***Practical 02***

Aim : Data Preprocessing.

import pandas as pd import numpy as np import matplotlib.pyplot as plt

# Load dataset (replace 'data.csv' with your dataset) df = pd.read\_csv('/content/data.csv')

# Handling missing values

# Using isnull() and notnull() to detect missing values print("Missing Values:\n", df.isnull().sum()) print("Non-Missing Values:\n", df.notnull().sum()) # Using dropna() to drop rows with missing values cleaned\_df = df.dropna()

# Using fillna() to fill missing values with a specified value

filled\_df = df.fillna(0) # Fill missing values with 0

# Using replace() to replace blank textual data with 'zzz' # Replace blanks in 'category' column with 'zzz' df['category'] = df['category'].fillna('zzz')

# Using interpolate() to interpolate missing numerical data

interpolated\_df = df.interpolate()

# Data Visualization

# Bar graph

df['category'].value\_counts().plot(kind='bar')

plt.title('Bar Graph')

plt.xlabel('Category')

plt.ylabel('Count')

plt.show()

# Scatterplot

plt.scatter(df['x'], df['y'])

plt.title('Scatterplot')

plt.xlabel('X')

plt.ylabel('Y')

plt.show()

# Line plot

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

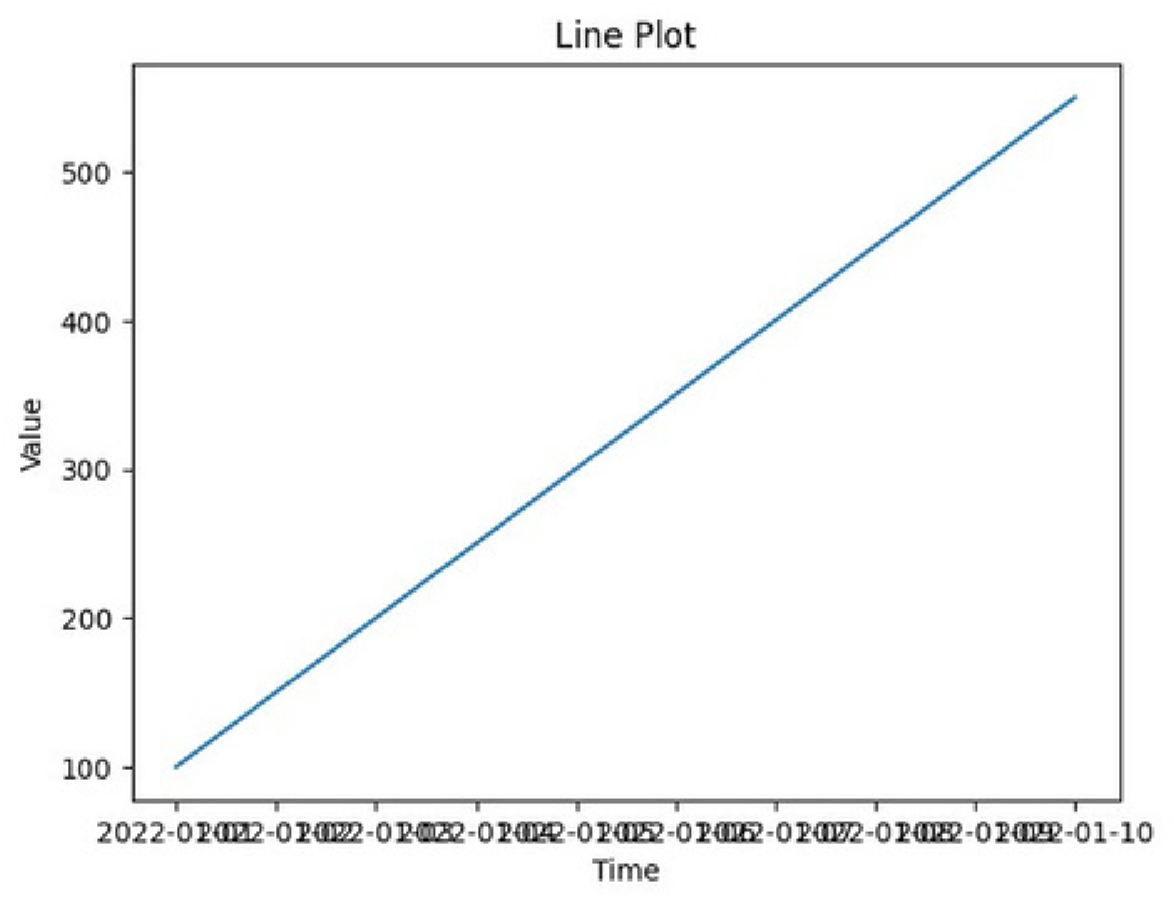
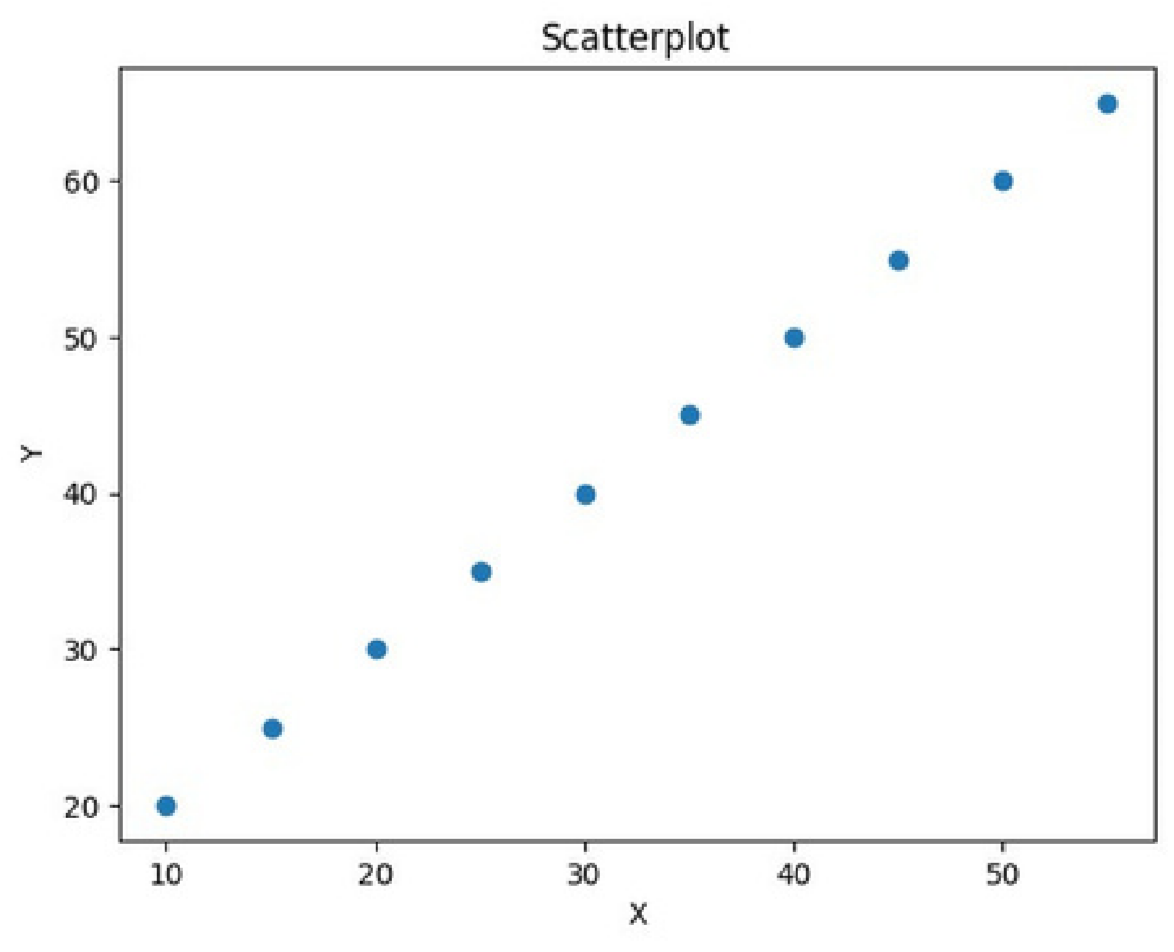
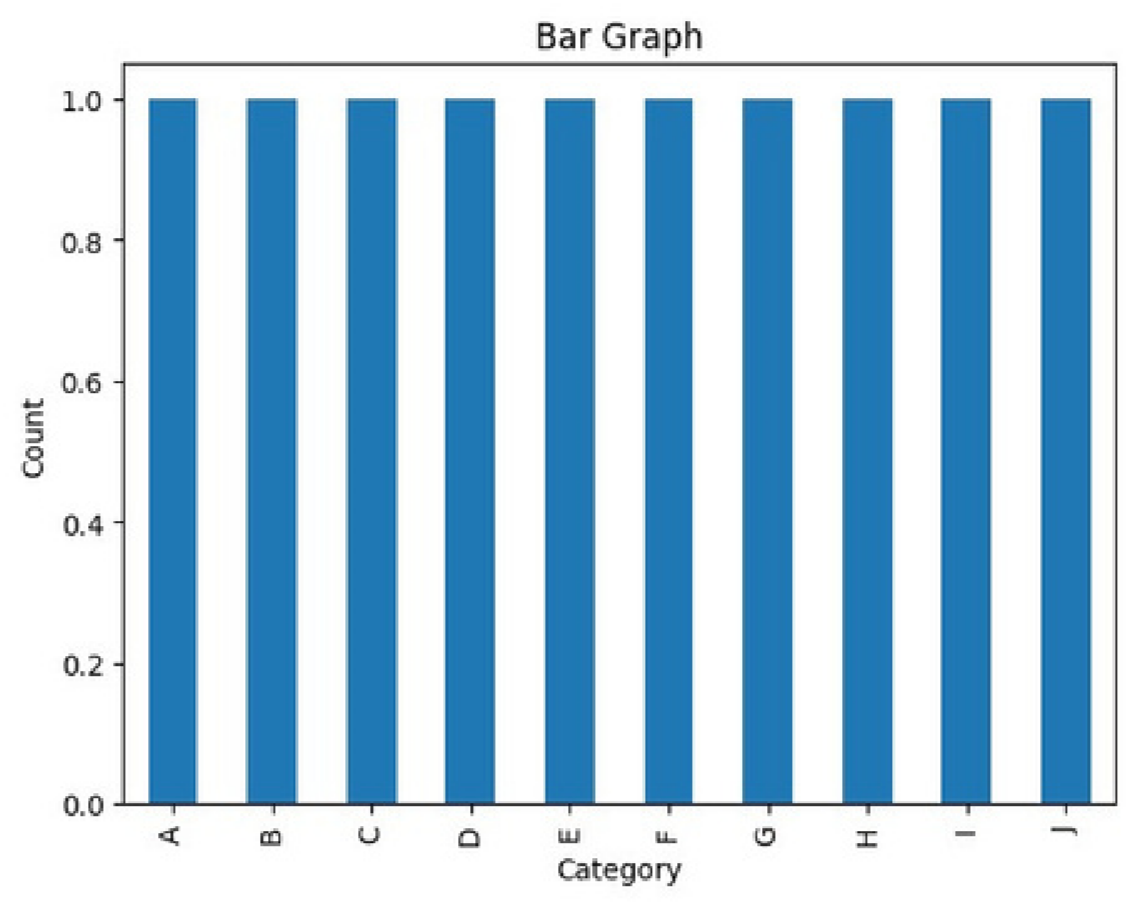
# Drop rows with NaN values in 'value' column

plt.plot(df\_cleaned['time'], df\_cleaned['value'])

plt.title('Line Plot')

plt.xlabel('Time')

plt.ylabel('Value')

plt.show() 