### Step 1: Setting up the environment

First, ensure you have pandas and matplotlib installed in your environment. If not, you can install them using pip:

### pip install pandas matplotlib seaborn

### Step 2: Loading the Dataset

We'll start by loading the Titanic dataset into a pandas DataFrame. You can download the dataset from here.

### import pandas as pd

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

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

### print(titanic.head())

### Step 3: Basic Data Exploration

Next, we'll perform some basic data exploration to understand the structure of the dataset.

### print(titanic.describe())

### print(titanic.isnull().sum())

### print(titanic['Survived'].value\_counts())

### Step 4: Data Visualization

We'll use matplotlib and seaborn to visualize various aspects of the dataset.

#### **4.1: Histograms**

Histograms are useful for understanding the distribution of numeric variables.

### import matplotlib.pyplot as plt

### titanic.hist(bins=20, figsize=(10, 10))

### plt.suptitle("Histograms of Titanic Dataset Features")

### plt.show()

#### **4.2: Bar Plot**

Bar plots can help visualize the counts of categorical variables.

### import seaborn as sns

### sns.countplot(x='Survived', data=titanic)

### plt.title('Survival Counts')

### plt.show()

### sns.countplot(x='Pclass', data=titanic)

### plt.title('Passenger Counts by Class')

### plt.show()

#### **4.3: Box Plot**

Box plots are useful for visualizing the distribution of numeric variables across different categories.

sns.boxplot(x='Pclass', y='Age', data=titanic)

plt.title('Age Distribution by Class')

plt.show()

sns.boxplot(x='Pclass', y='Fare', data=titanic)

plt.title('Fare Distribution by Class')

plt.show()

#### **4.4: Heatmap**

Heatmaps can show correlations between different numeric variables.

corr\_matrix = titanic.corr()

plt.figure(figsize=(10, 8))

sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Correlation Heatmap')

plt.show()

### Step 5: Summary and Insights

Finally, we summarize the findings and insights from our analysis.

1. **Survival Rates**: The bar plot of survival counts shows the number of survivors and non-survivors. It can be seen that more people did not survive than survived.
2. **Class Distribution**: The bar plot of passengers by class shows the distribution of passengers across different classes, with most passengers being in the third class.
3. **Age and Fare Distribution**: The box plots of age and fare distribution by class show that passengers in higher classes tended to be older and paid higher fares.
4. **Correlation Analysis**: The heatmap of the correlation matrix reveals that there is a positive correlation between Fare and Survived, indicating that higher fares were associated with higher survival rates.