Task no 1

!pip install seaborn matplotlib pandas

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df = sns.load\_dataset('titanic')

df.head()

# Missing values

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

# Drop columns with too many nulls

df.drop(['deck'], axis=1, inplace=True)

# Fill missing age with median

df['age'].fillna(df['age'].median(), inplace=True)

# Drop remaining nulls

df.dropna(inplace=True)

# Remove duplicates

df.drop\_duplicates(inplace=True)

# Bar plot for categorical variable

sns.countplot(x='class', data=df)

plt.title('Passenger Class Distribution')

plt.show()

# Histogram for numerical feature

df['age'].hist(bins=20)

plt.title('Age Distribution')

plt.xlabel('Age')

plt.ylabel('Count')

plt.show()

# Correlation heatmap

sns.heatmap(df.corr(numeric\_only=True), annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap')

plt.show()

print("Survival rate by class:\n", df.groupby('class')['survived'].mean())

print("Survival rate by gender:\n", df.groupby('sex')['survived'].mean())