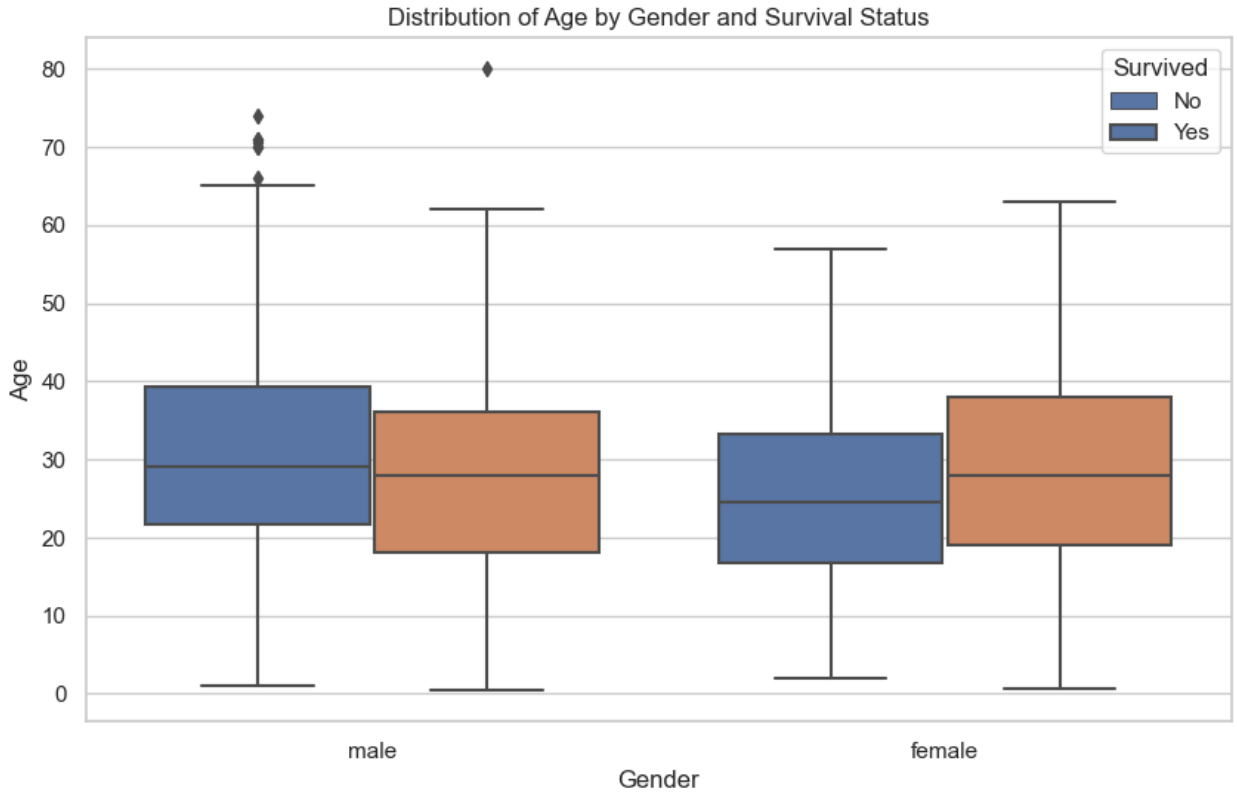
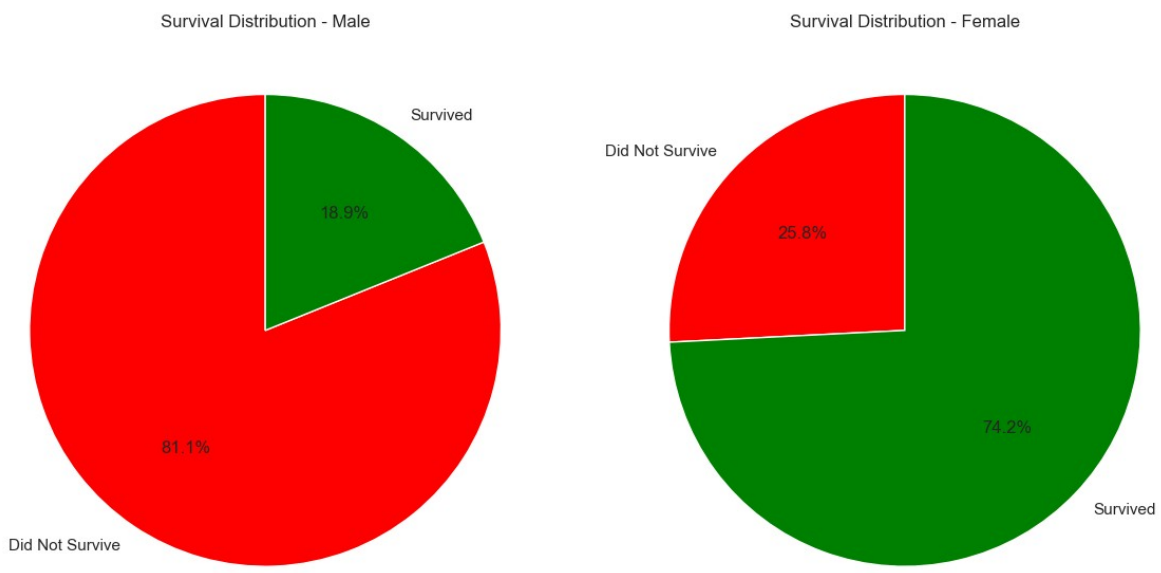
```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
titanic = sns.load_dataset('titanic')
titanic.head()
```

```
   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
```

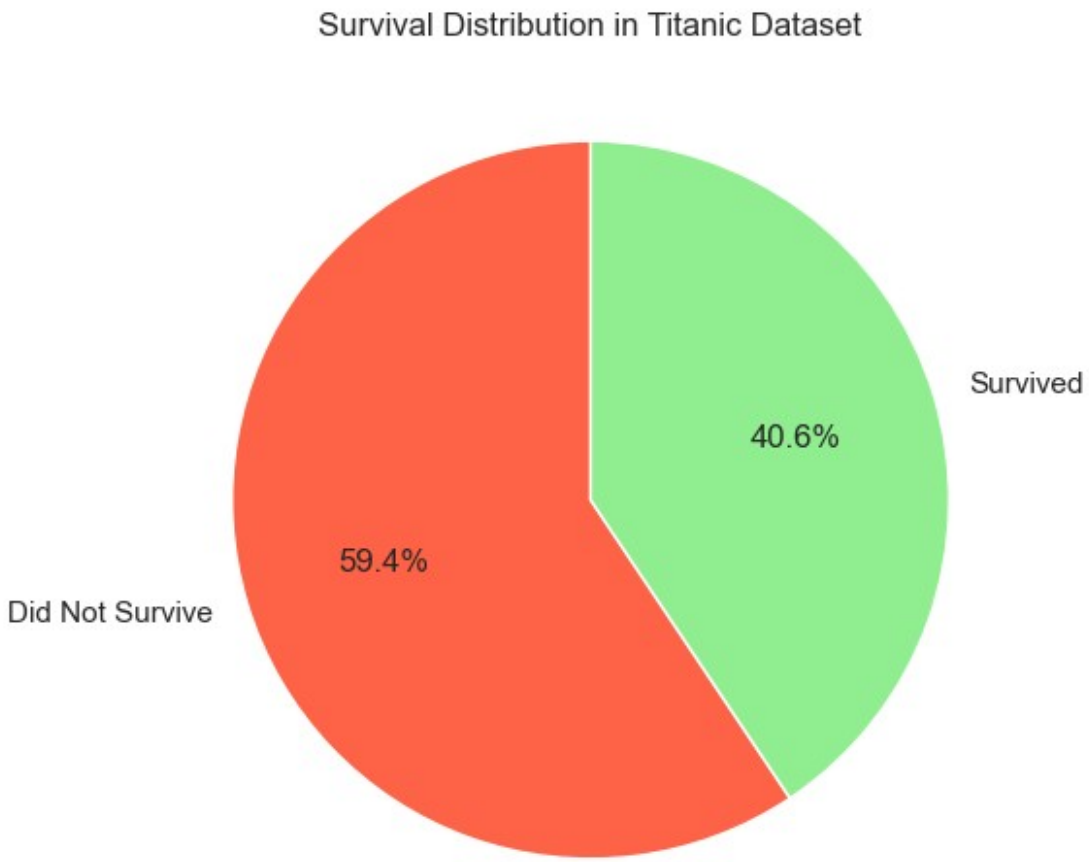
```
# Set the style for the plot
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.boxplot(data=titanic, x='sex', y='age', hue='survived')
plt.title('Distribution of Age by Gender and Survival Status')
plt.xlabel('Gender')
plt.ylabel('Age')
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.show()
```




```
