



```
In [1]: print("My Jupyter is working!")
```

My Jupyter is working!

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")
```

Matplotlib is building the font cache; this may take a moment.

```
In [3]: df = sns.load_dataset("titanic")
df.head()
```

```
Out[3]:   survived  pclass    sex   age  sibsp  parch     fare  embarked  class  who
          0         0      3  male  22.0      1      0    7.2500        S  Third  man
          1         1      1  female  38.0      1      0   71.2833        C  First  woman
          2         1      3  female  26.0      0      0    7.9250        S  Third  woman
          3         1      1  female  35.0      1      0   53.1000        S  First  woman
          4         0      3  male  35.0      0      0    8.0500        S  Third  man
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   survived    891 non-null    int64  
 1   pclass      891 non-null    int64  
 2   sex         891 non-null    object  
 3   age         714 non-null    float64 
 4   sibsp       891 non-null    int64  
 5   parch       891 non-null    int64  
 6   fare         891 non-null    float64 
 7   embarked    889 non-null    object  
 8   class        891 non-null    category
 9   who          891 non-null    object  
 10  adult_male   891 non-null    bool   
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object  
 13  alive        891 non-null    object  
 14  alone        891 non-null    bool  
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```

```
In [5]: df.describe()
```

Out[5]:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

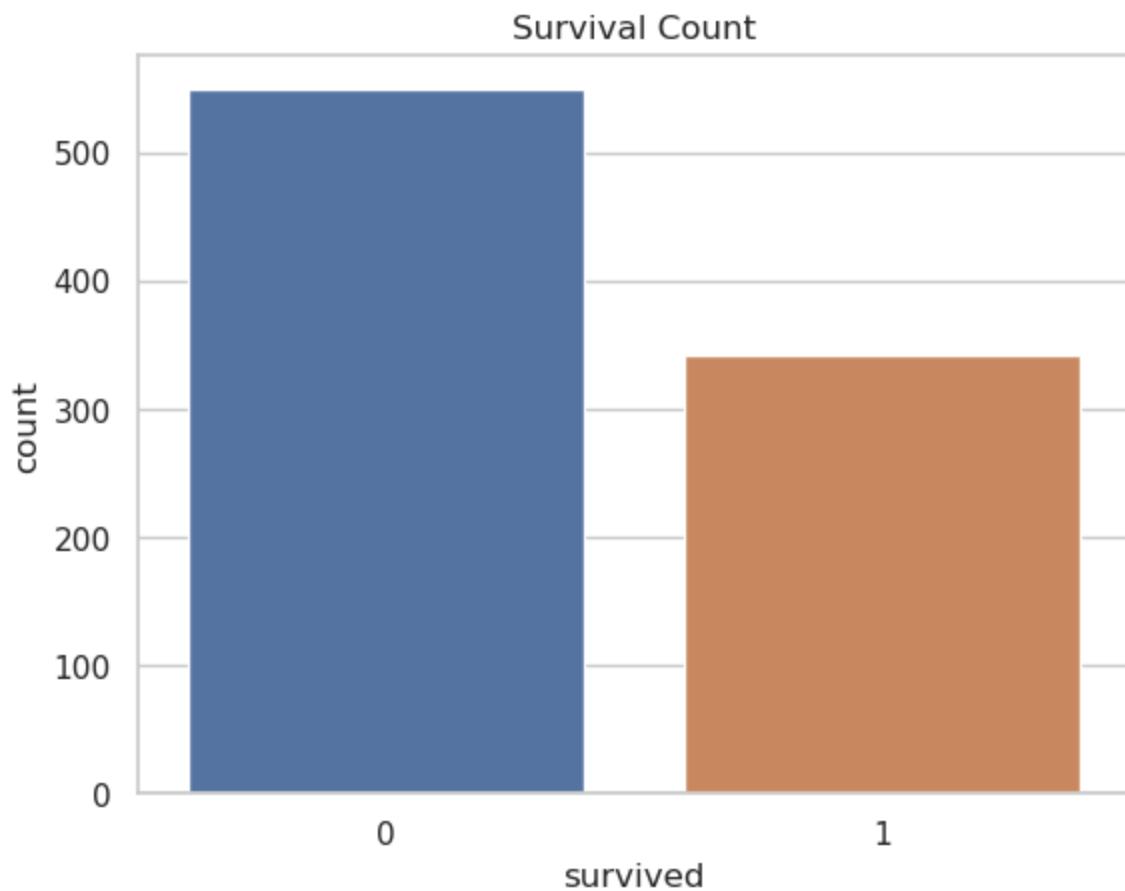
In [6]: `df.isnull().sum()`

Out[6]:

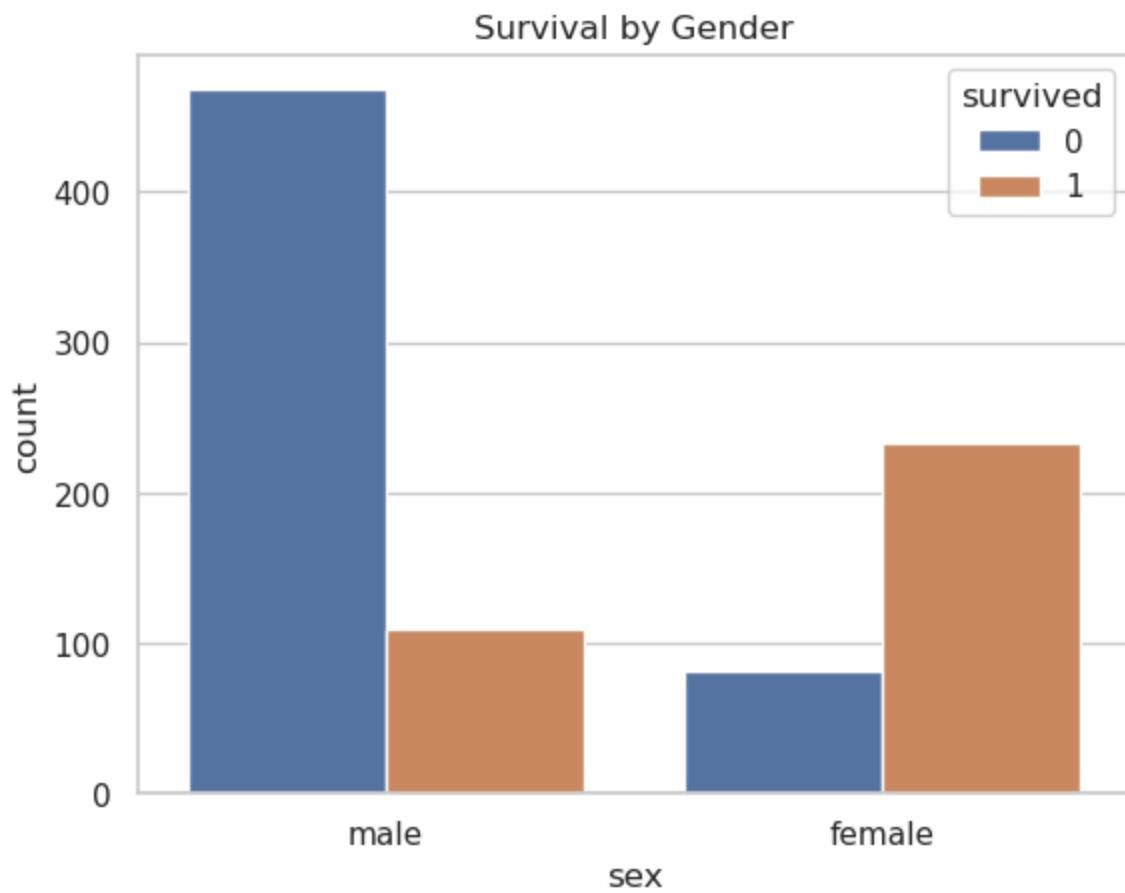
survived	0
pclass	0
sex	0
age	177
sibsp	0
parch	0
fare	0
embarked	2
class	0
who	0
adult_male	0
deck	688
embark_town	2
alive	0
alone	0

`dtype: int64`

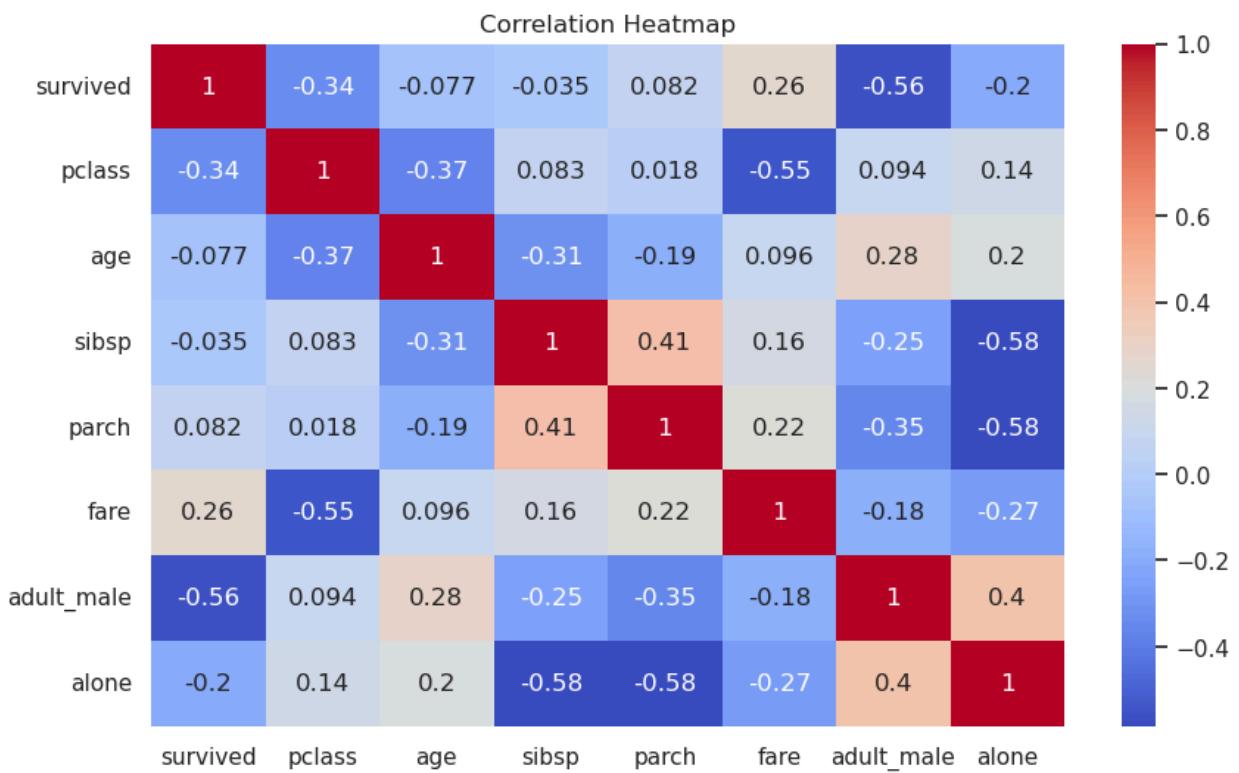
In [7]: `sns.countplot(x="survived", data=df)`
`plt.title("Survival Count")`
`plt.show()`



