

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

# Display settings
sns.set(style="whitegrid")
plt.style.use('ggplot')
%matplotlib inline
```

```
In [ ]: df = pd.read_csv("train.csv") # Update with correct path if needed
df.head()
```

```
In [ ]: # Shape and Data Types
print("Dataset Shape:", df.shape)
df.info()

# Summary statistics
df.describe()

# Check missing values
df.isnull().sum()
```

```
In [ ]: # Categorical: Survival Count
sns.countplot(data=df, x='Survived')
plt.title('Survival Count')
plt.show()

# Categorical: Passenger Class
sns.countplot(data=df, x='Pclass')
plt.title('Passenger Class Distribution')
plt.show()

# Numerical: Age Distribution
sns.histplot(data=df, x='Age', kde=True)
plt.title('Age Distribution')
plt.show()

# Fare
sns.histplot(data=df, x='Fare', kde=True)
plt.title('Fare Distribution')
plt.show()
```

```
In [ ]: # Survival by Gender
sns.countplot(data=df, x='Sex', hue='Survived')
plt.title('Survival by Gender')
plt.show()

# Survival by Class
sns.countplot(data=df, x='Pclass', hue='Survived')
plt.title('Survival by Passenger Class')
plt.show()

# Age vs Survival
sns.boxplot(data=df, x='Survived', y='Age')
plt.title('Age vs Survival')
plt.show()
```

```
# Fare vs Survival
sns.boxplot(data=df, x='Survived', y='Fare')
plt.title('Fare vs Survival')
plt.show()
```

```
In [ ]: # Correlation Matrix
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()

# Pairplot
sns.pairplot(df[['Survived', 'Pclass', 'Age', 'Fare']], hue='Survived')
plt.show()
```

```
In [ ]: # Fill Age with median
df['Age'].fillna(df['Age'].median(), inplace=True)

# Drop Cabin due to many nulls
df.drop(columns=['Cabin'], inplace=True)

# Drop rows with any remaining nulls
df.dropna(inplace=True)
```

Insights:

- Survival rate was higher among females.
- Younger passengers had a slightly higher survival rate.
- Higher fare and upper-class (Pclass=1) passengers had better survival chances.
- Strong correlation between Pclass and Fare.

These trends suggest socio-economic status played a significant role in survival chances.