gender = titanic.groupby(['sex', 'survived']).size().unstack()
labels = ['Did Not Survive', 'Survived']
colors = ['red', 'green']
# Plot Pie Charts
fig, axes = plt.subplots(1, 2, figsize=(12, 6))
for i, sex in enumerate(['male', 'female']):
    axes[i].pie(gender_survival.loc[sex], labels=labels,
autopct='%1.1f%%', colors=colors, startangle=90)
    axes[i].set_title(f'Survival Distribution - {sex.capitalize()}')

plt.tight_layout()
plt.show()
```



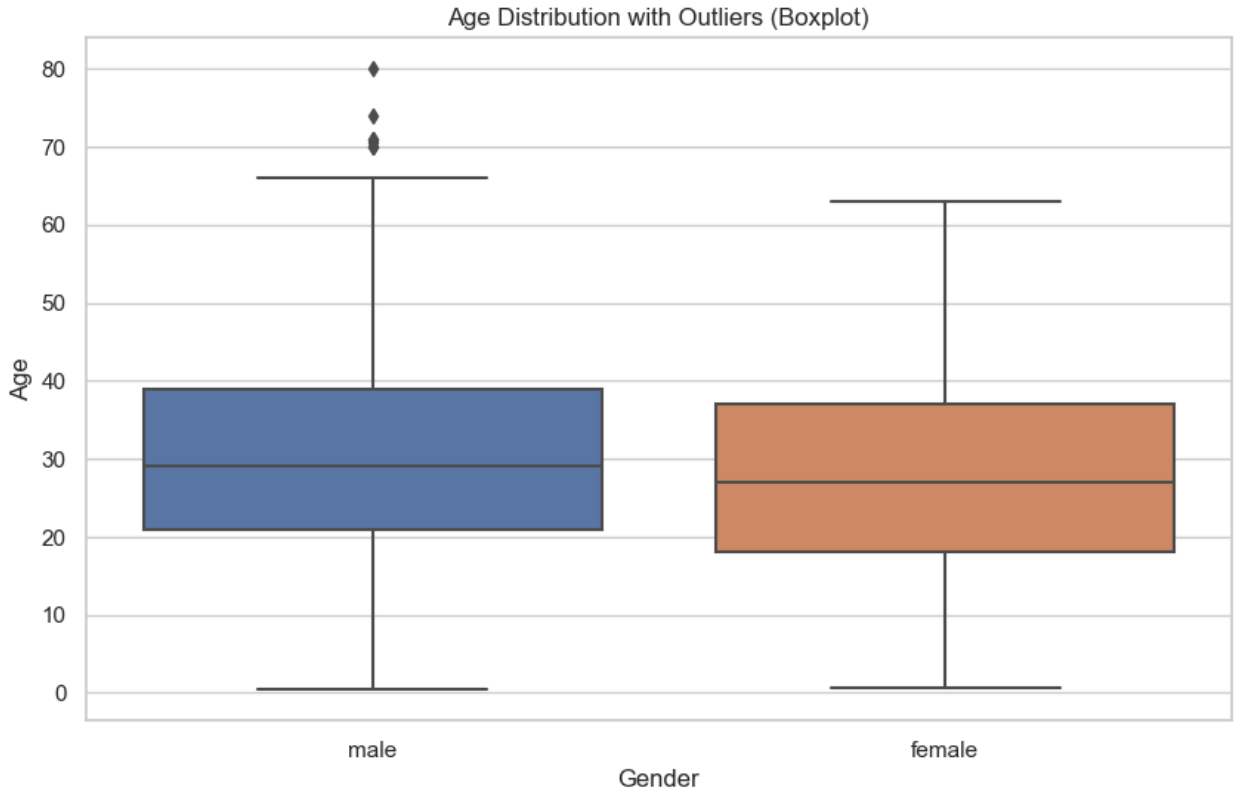

```
titanic_cleaned = titanic.dropna(subset=['survived'])
survival_counts = titanic_cleaned['survived'].value_counts()
plt.figure(figsize=(6,6))
