MENTAL HEALTH ANALYSIS

Using pandas

Introduction to dataset:

The dataset used in this analysis is Mental Health dataset , sourced from kaggle. A diverse dataset of 1000 individuals across professions, countries, and lifestyles. This dataset representing a wide range of ages, genders, occupations, and countries. It aims to shed light on the various factors affecting mental health, offering valuable insights into stress levels, sleep patterns, work-life balance, and physical activity.

Columns:

- **Age** : The age of the individual
- **Gender :** The gender of the individual (e.g., Male, Female, Non-Binary).
- **Occupation** : The profession or job title of the individual (e.g., Engineering, Healthcare, Sales, etc.).
- **Country** : The country of residence of the individual (e.g.USA, UK, etc.).
- **Mental Health Condition** : The mental health status or condition of the individual.
- **Severity** : The severity level of the individual's mental health condition.
- **Consultation History** : Indicates whether the individual has consulted a mental health professional.
- **Stress Level :** Reported stress level of the individual.
- **Sleep Hours :** Average hours of sleep per night of the individual.
- **Work Hours :** The total number of hours worked by the individual.
- **Physical Activity Hours :** The number of hours the individual exercise.

Aim:

The aim of this dataset is to explore the relationships between lifestyle factors, demographics, and mental health conditions. It seeks to understand how stress levels, sleep hours, work hours, physical activity, and consultation history impact mental well-being, identifying patterns that can guide strategies to improve mental health.

Objective:

The objective is to provide insights for mental health research and intervention, highlighting factors like occupation, physical activity, and sleep in managing stress and mental health. This data can help develop targeted strategies for better mental health support and healthier work-life balance.

Overview of Data :

The Mental Health Dataset, sourced from Kaggle, contains 1,000 rows and 12 columns. The columns include User ID , Age , Gender, Occupation, Country, Mental Health Condition, Severity, Consultation History, Stress Level, Sleep Hours, Work Hour, Physical Activity Hours. It provides insights into the relationship between demographics, lifestyle factors, and mental health conditions.

User_ID	int64
Age	int64
Gender	object
Occupation	object
Country	object
Mental_Health_Condition	object
Severity	object
Consultation_History	object
Stress_Level	object
Sleep_Hours	float64
Work_Hours	int64
Physical_Activity_Hours	int64

Steps involved in analysis:

1. **Data Cleaning and Preprocessing :**

   - A total of 501 missing values were found in the Severity column, which were filled based on the values of Stress Level , Consultation History and Physical Activity Hours.
   - Specifically, missing Severity values were filled by assigning 'High' for individuals with high stress, 'Medium' for those with medium stress and low physical activity, or those who had a history of consultation, and 'Low' otherwise.

- User ID column was dropped.
- Additionally, the dataset was examined for duplicates and outliers, but no duplicates or significant outliers were found, as confirmed through the use of box plots.

- *df.groupby('Stress_Level')['Severity'].value_counts()*
- *df.groupby('Consultation_History')['Severity'].value_counts()*
- *df.groupby('Physical_Activity_Hours')['Severity'].value_counts()*

- *def severity(row):*

```
        if pd.isna(row['Severity']):
            if row['Stress_Level'] == 'High':
                return 'High'
            elif row['Stress_Level'] == 'Medium' and
row['Physical_Activity_Hours'] < 2:
                return 'Medium'
            elif row['Consultation_History'] == 'Yes':
                return 'Medium'
            else:
                return 'Low'
        else:
            return row['Severity']  # Retain existing values
    df['Severity'] = df.apply(func=severity,axis=1)
```

- *df.duplicated().sum()*

- *df.drop(columns='User_ID',inplace=True)*

- *for col in df:*

    if df[col].dtype=='int' or df[col].dtype=='float':

        print(col)

        print()

        plt.figure(figsize=(7,4))

        plt.boxplot(df[col])

        plt.title(col)

        plt.show()

