Analyzing Factors Affecting OCD Severity and Medication Trends

INTRODUCTION:

Obsessive-Compulsive Disorder (OCD) is a chronic mental health condition characterized by recurring obsessions and compulsions. The dataset contains demographic, diagnostic, and treatment-related information about OCD patients. The goal is to analyze this dataset to understand the factors affecting OCD severity, identify patterns in medication usage, and address missing information to improve data completeness for better decision-making.

Objective:

To identify patterns, correlations, and potential predictors of OCD severity using patient demographics, symptoms, and treatment data. Additionally, explore how medication use varies across different patient groups.

Problem Statement:

What factors influence OCD severity (measured by Y-BOCS scores)?

Are there any patterns in medication usage based on demographics, symptoms, or OCD severity?

CODE:

Step 1: Import Libraries and read the data

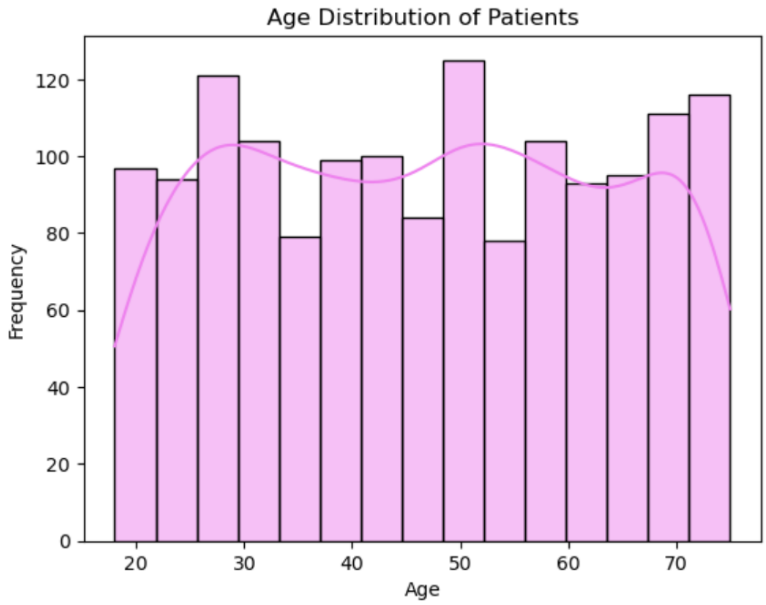
```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

data = pd.read_csv("OCD_Patient_Dataset.csv")

data

data.columns

data.shape

data.describe()

data.info()
```