plt.pie(survival_counts, labels=['Did Not Survive', 'Survived'],
autopct='%1.1f%%', colors=['#FF6347', '#90EE90'], startangle=90)
plt.title('Survival Distribution in Titanic Dataset')
plt.show()
```

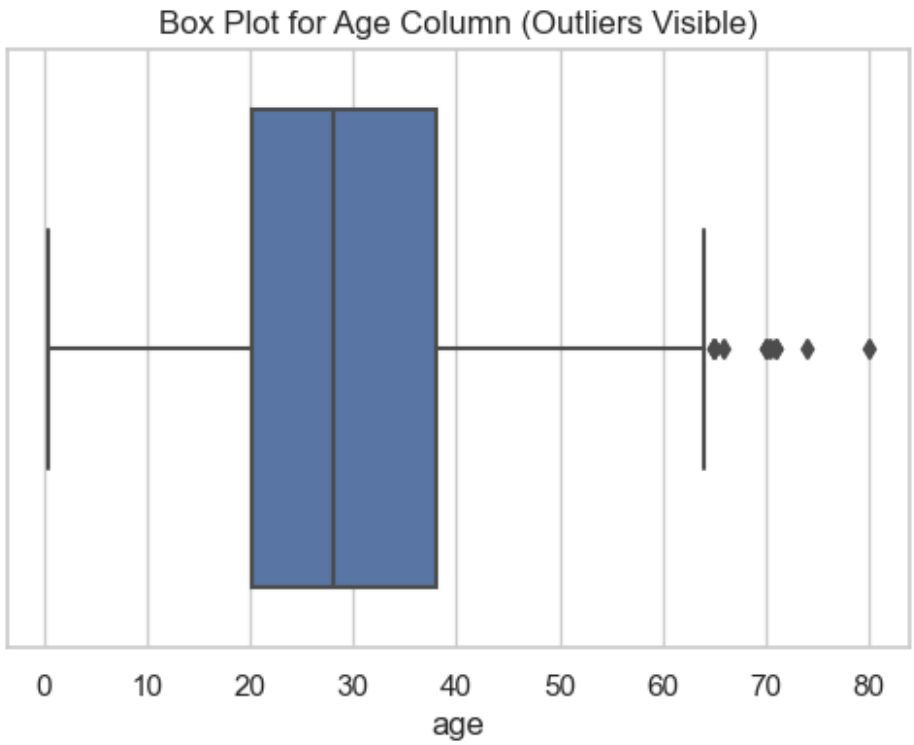



```
Q1 = titanic_cleaned['age'].quantile(0.25)  # 25th percentile
Q3 = titanic_cleaned['age'].quantile(0.75)  # 75th percentile
IQR = Q3 - Q1  # Interquartile range

lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

outliers = titanic_cleaned[(titanic_cleaned['age'] < lower_bound) |
(titanic_cleaned['age'] > upper_bound)]
# Print out the outliers
print("Outliers:")
print(outliers[['age', 'sex', 'survived']])

plt.figure(figsize=(10, 6))
sns.boxplot(x='sex', y='age', data=titanic_cleaned)

plt.title('Age Distribution with Outliers (Boxplot)')
plt.xlabel('Gender')
plt.ylabel('Age')
```

```
# Display the plot
plt.show()
```

```
Outliers:
      age   sex  survived
33   66.0  male         0
54   65.0  male         0
96   71.0  male         0
116  70.5  male         0
280  65.0  male         0
456  65.0  male         0
493  71.0  male         0
630  80.0  male         1
672  70.0  male         0
745  70.0  male         0
851  74.0  male         0
```




Box Plot for Age Column (Outliers Visible)
0
10
20
30
40
50
60
70
80
age