

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

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

```
In [2]: titanic = sns.load_dataset("titanic")
titanic.head()
```

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



```
In [3]: titanic.info()
titanic.describe()
```

```
<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[3]:

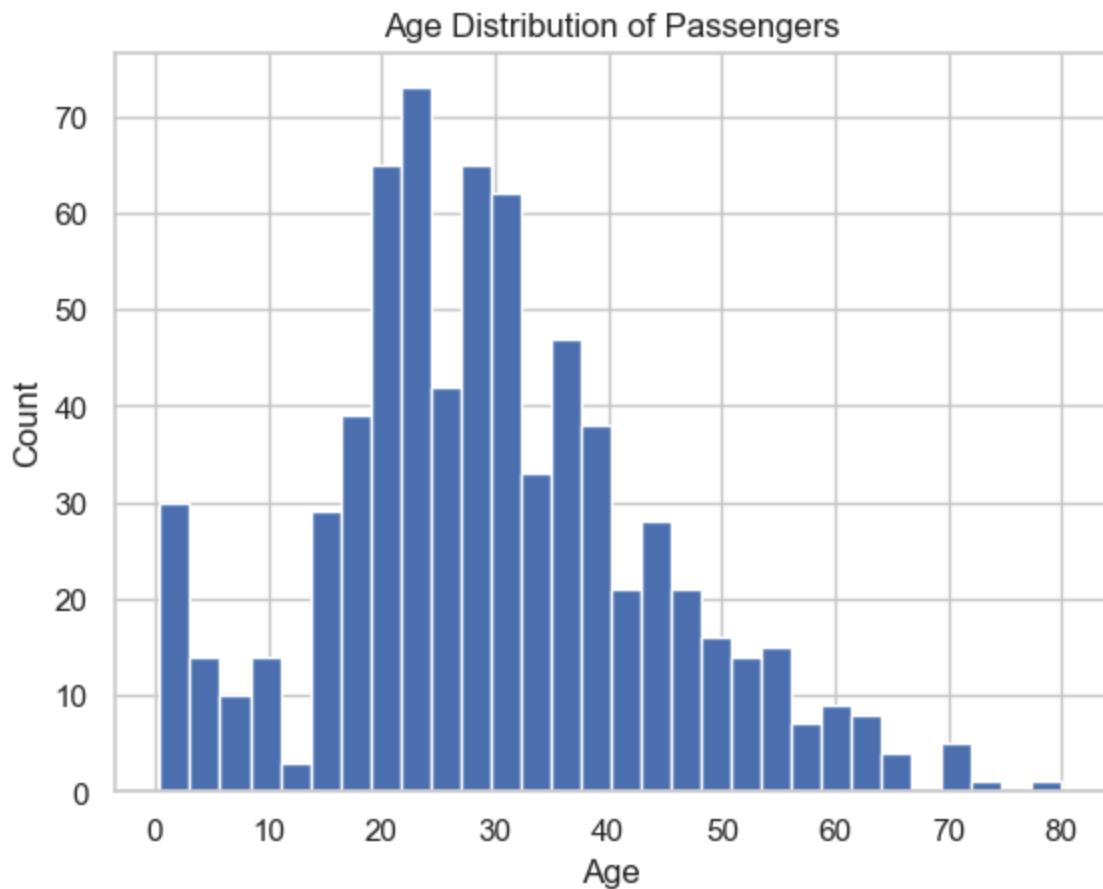
	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	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

Initial Observations

