**SET A**

**Q no. 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

import numpy as np

df=pd.read\_csv("titanic.csv")

print("Basic Information:")

print(df.info())

print("\nDataset Preview:")

print(df.head())

print("\nMissing Values Count:")

print(df.isnull().sum())

**Q no. 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 seaborn as sns

import matplotlib.pyplot as plt

# Load the Titanic dataset

df=pd.read\_csv("titanic.csv")

# Visualization 1: Bar plot for Sex

plt.figure(figsize=(8, 5))

sns.countplot(x="Sex", data=df, palette="pastel")

plt.title("Bar Plot of Sex")

plt.xlabel("Sex")

plt.ylabel("Count")

plt.show()

# Visualization 2: Histogram for Age

plt.figure(figsize=(8, 5))

sns.histplot(df["Age"].dropna(), bins=20, kde=True, color="skyblue")

plt.title("Histogram of Age")

plt.xlabel("Age")

plt.ylabel("Frequency")

plt.show()

# Visualization 3: Boxplot for Fare

plt.figure(figsize=(8, 5))

sns.boxplot(x=df["Fare"], color="lightgreen")

plt.title("Boxplot of Fare")

plt.xlabel("Fare")

plt.show()

# Visualization 4: Scatter plot for Age vs Fare

plt.figure(figsize=(8, 5))

sns.scatterplot(x="Age", y="Fare", data=df, hue="Pclass", palette="viridis", alpha=0.7)

plt.title("Scatter Plot of Age vs Fare")

plt.xlabel("Age")

plt.ylabel("Fare")

plt.legend(title="Pclass")

plt.show()

# Visualization 5: Heatmap of the correlation matrix

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

correlation\_matrix = df.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap="coolwarm", fmt=".2f")

plt.title("Heatmap of Correlation Matrix")

plt.show()

**SET B**

**Q no. 1) Write a python code shows the distribution of wine quality, calculates skewness and kurtosis, and applies a logtransformation to the alcohol column. (use red wine quality dataset).**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import numpy as np

from scipy.stats import skew, kurtosis

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

sns.countplot(x="quality", data=df, palette="pastel")

plt.title("Distribution of Wine Quality")

plt.xlabel("Quality")

plt.ylabel("Count")

plt.show()

print(f"Skewness of Alcohol: {skew(df['alcohol']):.2f}")

print(f"Kurtosis of Alcohol: {kurtosis(df['alcohol']):.2f}")

**Q no. 2)** **Write a python code to creates visualizations using Matplotlib (line plot), Seaborn (pairplot),and Plotly (interactive scatter plot). Use penguins dataset.**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

df = pd.read\_csv("penguins.csv")

# 1. Line Plot using Matplotlib

plt.plot(df['bill\_length\_mm'], label='Bill Length (mm)', color='blue')

plt.plot(df['bill\_depth\_mm'], label='Bill Depth (mm)', color='orange')

plt.title("Bill Measurements of Penguins")

plt.xlabel("Index")

plt.ylabel("Measurements (mm)")

plt.legend()

plt.show()

# 2. Pair Plot using Seaborn

sns.pairplot(df, hue='species')

plt.show()

# 3. Interactive Scatter Plot using Plotly

fig = px.scatter(df, x='flipper\_length\_mm', y='body\_mass\_g', color='species', title="Flipper Length vs

Body Mass")

fig.show()