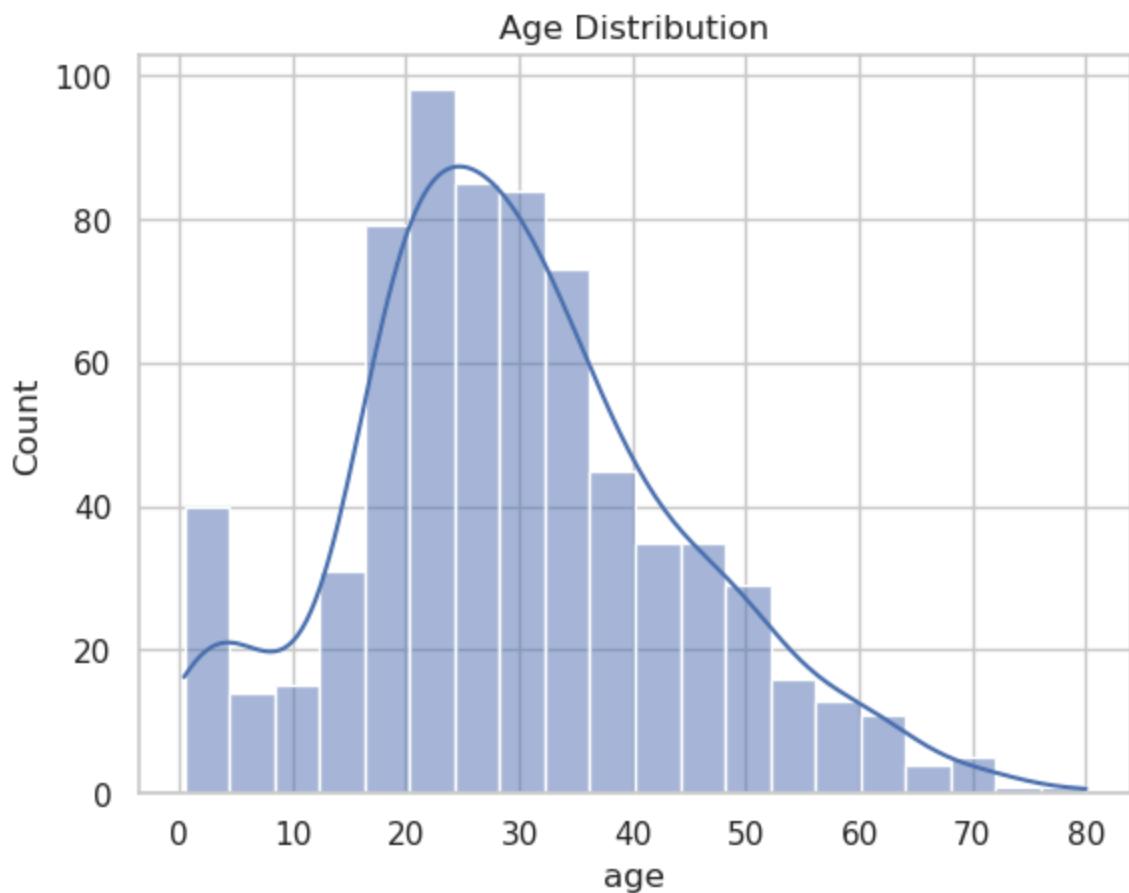
```
In [8]: sns.countplot(x="sex", hue="survived", data=df)
plt.title("Survival by Gender")
plt.show()
```



