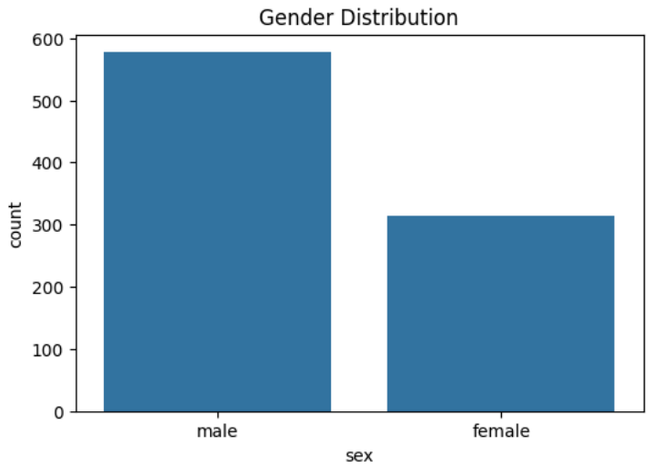
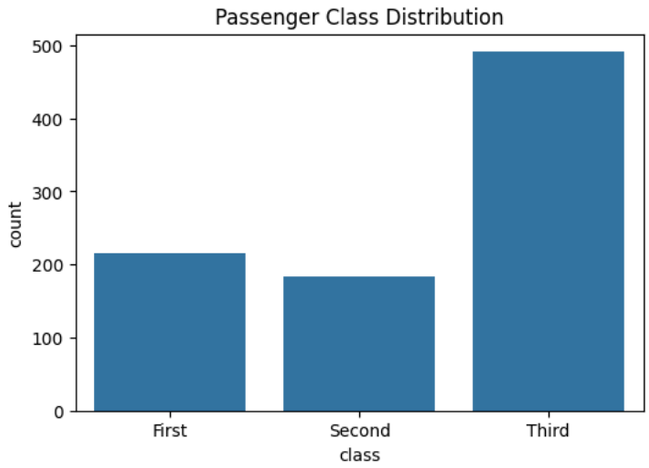
```
# Titanic EDA Project


# Step 1: Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Step 2: Load dataset (from seaborn's built-in Titanic dataset)
df = sns.load_dataset("titanic")

# Step 3: Basic Info
print("Shape of dataset:", df.shape)
print("\nMissing values:\n", df.isnull().sum())
print("\nFirst 5 rows:\n", df.head())
```

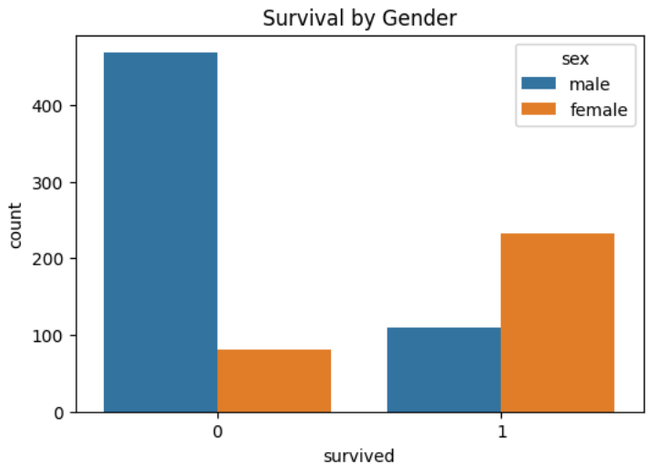
```
Shape of dataset: (891, 15)

Missing values:
 survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64

First 5 rows:
    survived  pclass     sex   age  sibsp  parch     fare embarked  class  \
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True
```

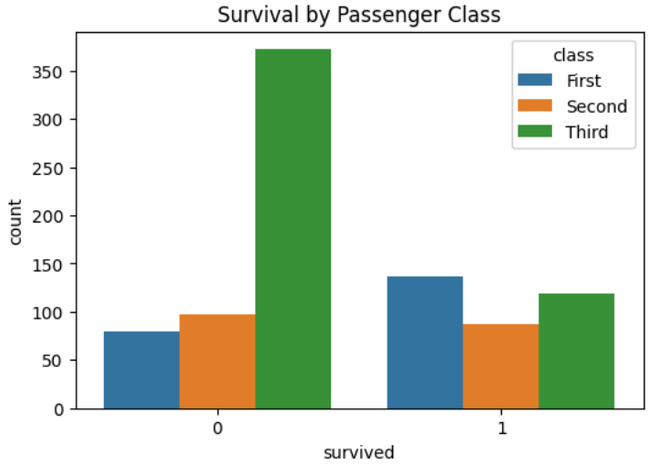
```
# Step 4: Univariate Analysis
plt.figure(figsize=(6,4))
sns.countplot(data=df, x="sex")
plt.title("Gender Distribution")
plt.show()

plt.figure(figsize=(6,4))
sns.countplot(data=df, x="class")
plt.title("Passenger Class Distribution")
plt.show()
```



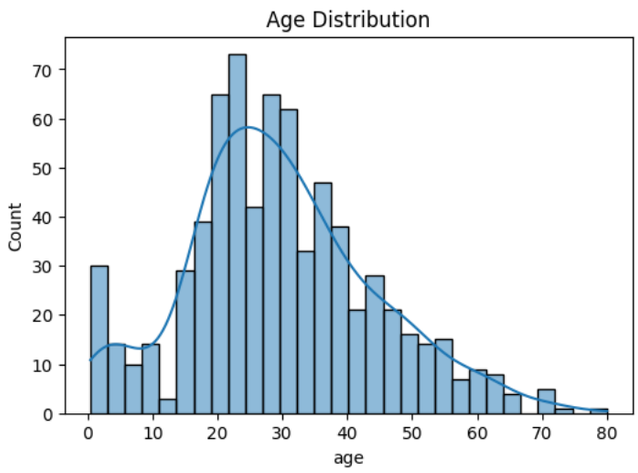



```
# Step 5: Survival Analysis
plt.figure(figsize=(6,4))
sns.countplot(data=df, x="survived", hue="sex")
plt.title("Survival by Gender")
plt.show()

plt.figure(figsize=(6,4))
sns.countplot(data=df, x="survived", hue="class")
plt.title("Survival by Passenger Class")
plt.show()
```





```
# Step 6: Age Distribution
plt.figure(figsize=(6,4))
sns.histplot(df['age'].dropna(), kde=True, bins=30)
plt.title("Age Distribution")
plt.show()
```




```
# Step 7: Correlation Heatmap
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```