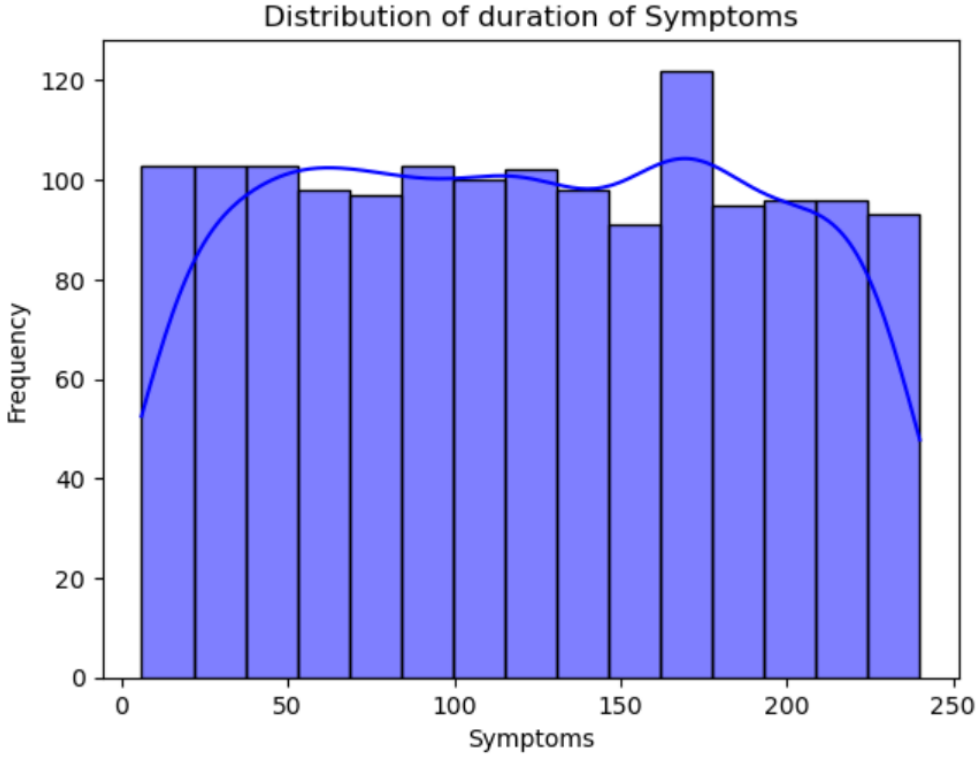
Step2: Handling missing values

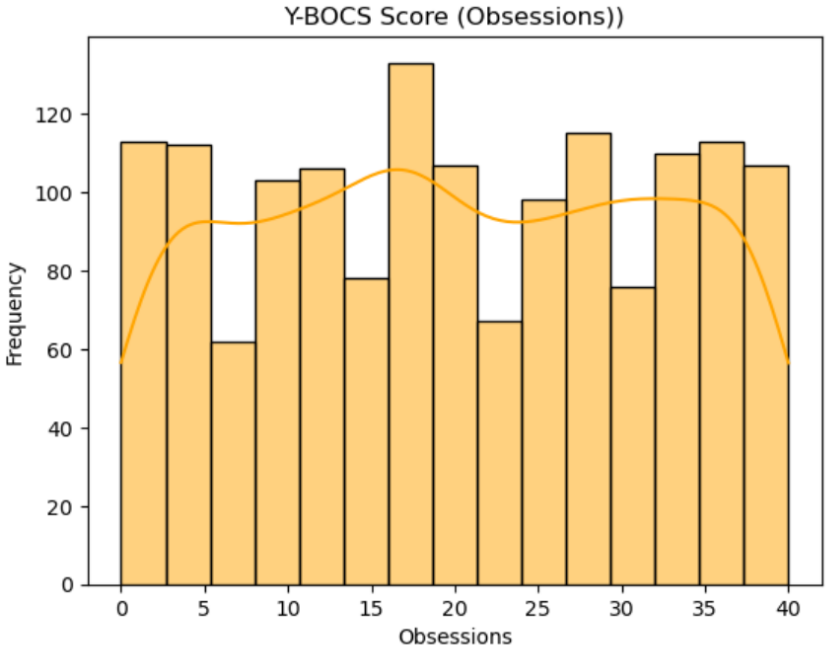
```
data.isnull().sum()
```

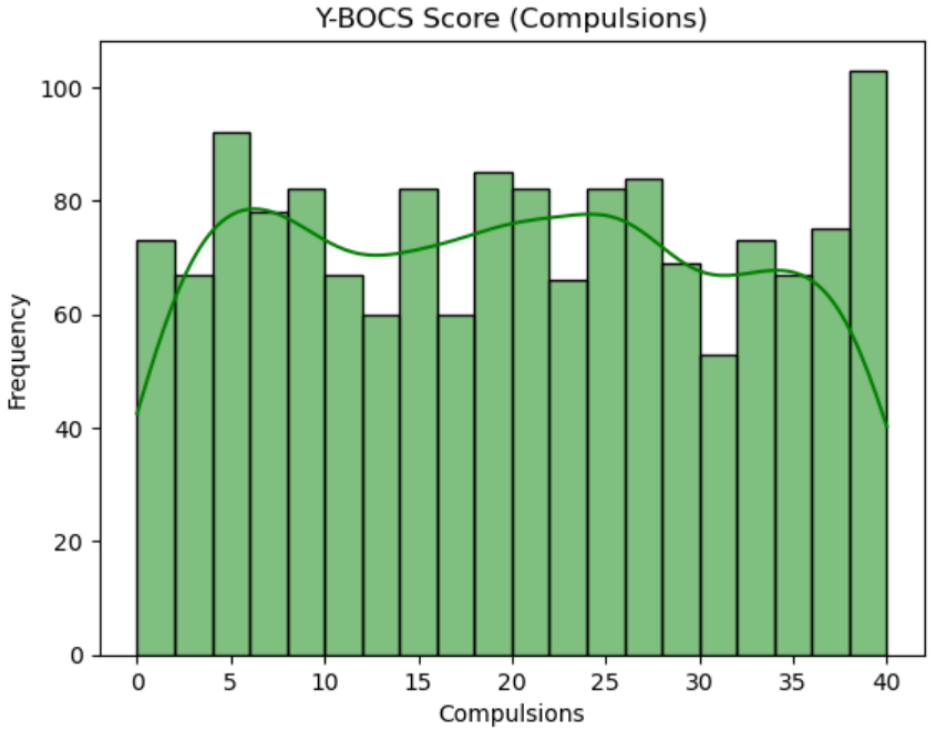
```
# Separate rows with missing values in specific columns
missing_values_rows = data[data[['Previous Diagnoses', 'Medications']].isna().any(axis=1)]
non_missing_values_rows = data.dropna(subset=['Previous Diagnoses', 'Medications'])

# Save missing rows separately (if needed)
missing_values_rows.to_csv('missing_values_rows.csv', index=False)
non_missing_values_rows.to_csv('non_missing_values_rows.csv', index=False)

# Display the counts
print(f"Rows with missing values: {missing_values_rows.shape[0]}")
print(f"Rows without missing values: {non_missing_values_rows.shape[0]}")
#Handling missing values
#The dataset dosent  seem much like a pattern , let us take it as unknown
data['Previous Diagnoses'].fillna('Unknown', inplace=True)
data['Medications'].fillna('Unknown', inplace = True)

data.head()
Step 3: Histogram
sns.histplot(data['Age'], bins=15, kde=True, color = 'violet')
plt.title('Age Distribution of Patients')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
sns.histplot(data['Duration of Symptoms (months)'], bins=15, kde=True, color = 'blue')
plt.title('Distribution of duration of Symptoms')
plt.xlabel('Symptoms')
plt.ylabel('Frequency')
plt.show()
sns.histplot(data['Y-BOCS Score (Obsessions)'], bins=15, kde=True, color = 'orange')
plt.title('Y-BOCS Score (Obsessions))')
plt.xlabel('Obsessions')
```