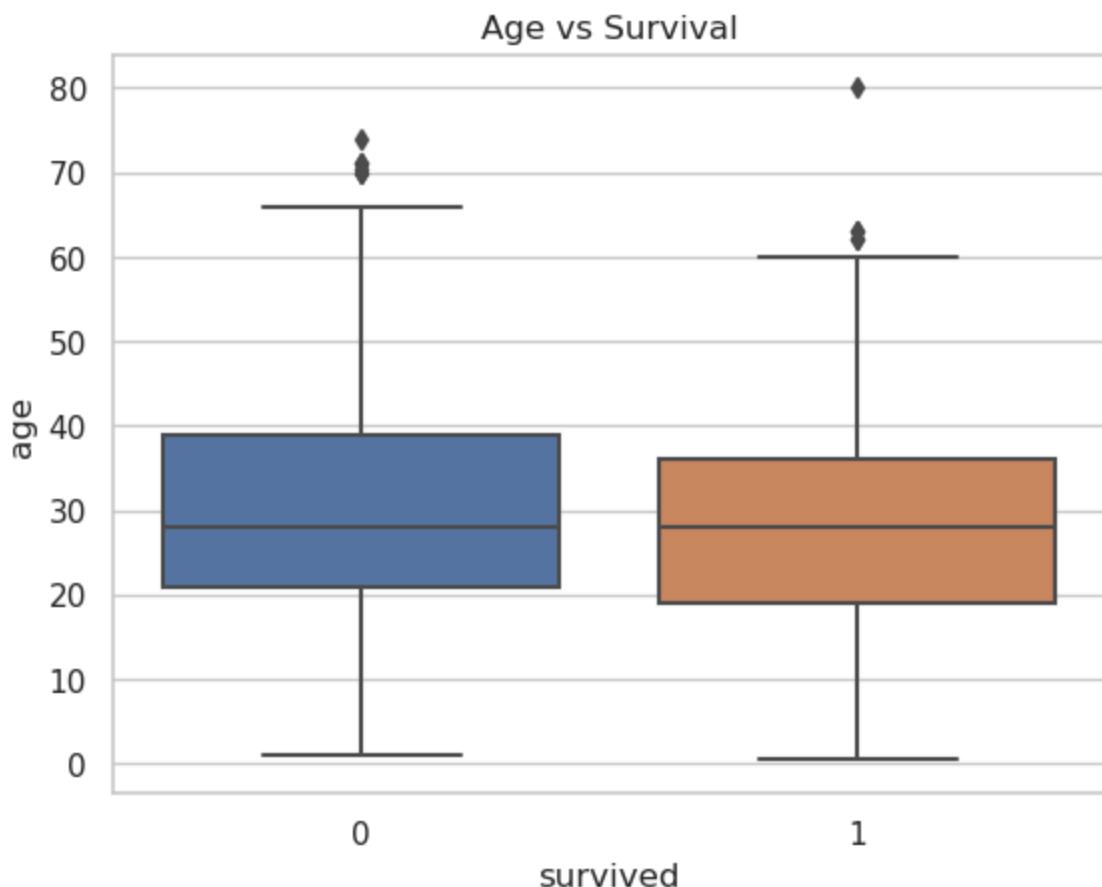
```
In [9]: plt.figure(figsize=(10,6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```



```
In [10]: sns.histplot(df["age"], kde=True)
plt.title("Age Distribution")
plt.show()
```



```
In [11]: sns.boxplot(x="survived", y="age", data=df)
plt.title("Age vs Survival")
plt.show()
```



Exploratory Data Analysis (EDA)

Titanic Dataset

Name: Krishita Banik Internship Task 5

1. Importing Libraries

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

2. Loading Dataset

```
In [15]: df = sns.load_dataset("titanic")
df.head()
```

Out[15]:

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

3. Basic Information

In [16]:

```
df.info()
df.describe()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   survived    891 non-null    int64  
 1   pclass      891 non-null    int64  
 2   sex         891 non-null    object  
 3   age         714 non-null    float64 
 4   sibsp       891 non-null    int64  
 5   parch       891 non-null    int64  
 6   fare         891 non-null    float64 
 7   embarked    889 non-null    object  
 8   class        891 non-null    category
 9   who          891 non-null    object  
 10  adult_male  891 non-null    bool   
 11  deck         203 non-null    category
 12  embark_town 889 non-null    object  
 13  alive        891 non-null    object  
 14  alone        891 non-null    bool  
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```

```
Out[16]: 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
```

4. Survival Count Analysis

Observation:

The number of passengers who did not survive is higher than those who survived. This indicates that the overall survival rate was low.

5. Final Summary of Findings

Through EDA on the Titanic dataset, survival was strongly influenced by gender and passenger class. Females and first-class passengers had significantly higher survival rates. Age had moderate impact, while fare showed positive correlation with survival. The dataset contained missing values which were handled appropriately.

Overall, clear patterns and trends were identified through statistical and visual exploration.