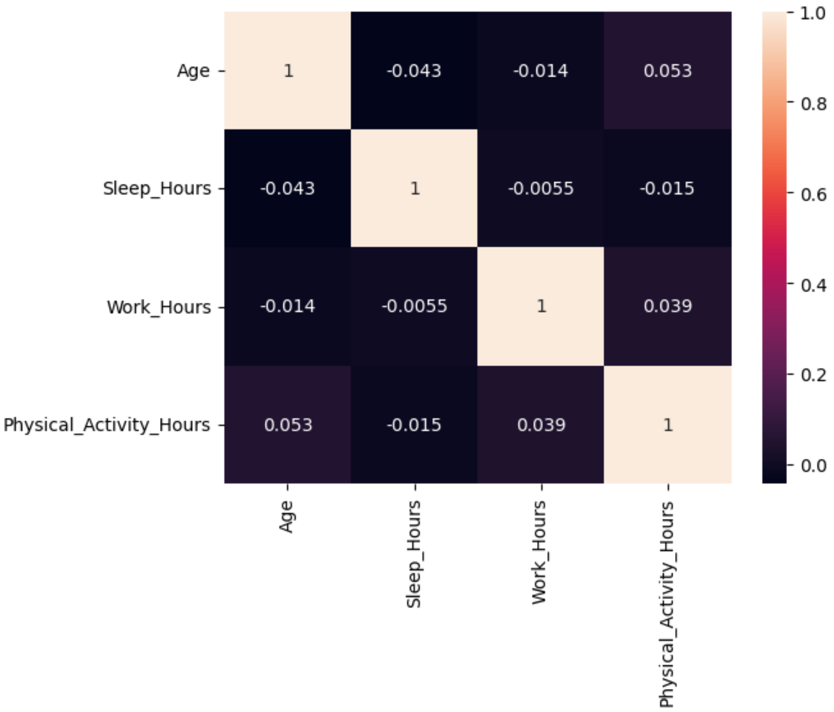
2. Feature Engineering :

- For feature engineering, the Age column was categorized into three age groups. Young Adults (0-30 years), Middle Aged (31-55 years), and Senior (56-80 years).

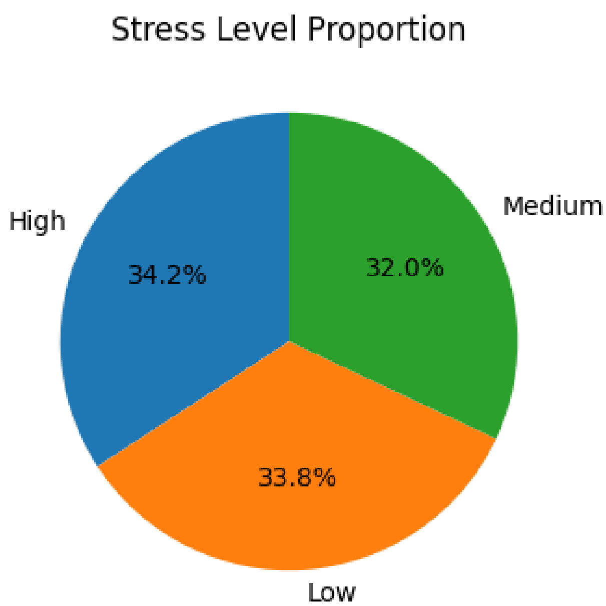
- *df['Age_cat'] = pd.cut(x=df['Age'], bins=[0,30,55,80], labels=['Young Adults', 'Middle Aged', 'Senior'], right=True)*

