import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

# Load the dataset

df = pd.read\_csv('Academic-Performance.csv')

display(df.head())

# Check for any missing values

print("Are there any missing values in the dataset? ", df.isnull().values.any())

# Sum of missing values in each column

missing\_values = df.isnull().sum()

print("Missing values in each column:\n", missing\_values)

# Handling missing values in 'math\_score'

if df['math\_score'].isnull().values.any():

median\_math = df['math\_score'].median()

df['math\_score'].fillna(median\_math, inplace=True)

print("Are there any missing values in 'math\_score' after filling? ", df['math\_score'].isnull().values.any())

# Handling missing values in 'science\_score' using interpolation

df['science\_score'] = df['science\_score'].interpolate()

print("Are there any missing values in 'science\_score' after interpolation? ", df['science\_score'].isnull().values.any())

# Dropping rows with missing 'history\_score'

df\_cleaned = df.dropna(subset=['history\_score'])

print("Missing values in each column after dropping rows with missing 'history\_score':\n", df\_cleaned.isnull().sum())

# Visualizing 'math\_score' for outliers

plt.figure(figsize=(10, 6))

plt.boxplot(df\_cleaned['math\_score'], vert=False)

plt.title('Boxplot of Math Score')

plt.xlabel('Score')

plt.grid(True)

plt.show()

# Scatterplot of 'math\_score'

plt.figure(figsize=(10, 6))

sns.scatterplot(x=df\_cleaned.index, y=df\_cleaned['math\_score'])

plt.title('Scatterplot of Math Score')

plt.xlabel('Index')

plt.ylabel('Score')

plt.grid(True)

plt.show()

# Detecting and handling outliers in 'math\_score' using IQR

Q1 = df\_cleaned['math\_score'].quantile(0.25)

Q3 = df\_cleaned['math\_score'].quantile(0.75)

IQR = Q3 - Q1

lower\_bound = Q1 - (1.5 \* IQR)

upper\_bound = Q3 + (1.5 \* IQR)

print("Lower bound:", lower\_bound)

print("Upper bound:", upper\_bound)

# Replace outliers with the mean of the 'math\_score'

mean\_math = df\_cleaned['math\_score'].mean()

df['math\_score'] = np.where((df['math\_score'] < lower\_bound) | (df['math\_score'] > upper\_bound), mean\_math, df['math\_score'])

# Scatterplot after handling outliers

df\_cleaned = df.dropna(subset=['math\_score'])

plt.figure(figsize=(10, 6))

sns.scatterplot(x=df\_cleaned.index, y=df\_cleaned['math\_score'])

plt.title('Scatterplot of Math Score after Handling Outliers')

plt.xlabel('Index')

plt.ylabel('Score')

plt.grid(True)

plt.show()

# Histogram of 'math\_score'

plt.figure(figsize=(10, 6))

df['math\_score'].hist(bins=10)

plt.xlabel('Math Score')

plt.ylabel('Number of Students')

plt.title('Histogram of Math Scores')

plt.show()