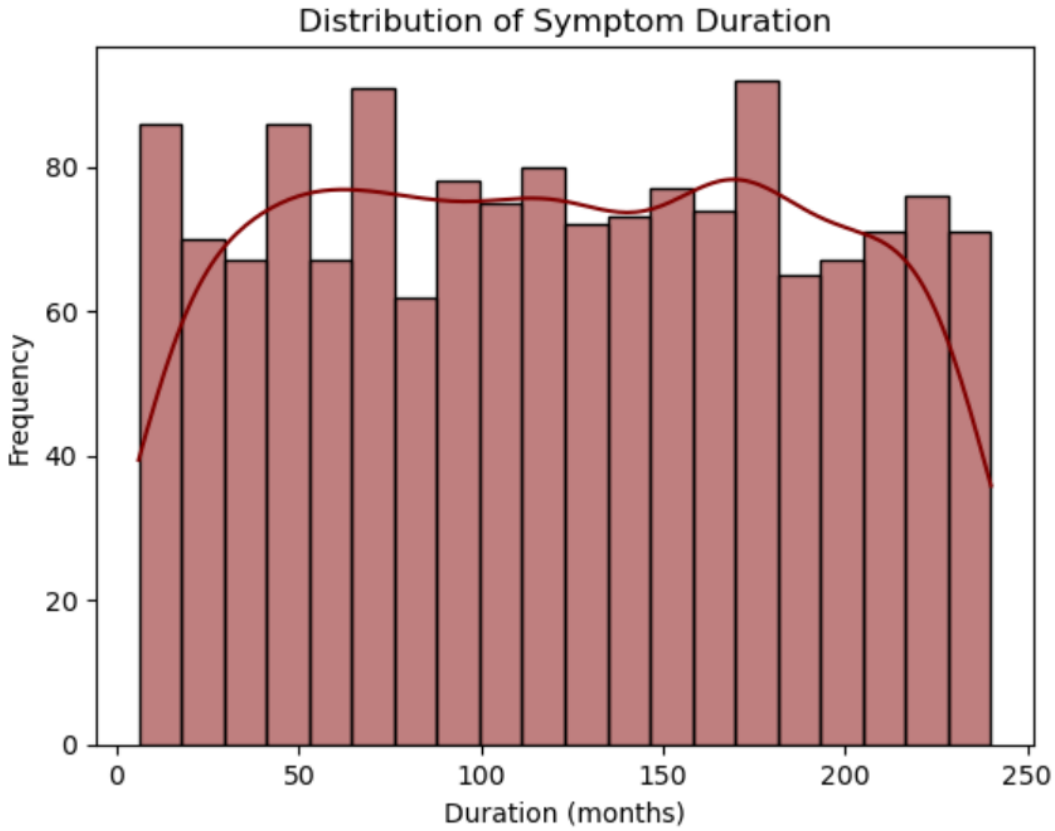
```
plt.ylabel('Frequency')
plt.show()
sns.histplot(data['Y-BOCS Score (Compulsions)'], bins=20, kde=True, color = 'green')
plt.title('Y-BOCS Score (Compulsions)')
plt.xlabel('Compulsions')
plt.ylabel('Frequency')
plt.show()
filtered_data = data[data['Medications'].str.lower() != 'none']

# Plot histogram for the 'Age' column
filtered_data['Age'].hist(bins=15, edgecolor='black', color='skyblue')
plt.title('Age Distribution for Patients on Medications')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
#Gender distribution
sns.countplot(x='Gender', data=data)
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
#Distribution of symptom duration
sns.histplot(data['Duration of Symptoms (months)'], bins=20, kde=True, color = "maroon")
plt.title('Distribution of Symptom Duration')
plt.xlabel('Duration (months)')
plt.ylabel('Frequency')
plt.show()
```

