Assignment 1 GitHub

October 24, 2025

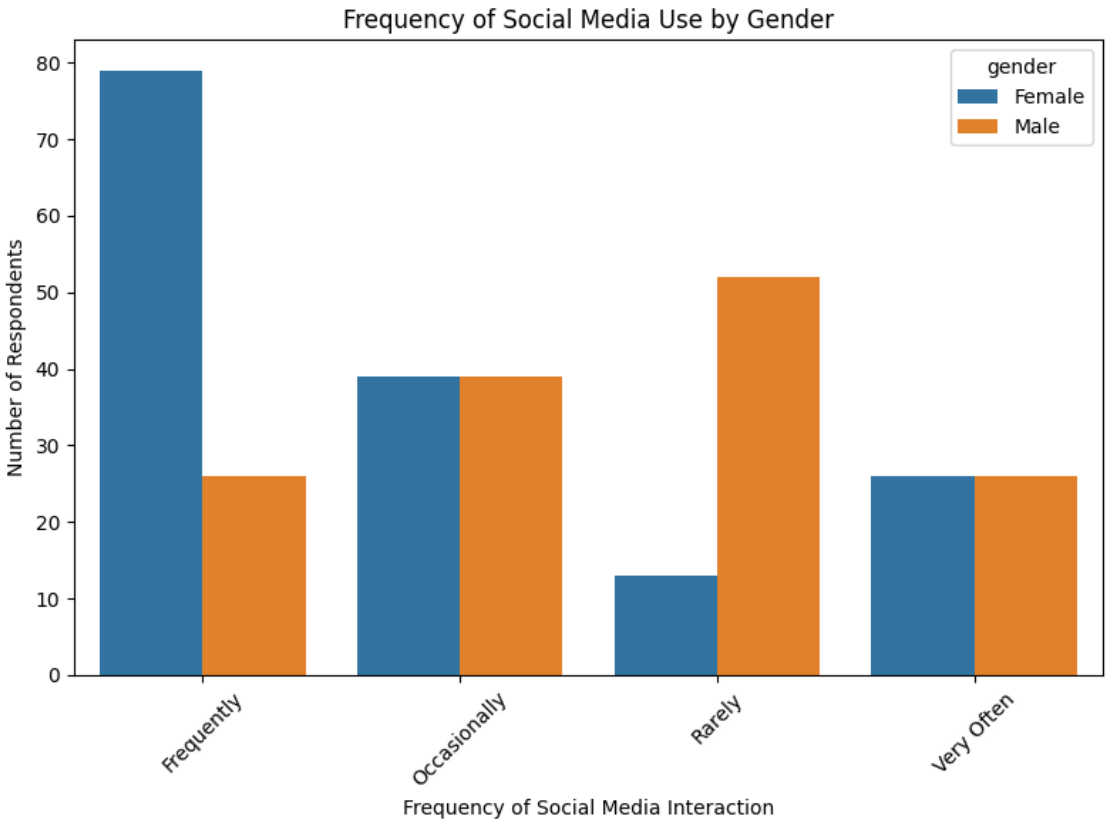
0.1 Introduction

I used the dataset titled "Mental Health and Social Media" from Kaggle. It is a a dataset containing self-reported information from individuals regarding their frequency of social media usage, gender, age, perceived mental health status and a rating for impact of social media on mental health.

0.2 Research Questions

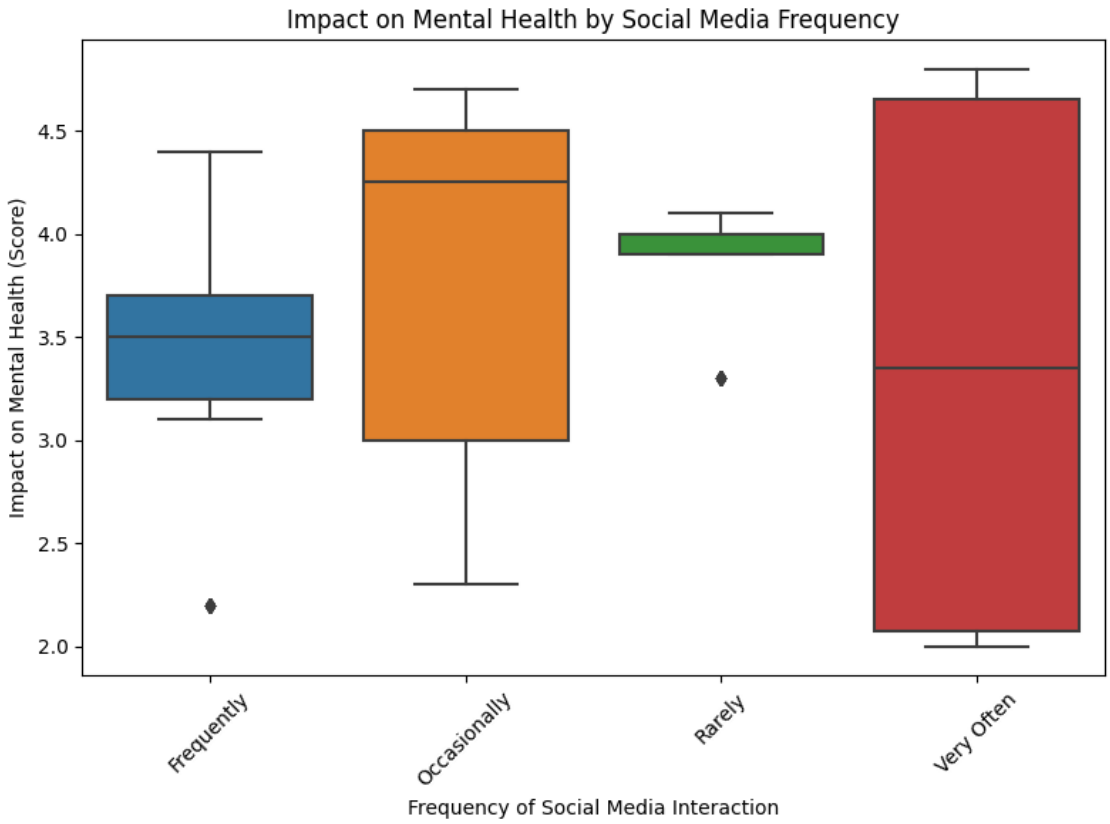
1. How does the frequency of social media interaction vary across genders?
2. Is there a relationship between frequency of use of social media and perceived mental health?
3. How does the impact score correlate with social media usage and frequency?

[41]:
```
# Histogram
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

file_path = "Mental Health and Social Media.xlsx"
df = pd.read_excel(file_path)

# Clean column names
df.columns = [col.strip().replace(" ", "_").lower() for col in df.columns]

# Graph Count of Social Media Use Frequency by Gender
plt.figure(figsize=(8,6))
sns.countplot(data=df, x = 'frequency_of_social_media_interaction',␣
 ↪hue='gender', order=df['frequency_of_social_media_interaction'].
 ↪value_counts().index)
plt.xticks(rotation=45)
plt.title('Frequency of Social Media Use by Gender')
plt.xlabel('Frequency of Social Media Interaction')
plt.ylabel('Number of Respondents')
plt.tight_layout()
plt.show()
```



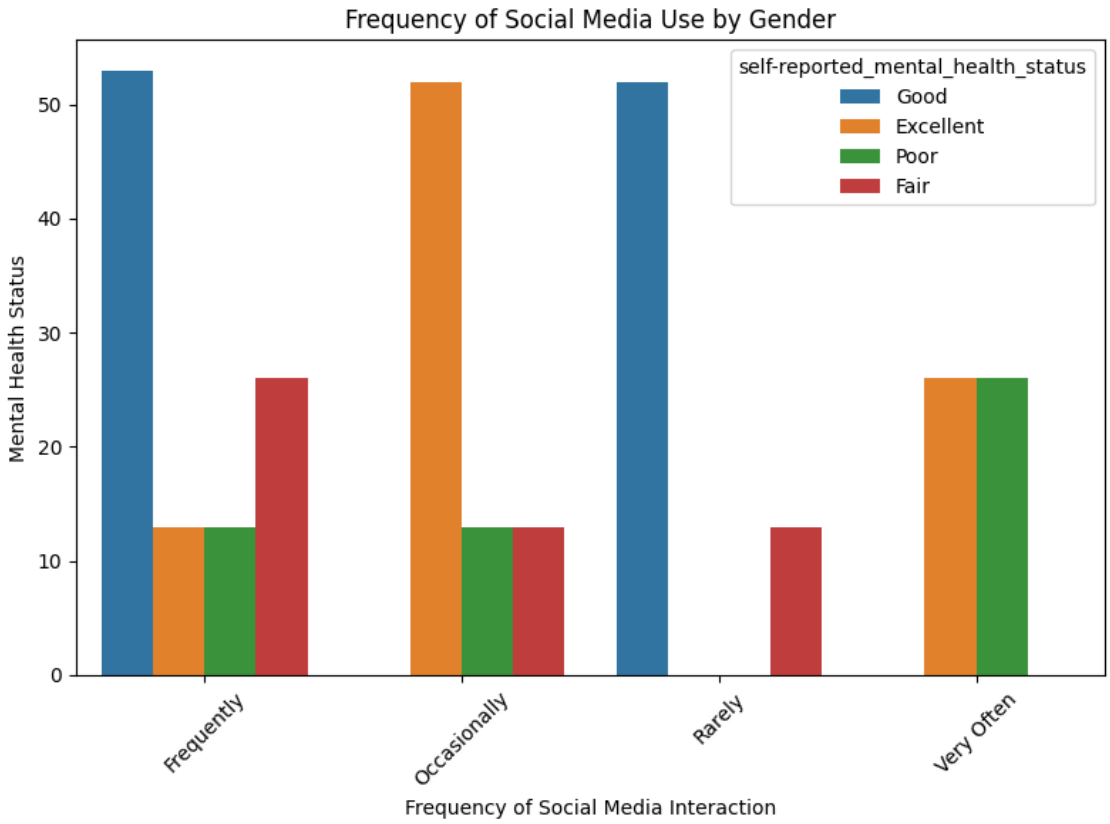
Women in this dataset report using social media more frequently than men. 'Frequently' and 'Occasionally' are the most common responses among female participants. Male respondents are more likely to report lower usage.

[44]:
```
# Boxplot
plt.figure(figsize=(8,6))
sns.boxplot(data=df, x = 'frequency_of_social_media_interaction', y =␣
 ↪'impact_on_mental_health_(score)',␣
 ↪order=df['frequency_of_social_media_interaction'].value_counts().index)
plt.xticks(rotation=45)
plt.title('Impact on Mental Health by Social Media Frequency')
plt.xlabel('Frequency of Social Media Interaction')
plt.ylabel('Impact on Mental Health (Score)')
plt.tight_layout()
plt.show()
```



