**Data Analysis with Pandas**

**Sample Dataset:**

StudentID,Gender,Math\_Score,Science\_Score,Literature\_Score

1,M,85,90,88

2,F,78,85,80

3,M,90,88,92

4,F,82,78,85

5,M,88,92,90

6,F,85,80,78

**Code:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Step 1: Load the dataset

file\_path = 'student\_scores.csv' # Replace with your dataset file path

df = pd.read\_csv(file\_path)

# Display the first few rows of the dataset to understand its structure

print("First 5 rows of the dataset:")

print(df.head())

# Step 2: Perform Data Exploration

# Filter data (example: filtering students with Math\_Score >= 80)

filtered\_data = df[df['Math\_Score'] >= 80]

# Sort data (example: sorting by Literature\_Score in descending order)

sorted\_data = df.sort\_values(by='Literature\_Score', ascending=False)

# Group data (example: calculate mean scores by Gender)

mean\_scores\_by\_gender = df.groupby('Gender').mean()

# Display the results

print("\nFiltered data (Math\_Score >= 80):")

print(filtered\_data)

print("\nSorted data (by Literature\_Score in descending order):")

print(sorted\_data)

print("\nMean scores by Gender:")

print(mean\_scores\_by\_gender)

# Step 3: Calculate Summary Statistics

# Summary statistics for numeric variables

summary\_stats = df.describe()

print("\nSummary statistics for numeric variables:")

print(summary\_stats)

# Step 4: Visualize Data

# Histogram of Math\_Score distribution

plt.figure(figsize=(8, 6))

sns.histplot(df['Math\_Score'], bins=10, kde=True, color='blue')

plt.title('Distribution of Math Scores')

plt.xlabel('Math Score')

plt.ylabel('Frequency')

plt.grid(True)

plt.show()

# Scatter plot of Math\_Score vs Science\_Score

plt.figure(figsize=(8, 6))

sns.scatterplot(x='Math\_Score', y='Science\_Score', data=df, hue='Gender', palette='Set2')

plt.title('Math Score vs Science Score')

plt.xlabel('Math Score')

plt.ylabel('Science Score')

plt.grid(True)

plt.show()