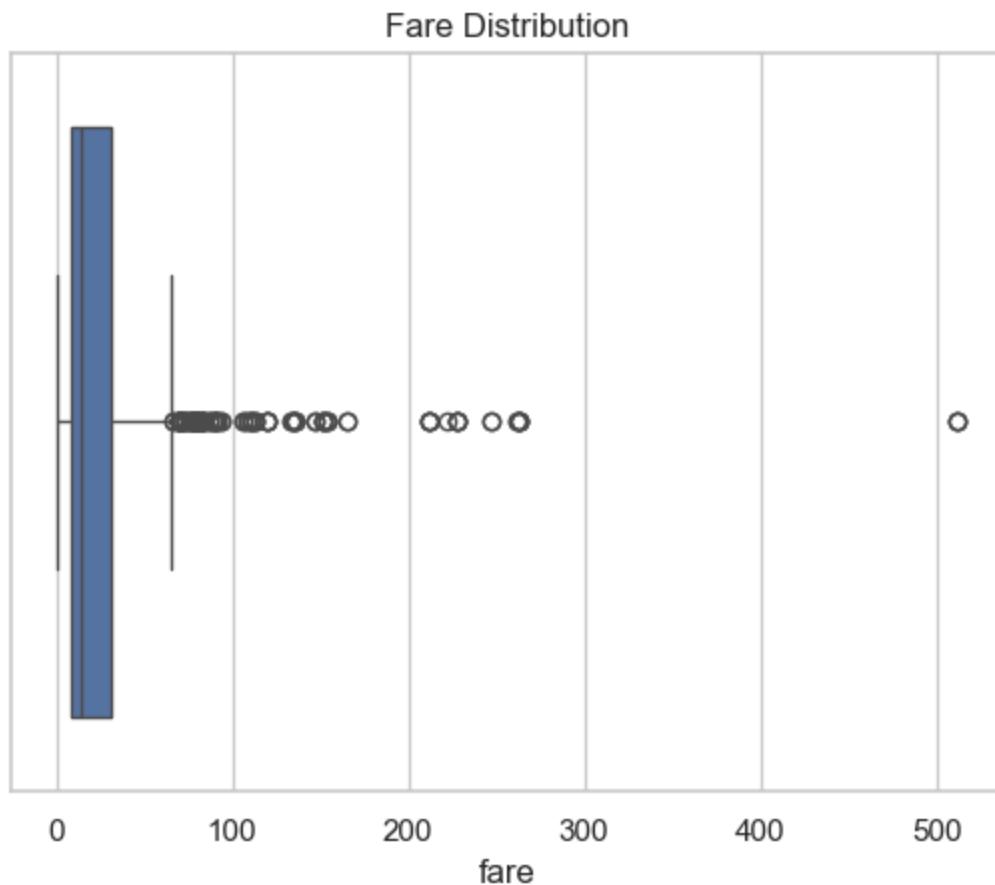
- The Titanic dataset contains 891 rows and 15 columns.
- Some columns such as age and deck contain missing values.
- The average passenger age is around 30 years.
- Fare values vary widely, indicating possible outliers.

In [4]:

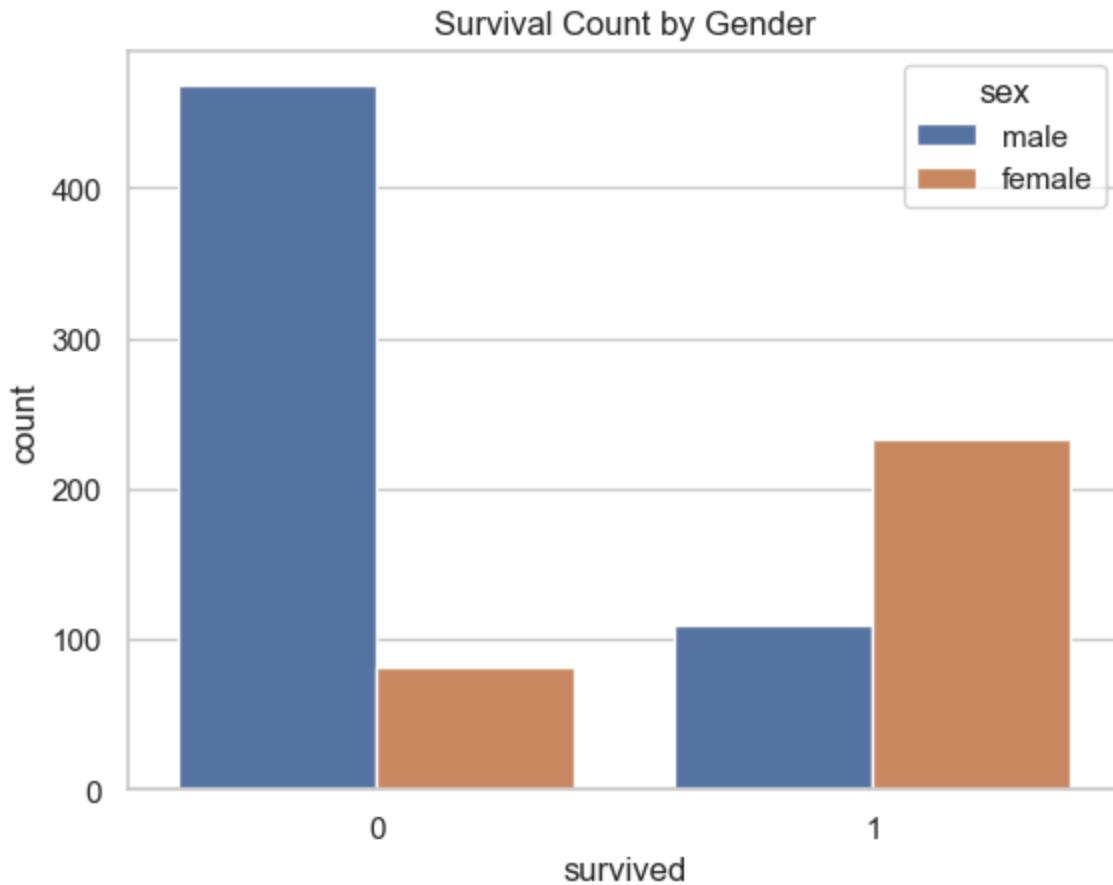
```
plt.figure()
titanic['age'].hist(bins=30)
plt.title("Age Distribution of Passengers")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```



```
In [5]: plt.figure()
sns.boxplot(x=titanic['fare'])
plt.title("Fare Distribution")
plt.show()
```



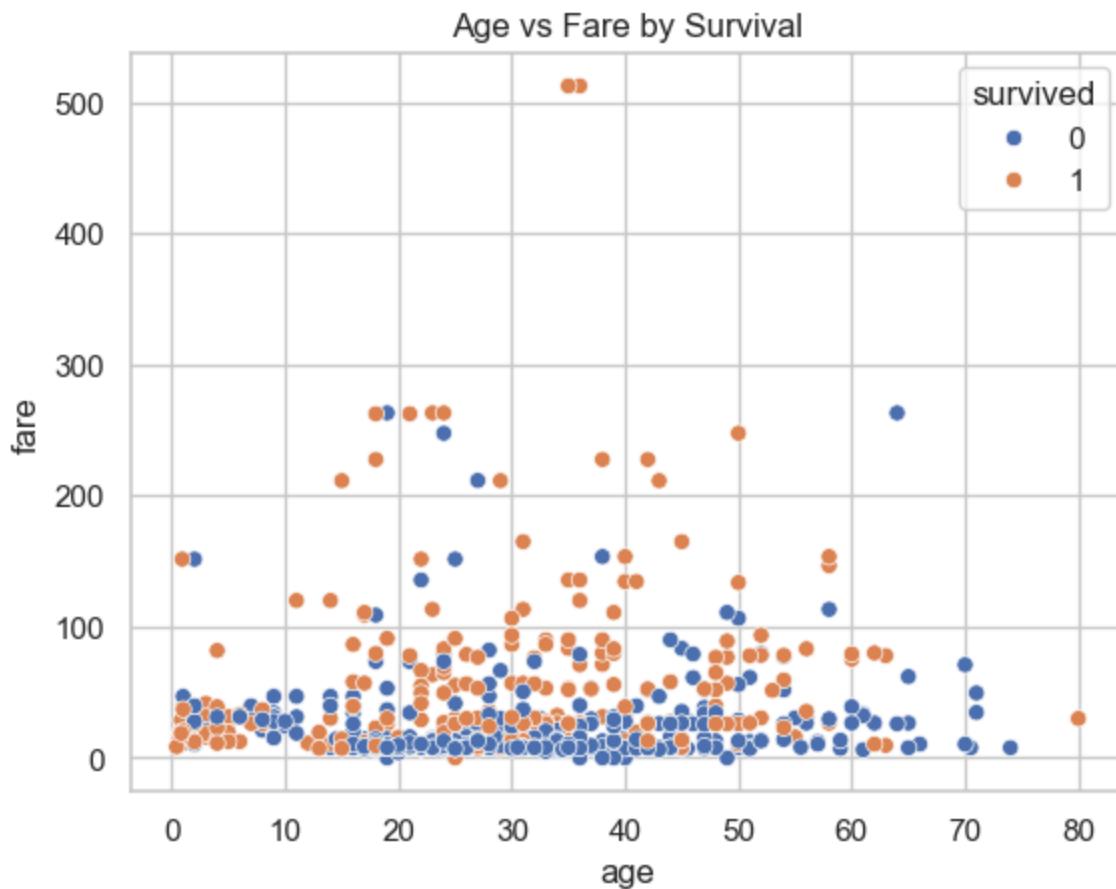
```
In [6]: plt.figure()
sns.countplot(x='survived', hue='sex', data=titanic)
plt.title("Survival Count by Gender")
plt.show()
```