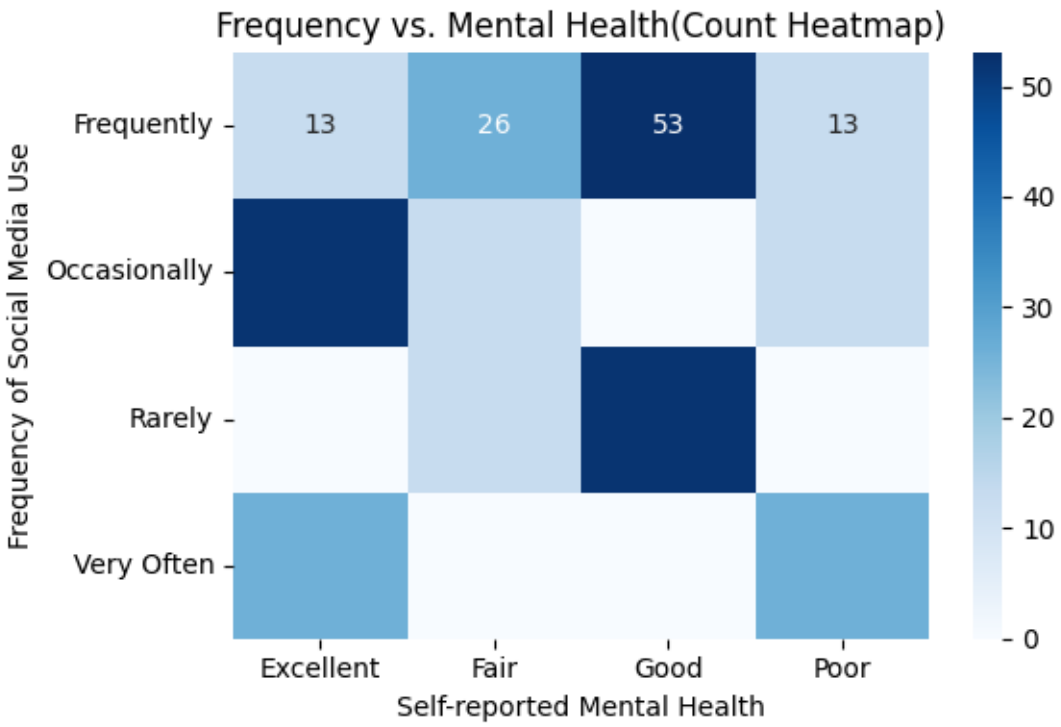
Those who use social media "Very Often" tend to report lower impact scores suggesting a more negative impact on mental health. Those who report using social media "Occasionally" or "Rarely" have higher impact scores, implying a lesser impact or positive impact. This suggests tha higher frequency of social media use may be linked to poorer mental health.

[47]:
```
# Bivariate Plot
plt.figure(figsize=(8,6))
sns.countplot(data=df, x = 'frequency_of_social_media_interaction',␣
 ↪hue='self-reported_mental_health_status',␣
 ↪order=df['frequency_of_social_media_interaction'].value_counts().index)
plt.xticks(rotation=45)
plt.title('Frequency of Social Media Use by Gender')
plt.xlabel('Frequency of Social Media Interaction')
plt.ylabel('Mental Health Status')
plt.tight_layout()
plt.show()
```



Respondent who use social media "Very Often" report a higher proportion of "Poor" or "Fair" mental health. Those using social media "Rarely" or "Occasionally" are more likely to report "Good" or "Excellent" mental health. This pattern reinforces the possible negative replationship between frequent social media use and mental well-being.

[50]:
```
# Heatmap
freq_order = ['Rarely', 'Occasionally', 'Frequently', 'Very Often']
mh_order = ['Poor', 'Fair', 'Good', 'Excellent']
pivot_counts = pd.crosstab(df['frequency_of_social_media_interaction'],␣
 ↪df['self-reported_mental_health_status'])
plt.figure(figsize=(6,4))
sns.heatmap(pivot_counts, annot=True, cmap='Blues', fmt='d')
plt.title('Frequency vs. Mental Health(Count Heatmap)')
plt.ylabel('Frequency of Social Media Use')
plt.xlabel('Self-reported Mental Health')
plt.tight_layout()
plt.show()
```



Respondents who had occasional use had the most positive (excellent mental health). Those with Frequent/Rare use had good mental health. Those with very often use were split between excellent and poor mental health. This suggests balanced social media use is best.

1 Conclusion

After analyzing the dataset on 'Mental Health and Social Media', I found: Women tend to use social media more frequently than men which is shown in the histogram. Occasional users rpeort the highest perceived impact on mental health whereas frequently and very frequent users report lower impact scores. Self-reported mental health tends to be worse among those who use social media most often as shown in the bivariate plot. The heatmap suggests balanced or occasional use of social media is best for the highest self-reported mental health.

Overall, my findings suggest that heavy social media use may be associated with decreased mental well- being. While correlation doesn't imply causation, the patterns indicate that reducing excessive usage may offer mental health benefits.

[ ]: