dsbdafinal9

May 4, 2025

[1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[3]:
```
# Load the Titanic dataset
df = pd.read_csv("/Users/anant/Downloads/titanic_train.csv.xls")
```

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 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          714 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  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

[5]:
```
# Display basic info (optional but useful for checking)
print(df[['Sex', 'Age', 'Survived']].info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   Sex       891 non-null    object
```

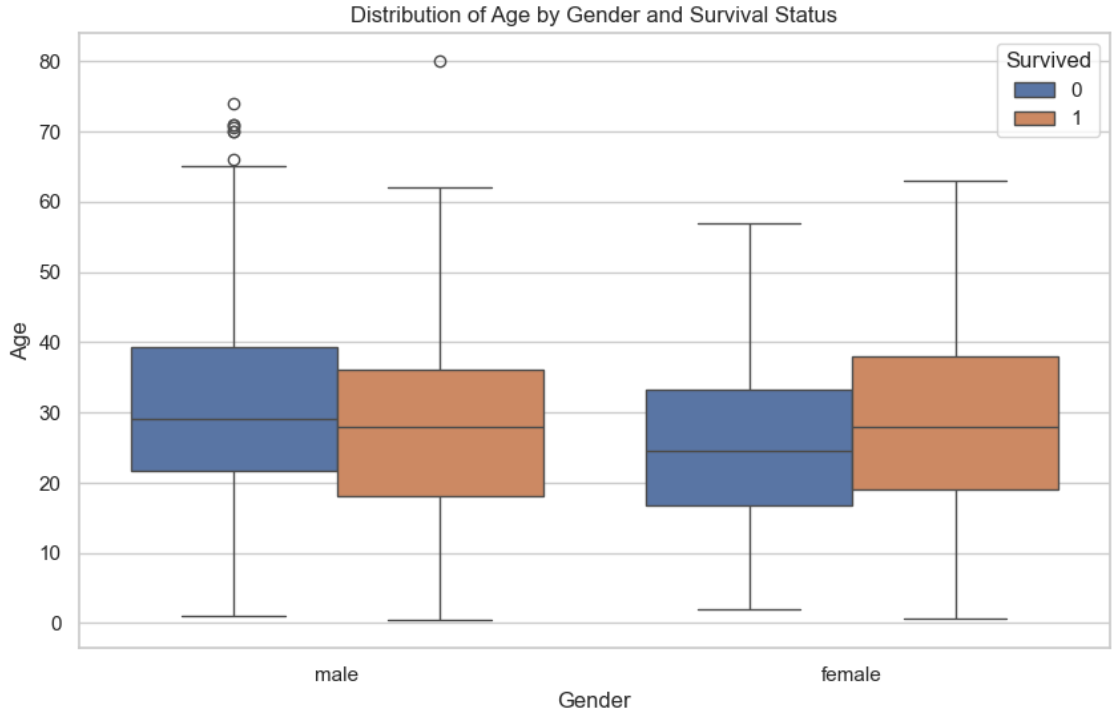
```
 1   Age       714 non-null    float64
 2   Survived  891 non-null    int64
dtypes: float64(1), int64(1), object(1)
memory usage: 21.0+ KB
None
```

[6]:
```
# Drop rows with missing age values
df = df.dropna(subset=['Age'])
```

[7]:
```
# Set the plot style
sns.set(style="whitegrid")
```

[8]:
```
# Create the box plot
plt.figure(figsize=(10, 6))
sns.boxplot(x='Sex', y='Age', hue='Survived', data=df)

# Add plot title and labels
plt.title('Distribution of Age by Gender and Survival Status')
plt.xlabel('Gender')
plt.ylabel('Age')
plt.show()
```



[ ]: