-25	15
3	Worsened	18-25	16
4	Worsened	26-40	1

   INSIGHTS :

   The younger group (18-25) showed a mix of opinions on the feedback, with "No Change" being slightly more common than "Worsened" or "Improved." In contrast, the small amount of feedback from the 26-40 age group was entirely negative ("Worsened"). This could point to a real difference in how these age groups perceive things, or it might just be because we have very little information from the older group. Getting more feedback from the 26-40 age range would help us understand their views better.

5. Which occupation spends the most time on social media on average?

   QUERY :

   ```
   select occupation,

   avg(

    case spending

     when 'Less than 1 hour' then 1

     when '1-2 hours' then 2

     when '2-3 hours' then 3

     when '3-5 hours' then 4

     when 'More than 5 hours' then 6

    else null

   end
   ```

```
) as avg_spending , count(*) as num_of_users from social_media
group by occupation
order by num_of_users desc
limit 1;
```

OUTPUT :

	occupation character varying	avg_spending numeric	num_of_users bigint
1	Student	3.8181818181818182	44

INSIGHTS :

"Student" is the occupation with the highest number of users recorded (44). Their average spending, calculated from the time they spend (though represented numerically here), is approximately 3.82. While this tells us about the average spending for the occupation with the most users, it doesn't directly answer which occupation *spends the most time* on average across all occupations in the dataset. To find that, we'd need to see the average spending for all the different occupations present in the data, not just the one with the most users.

6. What is the distribution of social media usage (spending time) among people in urban vs suburban areas?

QUERY :

```
select place,
sum(
 case spending
  when 'Less than 1 hour' then 1
  when '1-2 hours' then 2
  when '2-3 hours' then 3
  when '3-5 hours' then 4
  when 'More than 5 hours' then 6
 else null
end
) as total_spending, count(*) as num_of_users from social_media
group by place
order by total_spending ;
```

OUTPUT :

	place character varying	total_spending bigint	num_of_users bigint
1	Suburban Area	28	7
2	Rural Area	67	17
3	Urban Area	94	25

QUERY :

```
select place,
sum(
 case spending
  when 'Less than 1 hour' then 1
  when '1-2 hours' then 2
  when '2-3 hours' then 3
  when '3-5 hours' then 4
  when 'More than 5 hours' then 6
 else null
end
) as total_spending, count(*) as num_of_users from social_media
where place = 'Urban Area' or place = 'Suburban Area'
group by place
order by total_spending ;
```

OUTPUT :

	place character varying	total_spending bigint	num_of_users bigint
1	Suburban Area	28	7
2	Urban Area	94	25

INSIGHTS :

we can see that users in "Urban Areas" have a much higher total spending (94) compared to those in "Suburban Areas" (28). This total spending is calculated based on the time users reported spending on social media. Additionally, there are significantly more users from "Urban Areas" (25) represented in this data compared to "Suburban Areas" (7). This suggests that either people in urban areas tend to spend more time on social media, or there are simply more social media users from urban areas in this particular dataset, leading to a higher overall total spending

Advanced Level

7. Which communication channel (Email, LinkedIn, WhatsApp) is most frequently used by people who reported a 'Worsened' impact?

QUERY :

```
select 'whatsapp_communication' as channel, count(*) as count from social_media
where feedback = 'Worsened' and whatsapp_communication = 'Daily'
union
select 'email_communication' as channel, count(*) as count from social_media
where feedback = 'Worsened' and email_communication = 'Daily'
union
select 'linkedin_communication' as channel, count(*) as count from social_media
where feedback = 'Worsened' and linkedin_communication = 'actively'
order by count desc;
```

OUTPUT :

	channel text	count bigint
1	whatsapp_communica...	10
2	email_communication	2
3	linkedin_communication	0

INSIGHTS :

The data shows that 'whatsapp_communication' is the most frequently used channel among them, with a count of 10. 'email_communication' is used by 2 people in this group, while 'linkedin_communication' shows 0 users who reported a 'Worsened' impact and actively use LinkedIn for communication. This suggests that WhatsApp is the dominant communication channel for individuals in this dataset who feel their experience has worsened.

8. Correlation between the most used platform and perceived impact—does using a certain platform lead to more 'Worsened' feedback?

QUERY :

```
select mostly_used, feedback, count(*) as user_count from social_media
group by mostly_used, feedback
order by user_count desc;
```

OUTPUT :

	mostly_used character varying	feedback character varying	user_count bigint
1	Facebook	Worsened	15
2	Facebook	Improved	11
3	Facebook	No Change	10
4	Others	No Change	3
5	Instagram	No Change	2
6	Instagram	Worsened	2
7	Instagram	Improved	2
8	LinkedIn	No Change	2
9	LinkedIn	Improved	1
10	Others	Improved	1

INSIGHTS :

Looking at which platform people use most and how they feel, Facebook has the most people saying things got "Worsened" (15). But Facebook also has the most people saying things "Improved" (11) or stayed the "Same" (10). This might just be because lots of people use Facebook.

For other platforms like Instagram, the number of people saying "Worsened" is the same as those saying "Improved" or "Same." To really know if one platform makes people feel worse, we'd need to look at the percentage of bad feedback for each platform. But just counting, Facebook has the most people who feel things got worse.

9. Is there a pattern in social media impact (positive/negative) based on time spent on social media?

QUERY :

```
select impact, spending, count(*) as users from social_media
group by spending, impact
order by spending, users desc;
```

OUTPUT :

	impact character varying	spending character varying	users bigint
1	Yes, negatively	1-2 hours	7
2	Not sure	1-2 hours	4
3	No, it has no effect	1-2 hours	3
4	Yes, positively	1-2 hours	2
5	Yes, negatively	3-5 hours	8
6	Not sure	3-5 hours	7
7	Yes, positively	3-5 hours	3
8	Yes, negatively	Less than 1 hour	1
9	Yes, negatively	More than 5 hours	7
10	Yes, positively	More than 5 hours	4
11	Not sure	More than 5 hours	3

INSIGHTS :

People who spend 1-2 hours and 3-5 hours on social media most often feel a negative impact. Even those who spend more than 5 hours also report negative impacts a lot, though some in that group feel positive or aren't sure. So, feeling negative about social media seems common no matter how much time people spend on it, but it's not the only feeling.

10. Which age group and occupation combination is most likely to report 'Improved' feedback?

QUERY :

```
select occupation, age_group, feedback, count(*) as users from social_media
where feedback = 'Improved'
group by occupation, age_group, feedback
order by users desc;
```

OUTPUT :

	occupation character varying	age_group character varying	feedback character varying	users bigint
1	Student	18-25	Improved	14
2	Freelancer	18-25	Improved	1

INSIGHTS :

The data for 'Improved' feedback, the combination of 'Student' in the '18-25' age group has the highest number of users reporting this positive change (14 users). Other combinations, like 'Freelancer' in the '18-25' age group, have much fewer users reporting 'Improved' feedback (only 1). Therefore, based on this data, students in the 18-25 age range are the most likely to report an 'Improved' experience.