**Titanic Dataset - Exploratory Data Analysis (EDA)**

**==================================================**

import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt

**Load dataset**

train\_df = pd.read\_csv(‘train.csv’)

**Step 1: Dataset Overview**

print(“Data Info:”)  
print(train\_df.info())  
print("\nDescriptive Statistics:")  
print(train\_df.describe(include=‘all’))

**Value Counts**

print("\nPclass Counts:")  
print(train\_df[‘Pclass’].value\_counts())

print("\nSex Counts:")  
print(train\_df[‘Sex’].value\_counts())

print("\nEmbarked Counts:")  
print(train\_df[‘Embarked’].value\_counts())

**Step 2: Correlation Heatmap**

plt.figure(figsize=(10, 6))  
corr = train\_df.corr(numeric\_only=True)  
sns.heatmap(corr, annot=True, cmap=‘coolwarm’, fmt=".2f")  
plt.title(‘Correlation Heatmap’)  
plt.tight\_layout()  
plt.show()

**Step 3: Trend Visualization**

fig, axs = plt.subplots(2, 2, figsize=(14, 10))

**Age Distribution**

sns.histplot(data=train\_df, x=‘Age’, hue=‘Survived’, multiple=‘stack’, ax=axs[0, 0], bins=30)  
axs[0, 0].set\_title(‘Age Distribution by Survival’)

**Fare vs Survival**

sns.boxplot(data=train\_df, x=‘Survived’, y=‘Fare’, ax=axs[0, 1])  
axs[0, 1].set\_title(‘Fare vs Survival’)

**Age by Pclass**

sns.boxplot(data=train\_df, x=‘Pclass’, y=‘Age’, ax=axs[1, 0])  
axs[1, 0].set\_title(‘Age by Passenger Class’)

**Sex vs Survival**

sns.countplot(data=train\_df, x=‘Sex’, hue=‘Survived’, ax=axs[1, 1])  
axs[1, 1].set\_title(‘Survival Count by Sex’)

plt.tight\_layout()  
plt.show()

**Step 4: Observations**

print("\nObservations:")  
print("- Females and children were more likely to survive.")  
print("- Higher fares were associated with better survival outcomes.")  
print("- 1st class passengers had a higher survival rate.")  
print("- Clear survival bias based on gender and class.")

**Step 5: Summary**

print("\nSummary of Findings:")  
print("""

1. **Sex**: Females had higher survival rates.
2. **Age**: Younger individuals (especially children) had better survival outcomes.
3. **Class**: 1st class passengers were more likely to survive.
4. **Fare**: Paying higher fares correlated with higher survival.
5. **Missing Data**: Age, Cabin, and Embarked columns had missing values.  
   “”")