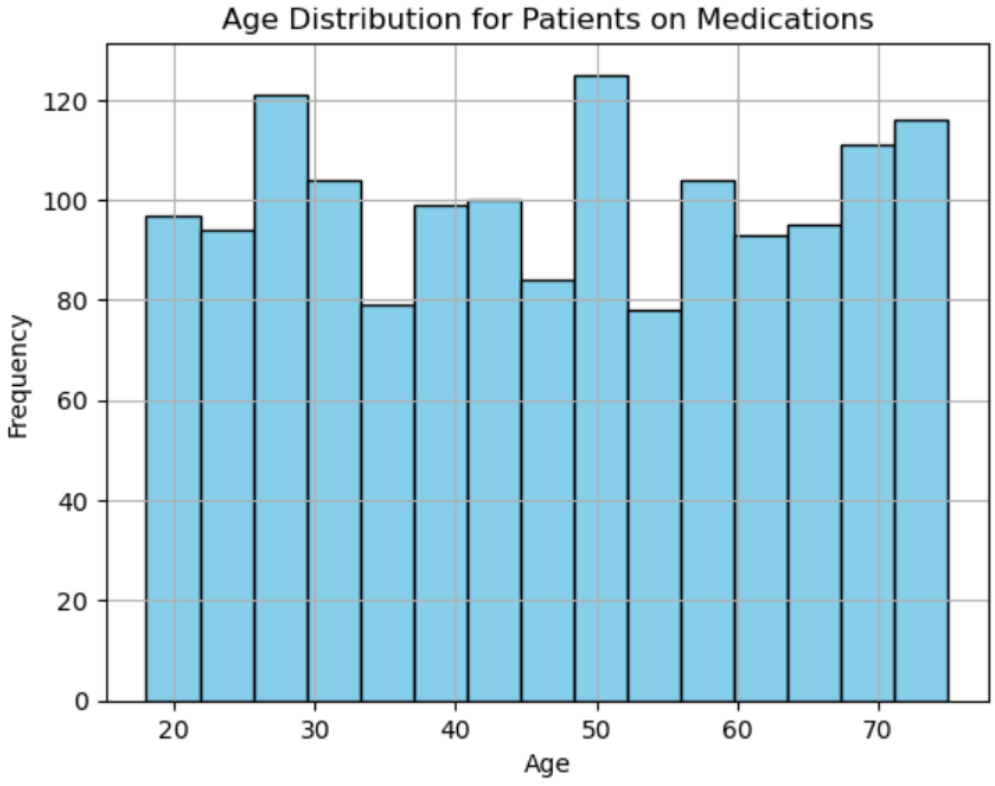

Age Distribution of Patients
Frequency
Age
0
20
40
60
80
100
120
20
30
40
50
60
70


Distribution of duration of Symptoms
Frequency
Symptoms
0
20
40
60
80
100
120
0
50
100
150
200
250


Y-BOCS Score (Obsessions))
Frequency
Obsessions
0
20
40
60
80
100
120
0
5
10
15
20
25
30
35
40


Y-BOCS Score (Compulsions)
Frequency
Compulsions
0
20
40
60
80
100
0
5
10
15
20
25
30
35
40