3. Analysis :

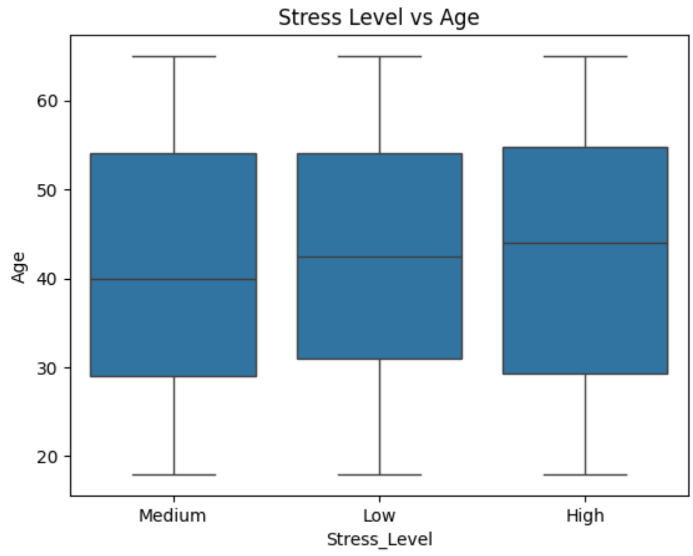
- **Correlation :** These variables are relatively independent of each other. Age doesn't seem to play a significant role in influencing sleep, work, or physical activity habits. There are only very weak correlations between the other variables, indicating that they don't have a strong impact on each other.



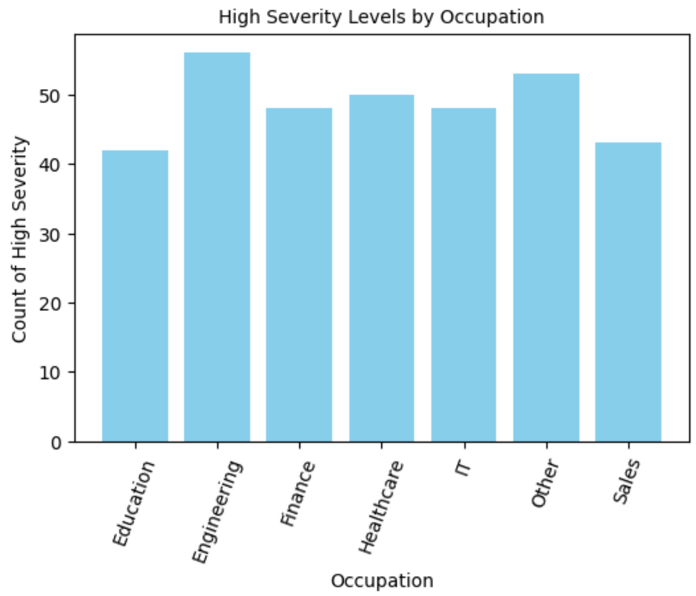
- **Stress Level Proportion :** The distribution of stress levels among individuals, revealing that a significant proportion (34.2%) experience high levels of stress. This is closely followed by those experiencing low stress (33.8%) and medium stress (32.0%).



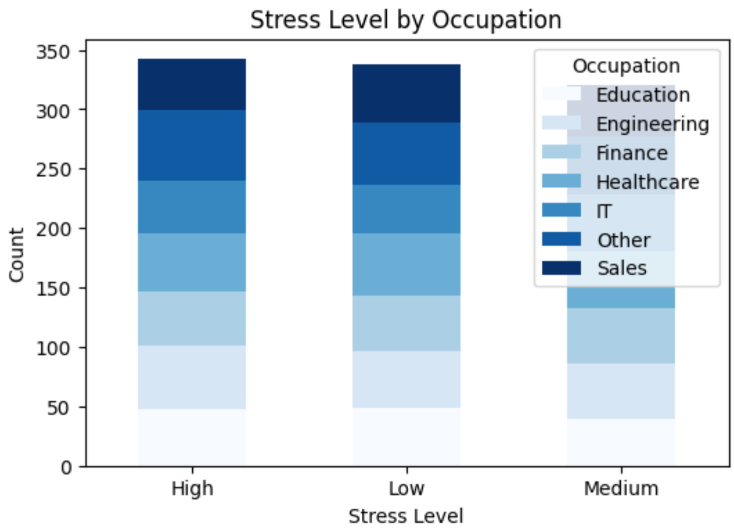
- **Stress level v/s Age :** The spread of ages within each stress level is relatively similar, suggesting that stress levels are not solely determined by age. Other factors likely contribute to an individual's stress experience.



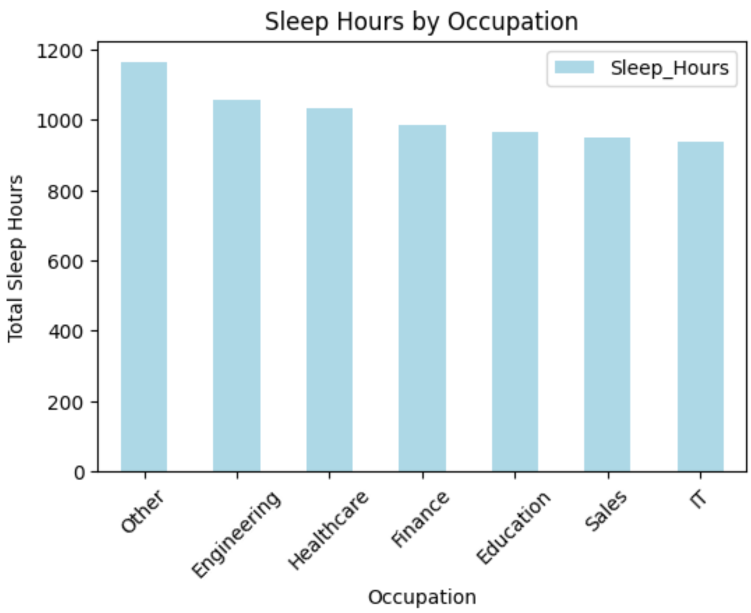
- **Severity v/s Occupation :** Engineering and Other occupations have the highest number of reported high severity levels, followed by Sales and IT. This suggests that certain occupations may be associated with a higher risk of experiencing high severity incidents.



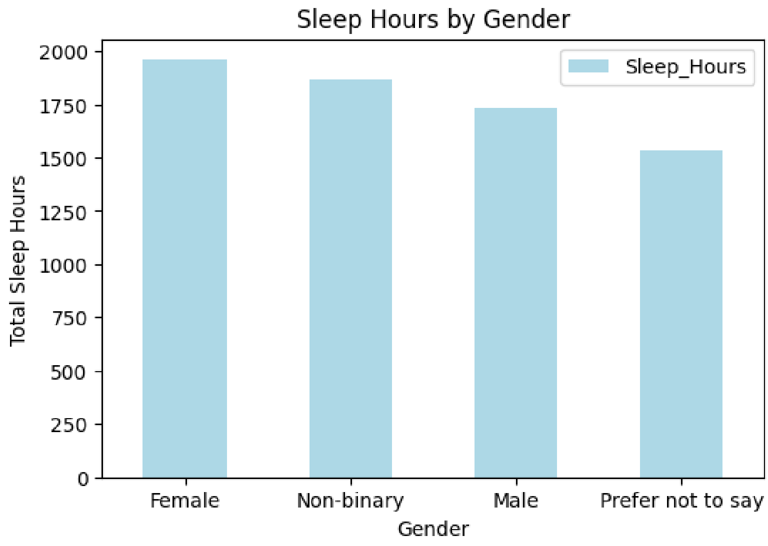
- **Stress v/s Occupation :** It reveals that "Other" occupation has the highest count of individuals experiencing high stress, while "Education" has the highest count for low stress. This suggests that certain occupations may be associated with distinct stress profiles.



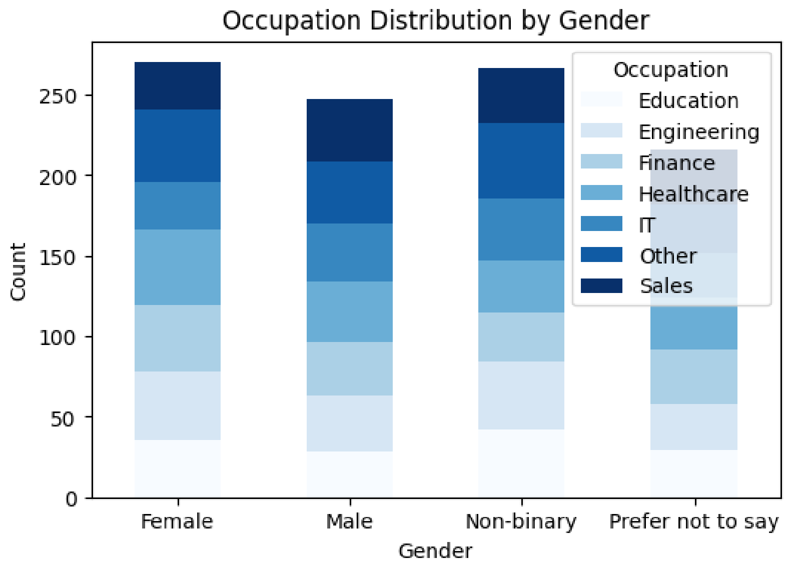
- **Sleep Hour v/s Occupation :** It shows that individuals in "Other" occupations tend to sleep the most, followed by those in Engineering and Healthcare.



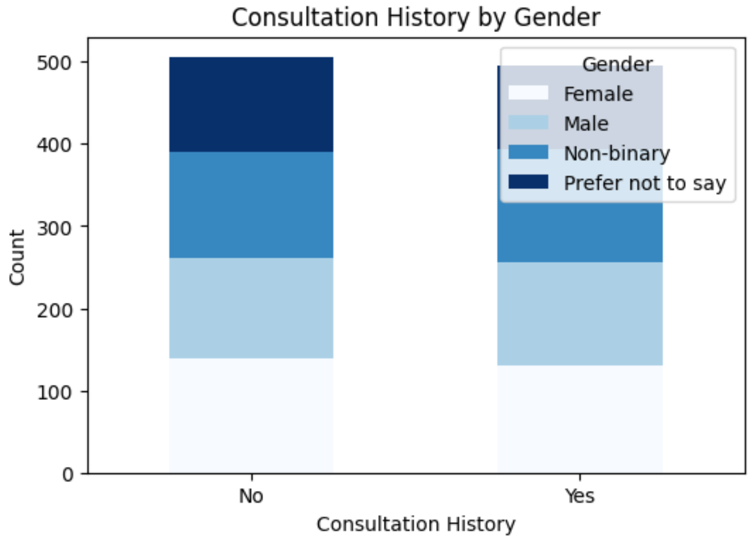
- **Sleep Hour v/s Gender :** It reveals that females tend to sleep the most, followed by non-binary individuals. Males and those who prefer not to say report the least amount of sleep.



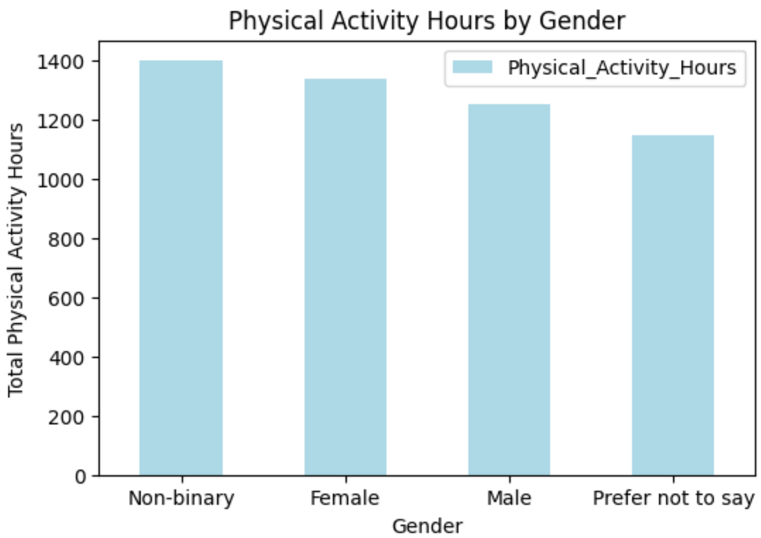
- **Occupation v/s Gender :** It reveals that "Other" is the most common occupation among all genders. Interestingly, "Education" and "Sales" appear to be less common among non-binary and those who prefer not to say.



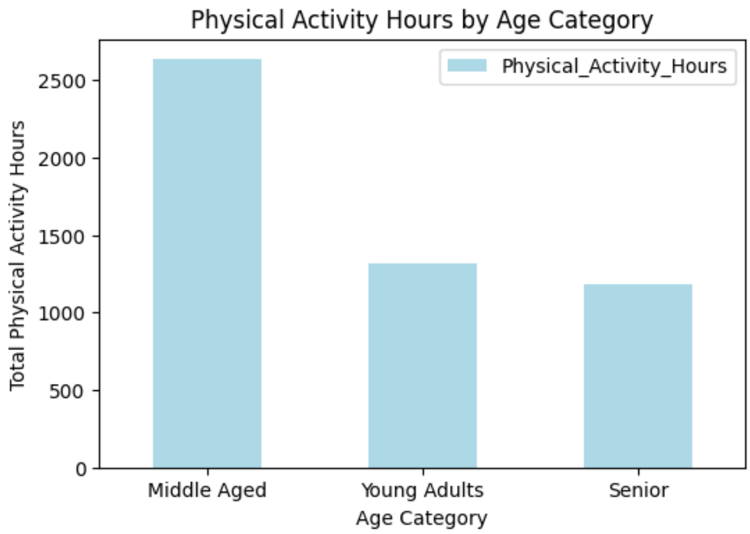
- **Consulation History v/s Gender :** It reveals that a higher proportion of females and those who prefer not to say have a consultation history compared to males and non-binary individuals.



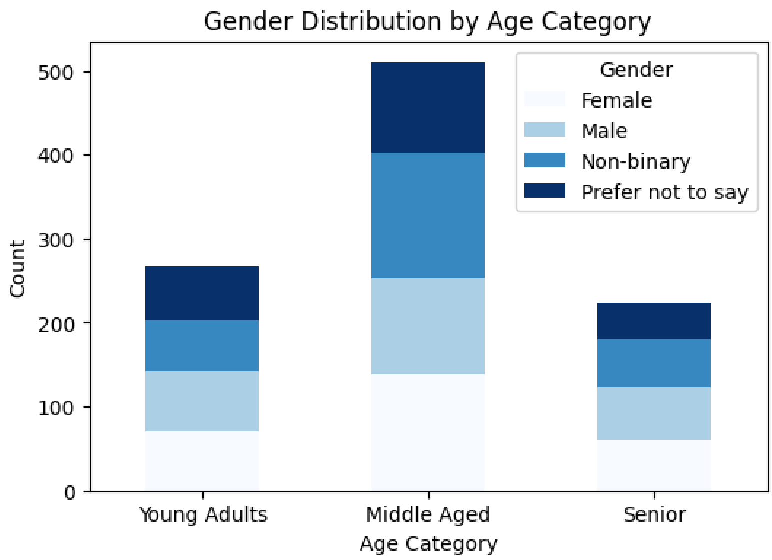
- **Physical Activity Hour v/s Gender :** Non-binary individuals tend to engage in the most physical activity, followed by females. Males and those who prefer not to say report the least amount of physical activity.



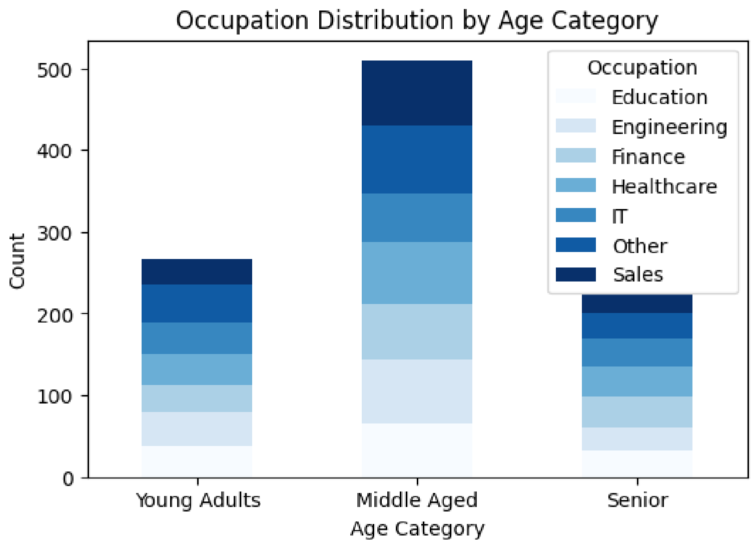
- **Physical Activity Hour v/s Age Category :** It reveals that middle-aged individuals tend to engage in the most physical activity, followed by young adults. Seniors report the least amount of physical activity.



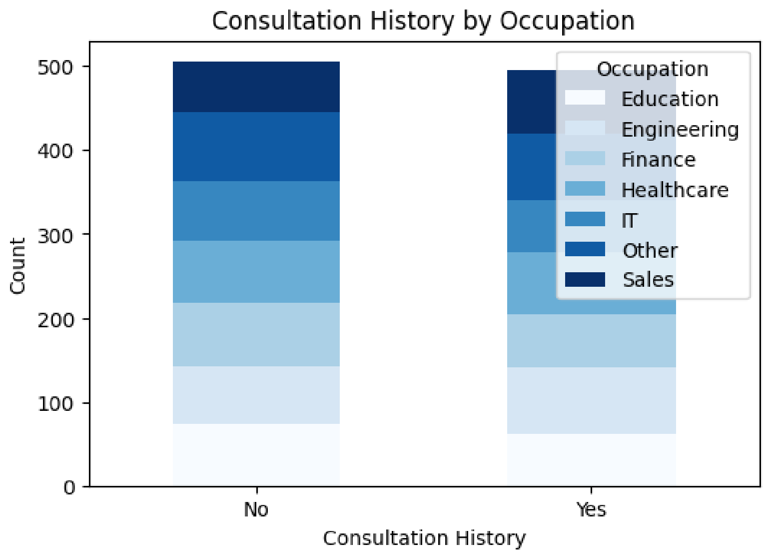
- **Gender v/s Age Category :** It reveals that females are the most common gender in all age categories. Interestingly, the proportion of non-binary individuals seems to be higher among young adults compared to middle-aged and senior groups.



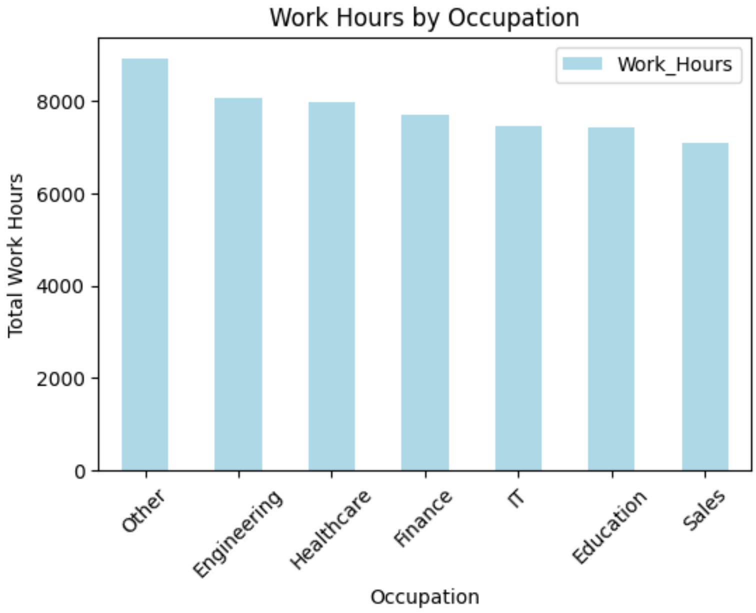
- **Occupation v/s Age Category :** It reveals that "Other" is the most common occupation in all age groups. Interestingly, "Education" and "Sales" appear to be less common among seniors compared to young adults and middle-aged individuals.



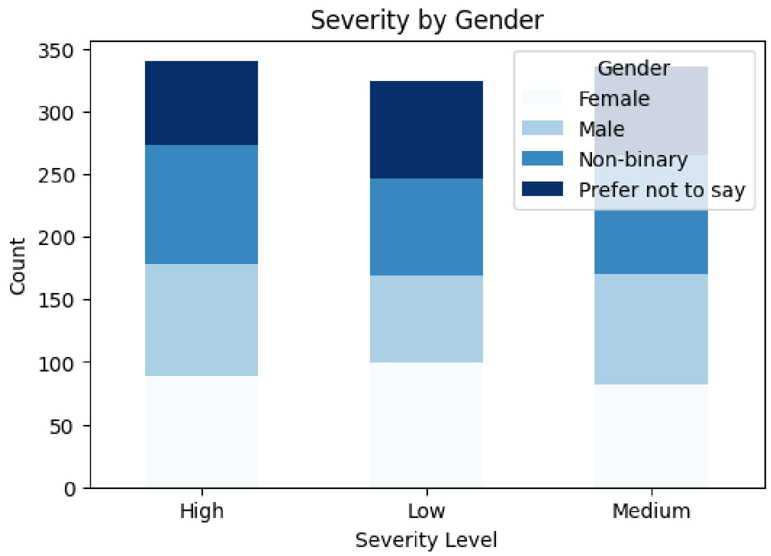
- **Consultation History v/s Occupation :** It reveals that a higher proportion of individuals in "Sales" and "Other" occupations have a consultation history compared to those in "Education" and "Healthcare".



- **Work Hour v/s Occupation :** It reveals that individuals in "Other" occupations tend to work the most, followed by those in Engineering. On the other hand, people in Sales report the least amount of work hours.



- **Severity v/s Gender :** It reveals that females report the highest number of high and low severity cases, while males report the highest number of medium severity cases.



Key Insights :

Gender and Health

- Females tend to sleep more and engage in more physical activity compared to males.
- Females and those preferring not to disclose gender are more likely to seek healthcare consultations.
- Gender differences exist in the severity of reported incidents.

Age and Health

- Middle-aged individuals are more physically active than younger and older adults.
- Gender distribution varies across age groups.

Occupation and Health

- Certain occupations, like "Other" and "Engineering," are associated with longer work hours and higher stress levels.
- "Sales" and "Other" occupations have higher rates of healthcare consultations.

Country and Healthcare Utilization

- Individuals from the US, UK, and Canada are more likely to seek healthcare consultations compared to those from other countries.

Conclusion :

The analysis of the provided data reveals distinct patterns in health-related behaviors across different demographic groups. Gender, age, occupation, and country significantly influence factors like sleep patterns, physical activity levels, healthcare utilization, and stress levels.

Promote Mental Health Awareness:

- Conduct regular campaigns to educate people about mental health conditions and encourage help-seeking behaviors.

Prioritize Sleep and Physical Activity:

- Encourage adequate sleep and regular physical activity as essential components of mental well-being.

Improve Access to Healthcare:

- Increase accessibility to mental health services, especially in underserved areas.

Tailored Interventions:

- Develop targeted interventions based on the specific needs of different demographic groups.
- Consider factors like gender, age, occupation, and cultural background when designing interventions.

Workplace Wellness Programs:

- Implement workplace programs to promote mental health, such as stress management workshops, mindfulness training, and employee assistance programs.