**SET A**

**1. Loads the Titanic dataset, shows basic information, checks for missing values, calculates**

**basic statistics, and imputes missing values for the Age column.**

import pandas as pd

url = "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"

titanic\_df = pd.read\_csv(url)

print(titanic\_df(head))

print("Basic Information:")

print(titanic\_df.info())

print("\nMissing Values:")

print(titanic\_df.isnull().sum())

print("\nBasic Statistics:")

print(titanic\_df.describe())

titanic\_df['Age'] = titanic\_df['Age'].fillna(titanic\_df['Age'].mean())

print("\nMissing Values After Imputation:")

print(titanic\_df.isnull().sum())

**2. Write a python code to produces the required visualizations: bar plot for sex, histogram for**

**age, boxplot for fare, scatter plot for age vs fare, and a heatmap of the correlation**

**matrix.(use Titanic Dataset).**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Load Titanic dataset

titanic\_df = pd.read\_csv('titanic.csv.csv')

# Basic information

print(titanic\_df.head())

plt.figure(figsize=(6,4))

sns.countplot(x="Sex", data=titanic\_df)

plt.title("Sex Distribution")

plt.show()

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

sns.histplot(titanic\_df["Age"],bins=20)

plt.title("Age Distribution")

plt.show()

plt.figure(figsize=(2,4))

sns.boxplot(titanic\_df["Fare"])

plt.title("Fare Distribution")

plt.show()

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

sns.scatterplot(x="Age", y="Fare", data=titanic\_df)

plt.title("Age vs Fare")

plt.show()

numeric\_df = titanic\_df.select\_dtypes(include=[np.number])

corr\_matrix = numeric\_df.corr()

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

sns.heatmap(corr\_matrix, annot=True, cmap="coolwarm")

plt.title("Correlation Matrix")

plt.show()

**SET B**

**1. Write a python code shows the distribution of wine quality, calculates skewness and kurtosis, and applies a log transformation to the alcohol column. (use red wine quality dataset).**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

wine\_df = pd.read\_csv("winequality-red.csv")

print(wine\_df.head())

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

sns.countplot(x="quality", data=wine\_df) # Specify x= for the 'quality' column and data=wine\_df

plt.title("Distribution of Wine Quality")

plt.show()

print("Skewness:", wine\_df["alcohol"].skew())

print("Kurtosis:", wine\_df["alcohol"].kurt())

wine\_df["alcohol\_log"] = np.log(wine\_df["alcohol"])

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

sns.histplot(wine\_df["alcohol\_log"], bins=20)

plt.title("Log-Transformed Alcohol Column")

plt.show()

**2. Write a python code to creates visualizations using Matplotlib (line plot), Seaborn (pairplot), and Plotly (interactive scatter plot). Use penguins dataset.**

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

import pandas as pd

penguins= sns.load\_dataset('penguins')

print(penguins.head())

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

plt.plot(penguins["bill\_length\_mm"], penguins["bill\_depth\_mm"])

plt.xlabel("Bill Length (mm)")

plt.ylabel("Bill Depth (mm)")

plt.title("Bill Length vs Depth")

plt.show()

sns.pairplot(penguins, hue="species")

plt.show()

fig = px.scatter(penguins, x="bill\_length\_mm", y="bill\_depth\_mm", color="species")

fig.show()