```
[1]: !pip install pandas numpy matplotlib seaborn scipy statsmodels
```

```
Requirement already satisfied: pandas in c:\users\dharshini k s\anaconda3\lib\site-packages (2.2.3)
Requirement already satisfied: numpy in c:\users\dharshini k s\anaconda3\lib\site-packages (2.1.3)
```

```
[2]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     %matplotlib inline
```

```
[6]: df=pd.read_csv("C:/Users/Dharshini K S/Downloads/titanic.csv")
```

```
[8]: # Show the first few rows of the DataFrame
     df.head()

     # Get information about the data types and missing values
     df.info()

     # Get statistical summary of numerical columns
     df.describe()

     # Check for unique values in categorical columns
     df['Sex'].value_counts()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
```

```
[8]: Sex
     male      577
     female    314
     Name: count, dtype: int64
```

```
[9]: # Plot a histogram for a numerical column
     sns.histplot(data=df, x='Age', kde=True)

     # Plot a countplot for a categorical column
     sns.countplot(data=df, x='Sex')

     # Plot a boxplot to check for outliers in a numerical column
     sns.boxplot(data=df, y='Fare')
```

```
[9]: <Axes: xlabel='Age', ylabel='Count'>
```

```
[10]: # Create a heatmap to visualize correlations between numerical variables
      corr_matrix = df.corr(numeric_only=True)
      sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')

      # Create a pairplot for a quick overview of relationships
      sns.pairplot(df[['Age', 'Fare', 'Survived']])

      # Use a violinplot to compare a numerical variable across categories
      sns.violinplot(x='Sex', y='Age', data=df)
```

```
[10]: <Axes: xlabel='Survived', ylabel='Count'>
```

```
[11]: # Example: Scatter plot of Age vs. Fare, colored by survival
      sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
```

```
[11]: <Axes: xlabel='Survived', ylabel='Count'>
```

This histogram visualizes the distribution of ages among Titanic passengers. We can see that the majority of passengers were young adults, mainly between 20 and 40 years old. There's also a smaller peak for very young children.The majority of passengers did not survive. Female passengers and children had a higher survival rate. Passengers in first class had a much higher chance of survival than those in third class.