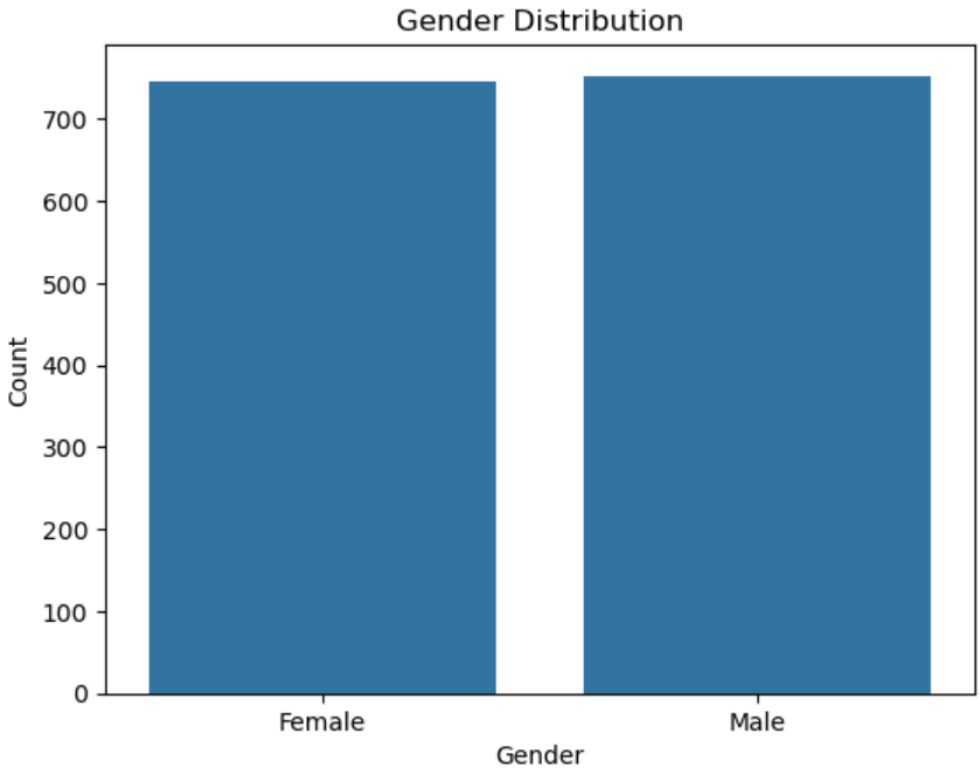


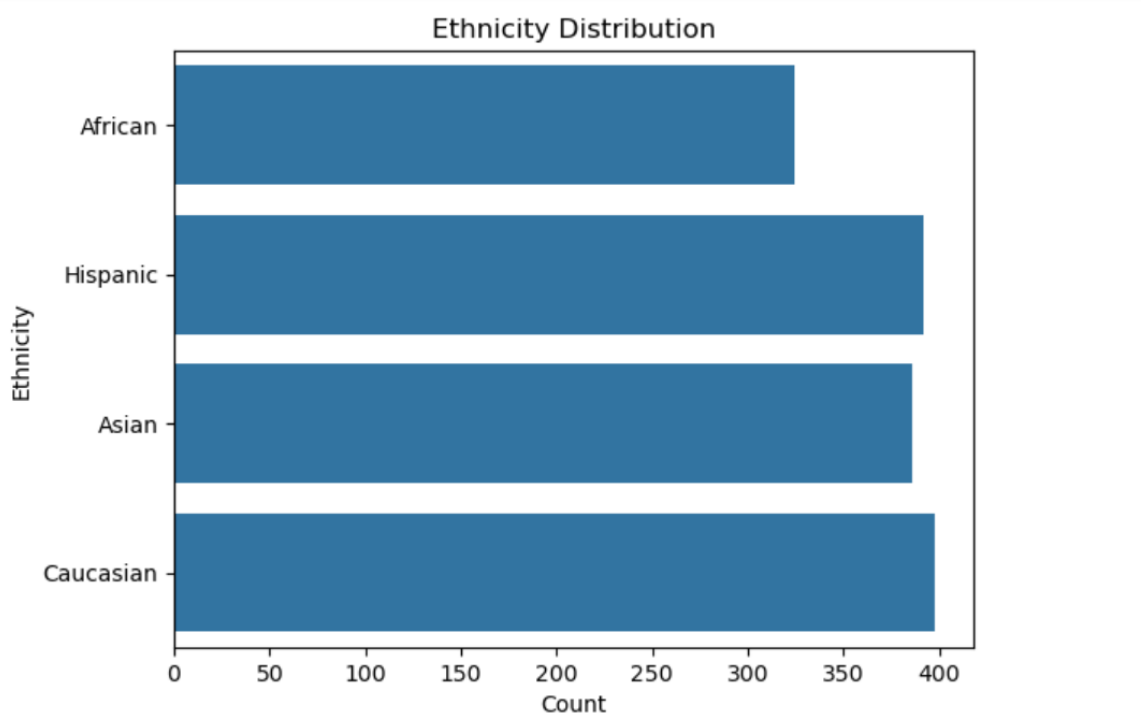


Observation: -- Age distribution for the patients who are on medications -- It is observed that that the highest are around age of 49 - 51 -- Next highest are around age of 27 - 29 -- Third highest are are 71 – 73

COUNTPLOT:

```
#Gender distribution
sns.countplot(x='Gender', data=data)
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
# Ethnicity distribution
sns.countplot(y='Ethnicity', data=data)
plt.title('Ethnicity Distribution')
plt.xlabel('Count')
plt.ylabel('Ethnicity')
plt.show()
```





BOXPLOT:

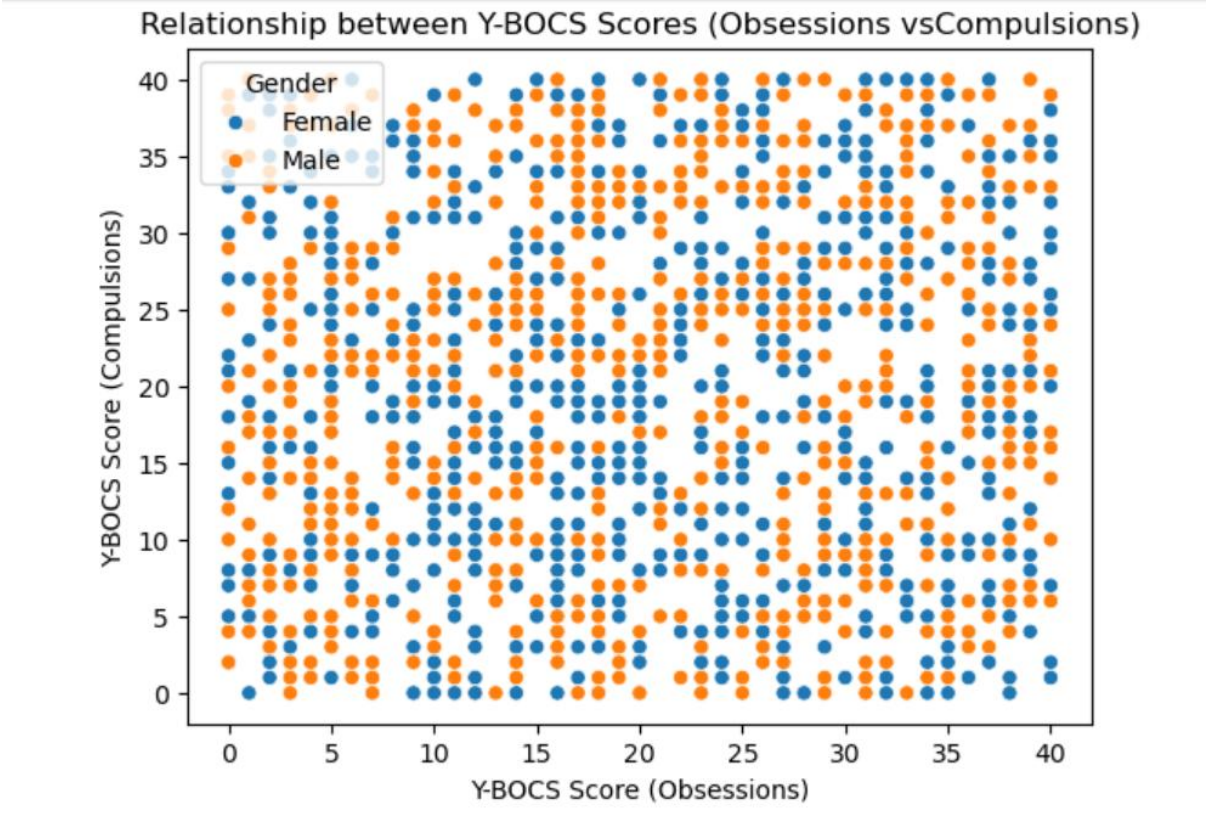
```
# Boxplot of Y-BOCS Scores by Gender
 sns.boxplot(x='Gender', y='Y-BOCS Score (Obsessions)', data=data)
 plt.title('Y-BOCS Obsession Scores by Gender')
 plt.xlabel('Gender')
 plt.ylabel('Y-BOCS Score (Obsessions)')
 plt.show()
```

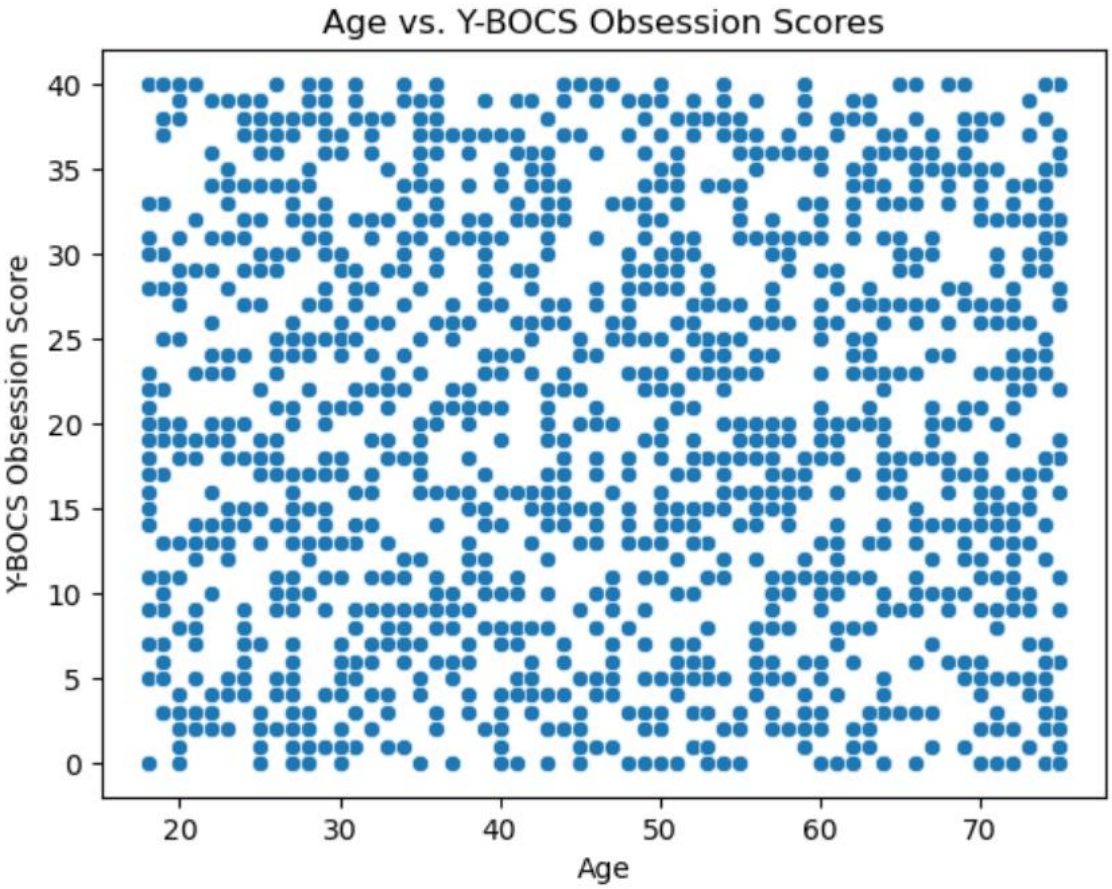
```
data['Severity Group'] = pd.cut(data['Y-BOCS Score (Obsessions)'] + data['Y-BOCS Score (Compulsions)'],
bins=[0, 14, 23, 40], labels=['Mild', 'Moderate', 'Severe'])
```

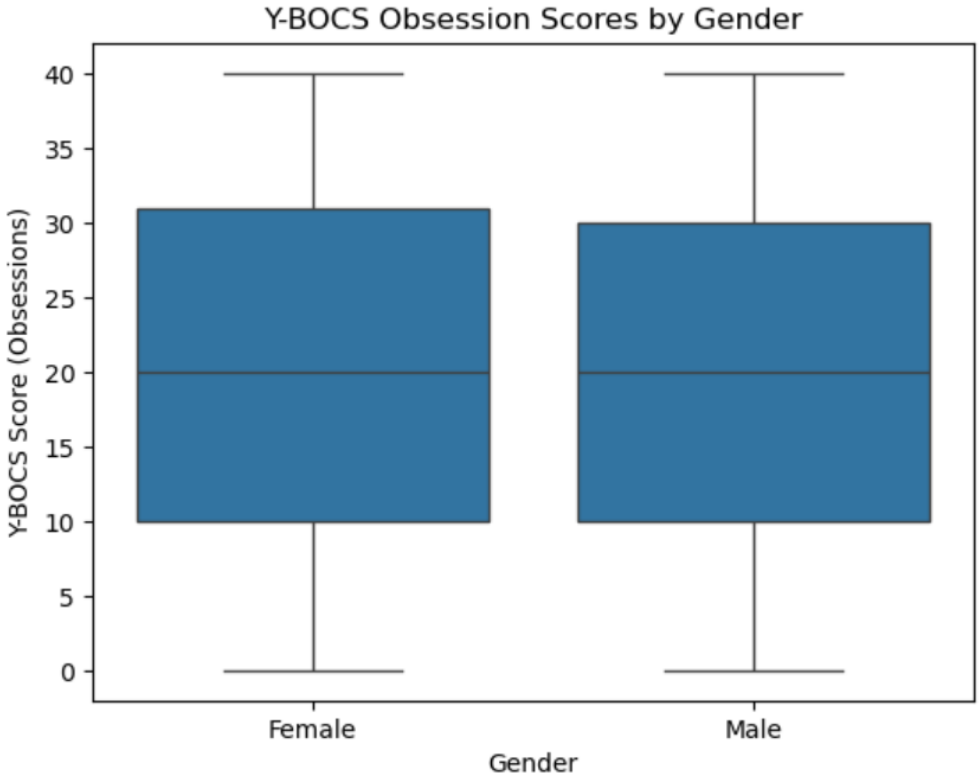
```
sns.boxplot(x='Severity Group', y='Age', data=data)
plt.title('Age Distribution Across OCD Severity Groups')
plt.xlabel('Severity Group')
plt.ylabel('Age')
plt.show()
```

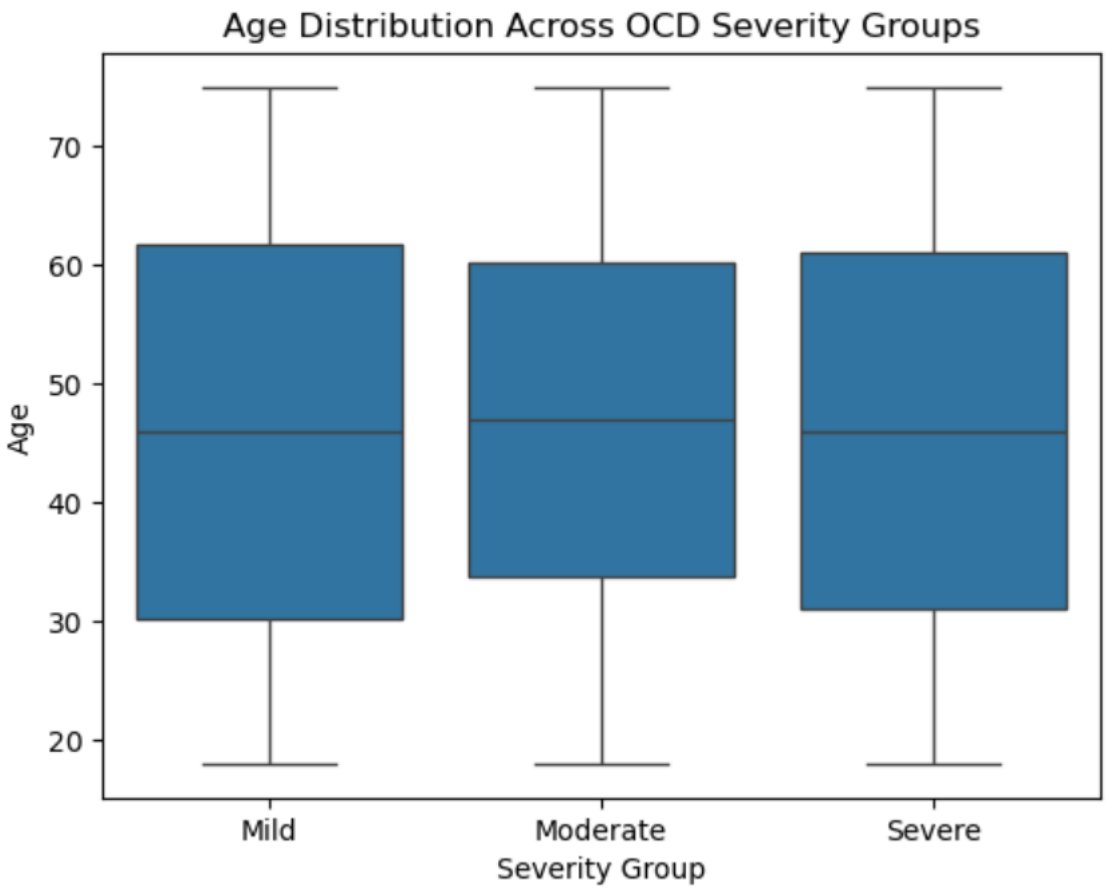
SCATTERPLOT:

```
# Relationship between Obsession and Compulsion Y-BOCS Scores

sns.scatterplot(x='Y-BOCS Score (Obsessions)', y='Y-BOCS Score (Compulsions)', hue='Gender', data=data)

plt.title('Relationship between Y-BOCS Scores (Obsessions vsCompulsions)')

plt.xlabel('Y-BOCS Score (Obsessions)')

plt.ylabel('Y-BOCS Score (Compulsions)')

plt.show()

sns.scatterplot(x='Age', y='Y-BOCS Score (Obsessions)', data=data)

plt.title('Age vs. Y-BOCS Obsession Scores')

plt.xlabel('Age')

plt.ylabel('Y-BOCS Obsession Score')

plt.show()
```



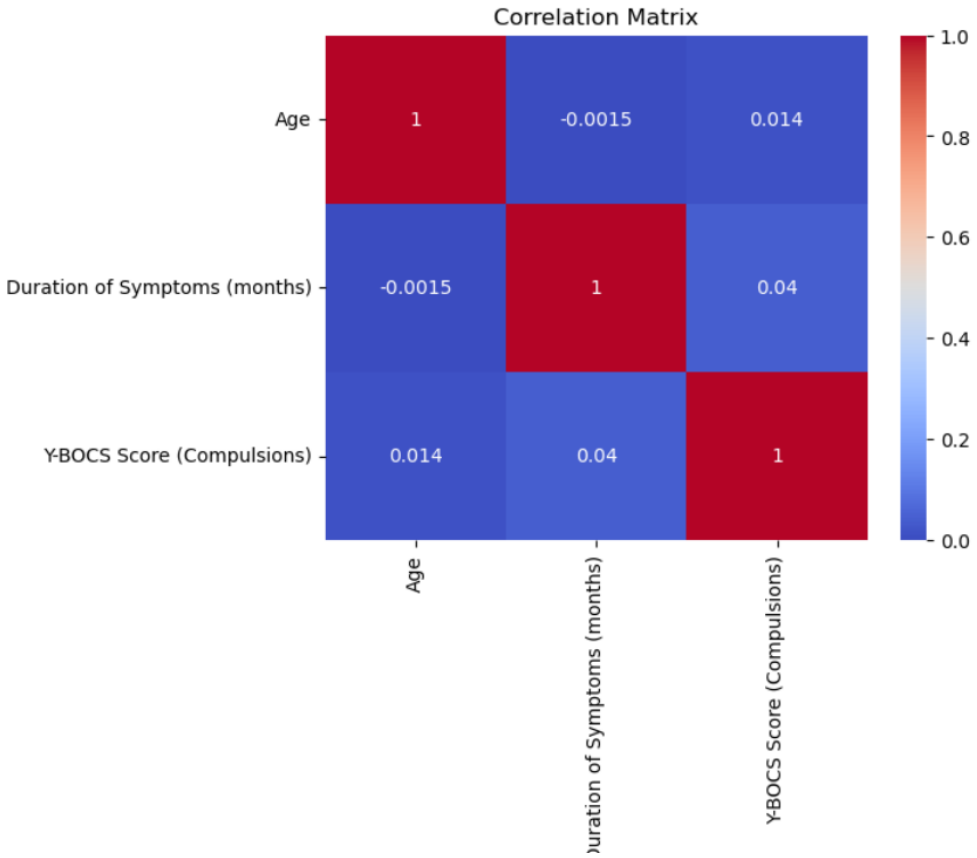

Age vs. Y-BOCS Obsession Scores
Y-BOCS Obsession Score
Age
0
5
10
15
20
25
30
35
40
20
30
40
50
60
70


Y-BOCS Obsession Scores by Gender
Y-BOCS Score (Obsessions)
Gender
Female
Male
0
5
10
15
20
25
30
35
40



HEATMAP:

```
corr_matrix = data[['Age', 'Duration of Symptoms (months)','Y-BOCS Score (Compulsions)']].corr()


sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



Observation:

The scatterplot shows a trend where younger patients have higher Y-BOCS scores: Younger individuals may experience more severe OCD symptoms. This could indicate that early-life stressors or developmental factors might influence symptom severity.

Patients with a family history of OCD have higher median Y-BOCS scores compared to those without. A genetic predisposition or familial environment might play a role in symptom severity.

Symptom Type (Obsession/Compulsion) vs. Severity:

Patients with specific obsession types (e.g., contamination fears) have higher Y-BOCS scores. Interpretation: Certain obsession types might be harder to manage, leading to more severe symptoms.

Duration of Symptoms vs. OCD Severity:

A positive correlation between the duration of symptoms and Y-BOCS scores. Longer symptom duration may worsen severity due to a lack of treatment or ineffective management.

Conclusion:

By analyzing the data and interpreting these observations, we can provide actionable insights to stakeholders, whether they are healthcare providers, researchers, or patients.

Focus on younger patients with high Y-BOCS scores, as they may need early intervention.

Tailor medication based on symptom type and severity.

Investigate why certain obsession types lead to higher severity.

Explore if medication patterns differ due to external factors like availability or socioeconomic status.

Highlight the importance of reporting symptoms early and accurately to get appropriate treatment.