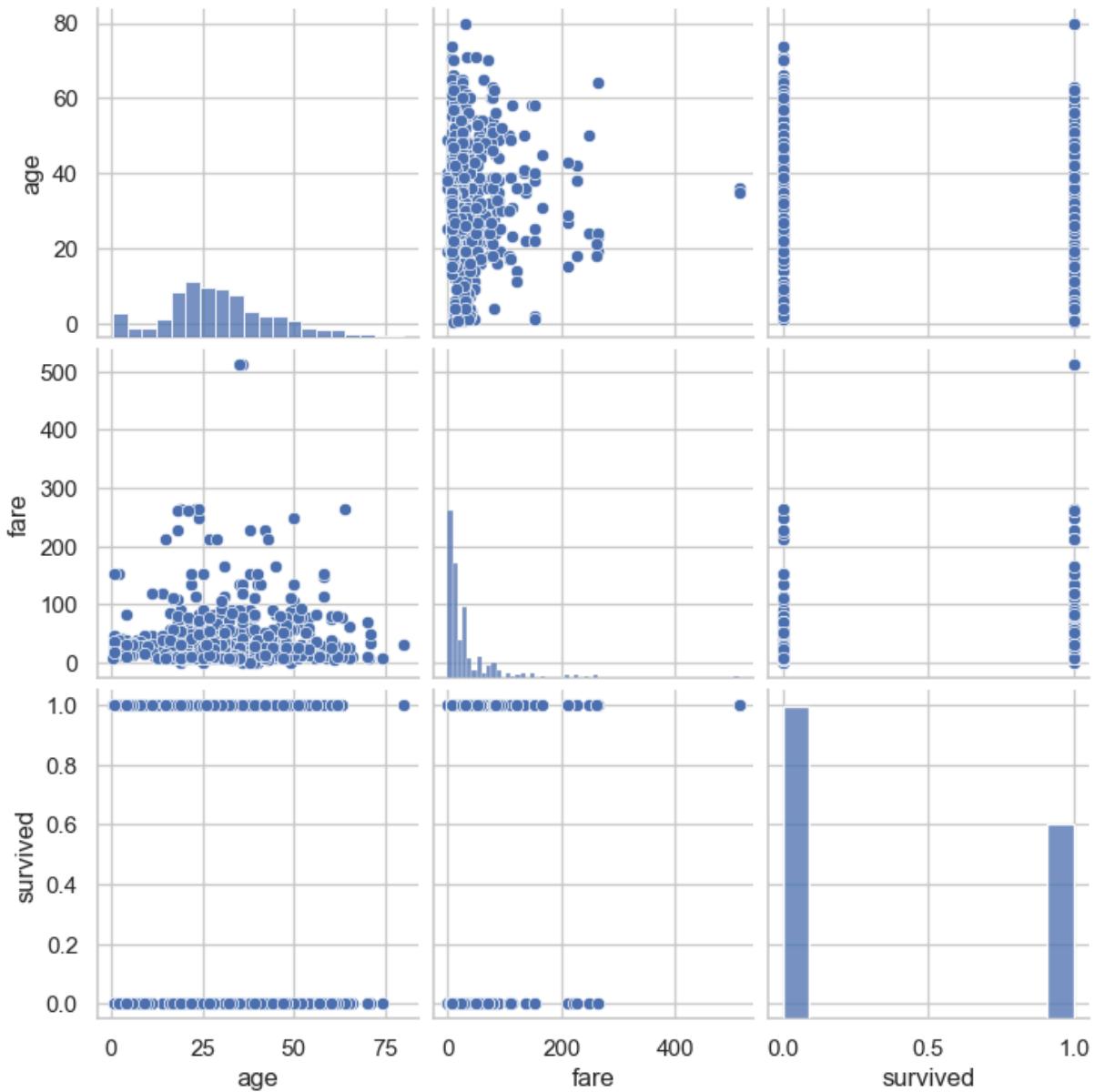
Visualization Observations

- Most passengers were between 20 and 40 years old.
- Fare distribution shows significant outliers, indicating a few passengers paid very high fares.
- Female passengers had a higher survival rate compared to male passengers.

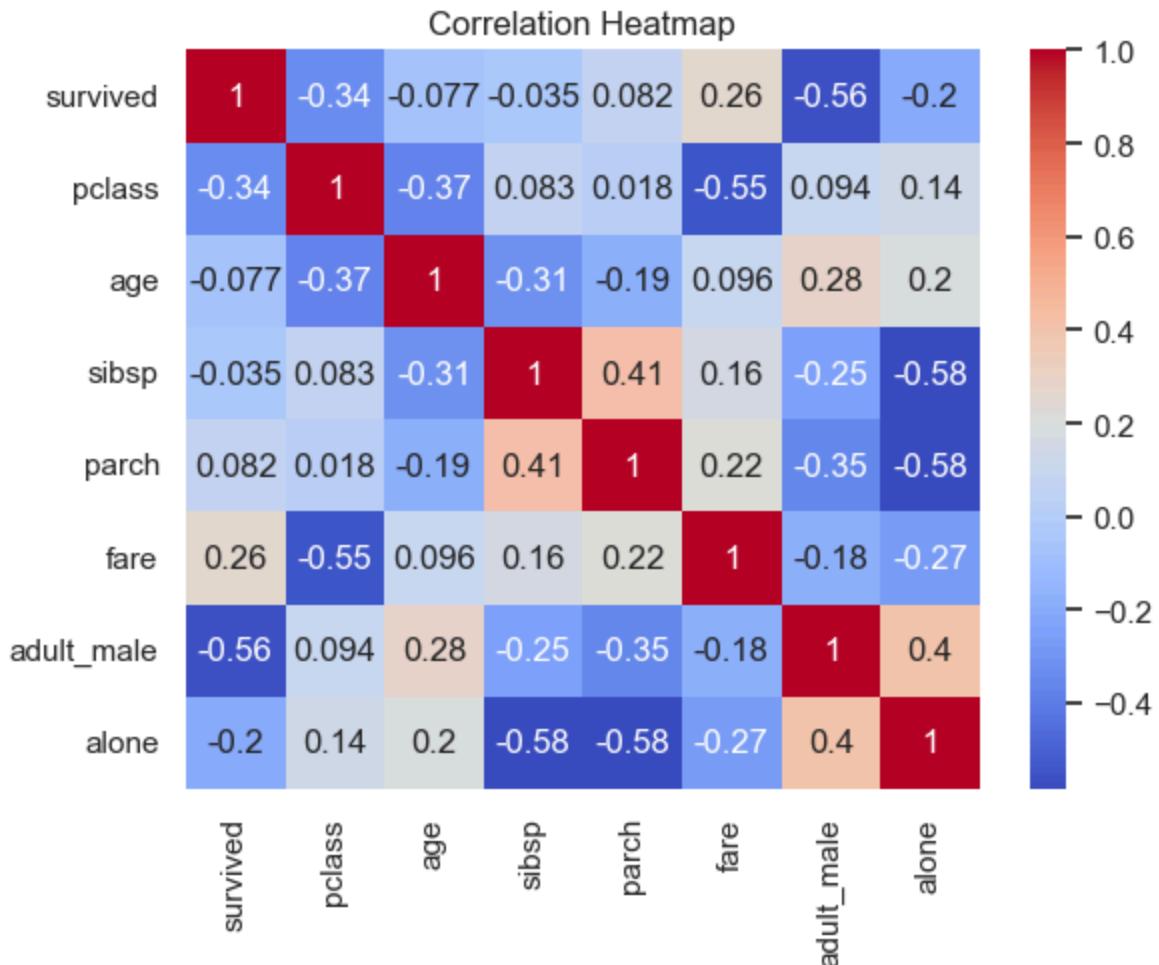
```
In [7]: plt.figure()
sns.scatterplot(x='age', y='fare', hue='survived', data=titanic)
plt.title("Age vs Fare by Survival")
plt.show()
```



```
In [8]: sns.pairplot(titanic[['age', 'fare', 'survived']])
plt.show()
```



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



Summary of Findings

- Female passengers had a significantly higher survival rate.
- Passengers who paid higher fares were more likely to survive.
- Most passengers were young adults between 20 and 40 years.
- Fare shows a positive correlation with survival.
- The dataset contains missing values, especially in the age and deck columns.

In []: