

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

# Make plots look nicer
sns.set(style='whitegrid')
```

```
In [3]: df = pd.read_csv(r"C:\Users\moham\OneDrive\Documents\tested.csv")
```

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

```
Out[4]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.82
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.00
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.68
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.66
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.28

```
In [5]: df.info() # Column names, types, missing values
df.describe() # Numbers summary (mean, min, max, etc.)
df.describe(include='object') # Text/categorical columns summary
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   PassengerId  418 non-null    int64
1   Survived     418 non-null    int64
2   Pclass       418 non-null    int64
3   Name         418 non-null    object
4   Sex          418 non-null    object
5   Age         332 non-null    float64
6   SibSp        418 non-null    int64
7   Parch        418 non-null    int64
8   Ticket       418 non-null    object
9   Fare         417 non-null    float64
10  Cabin        91 non-null     object
11  Embarked     418 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

```
Out[5]:
```

	Name	Sex	Ticket	Cabin	Embarked
<b>count</b>	418	418	418	91	418
<b>unique</b>	418	2	363	76	3
<b>top</b>	Kelly, Mr. James	male	PC 17608	B57 B59 B63 B66	S
<b>freq</b>	1	266	5	3	270

```
In [6]: df['Sex'].value_counts()
df['Pclass'].value_counts()
```

```
Out[6]: Pclass
3      218
1      107
2       93
Name: count, dtype: int64
```

```
In [7]: df.isnull().sum()
```

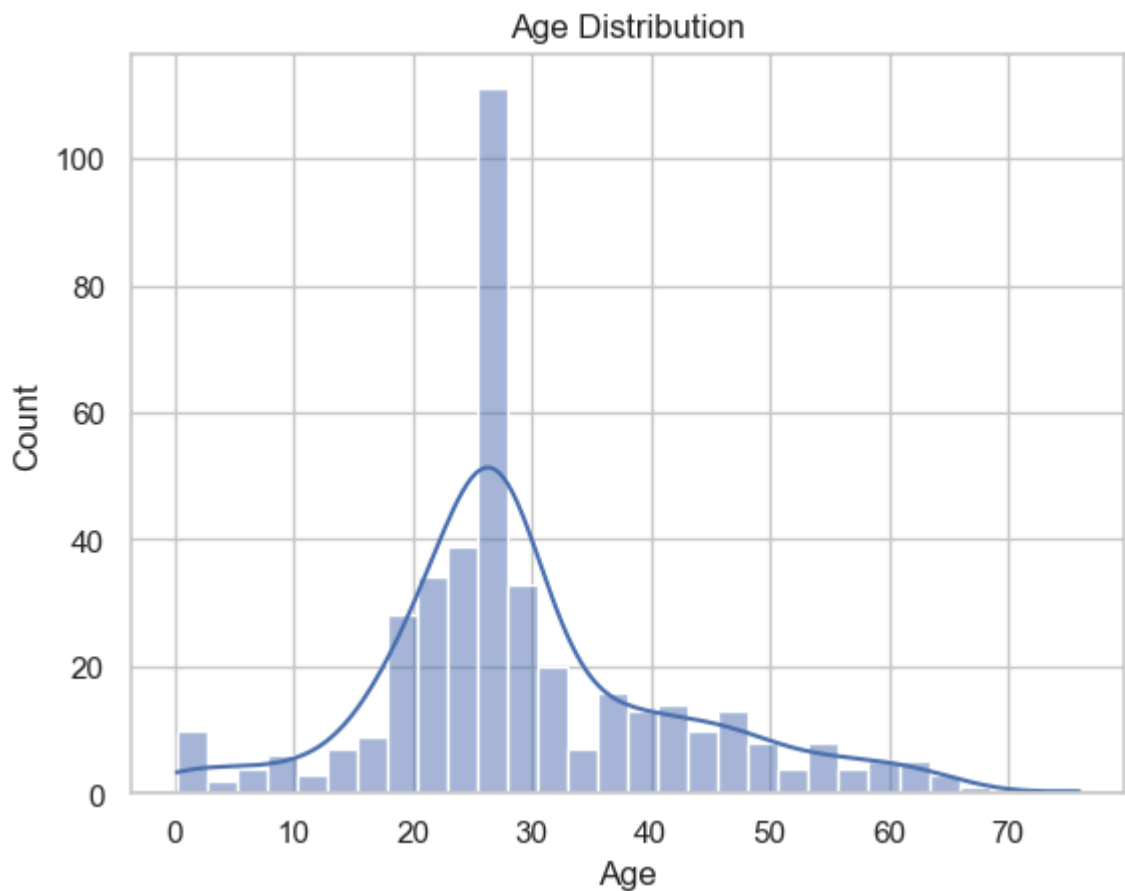
```
Out[7]: PassengerId      0
Survived                0
Pclass                  0
Name                    0
Sex                     0
Age                    86
SibSp                   0
Parch                   0
Ticket                  0
Fare                    1
Cabin                 327
Embarked                0
dtype: int64
```

```
In [8]: # Fill 'Age' with median value
median_age = df['Age'].median()
df['Age'] = df['Age'].fillna(median_age)

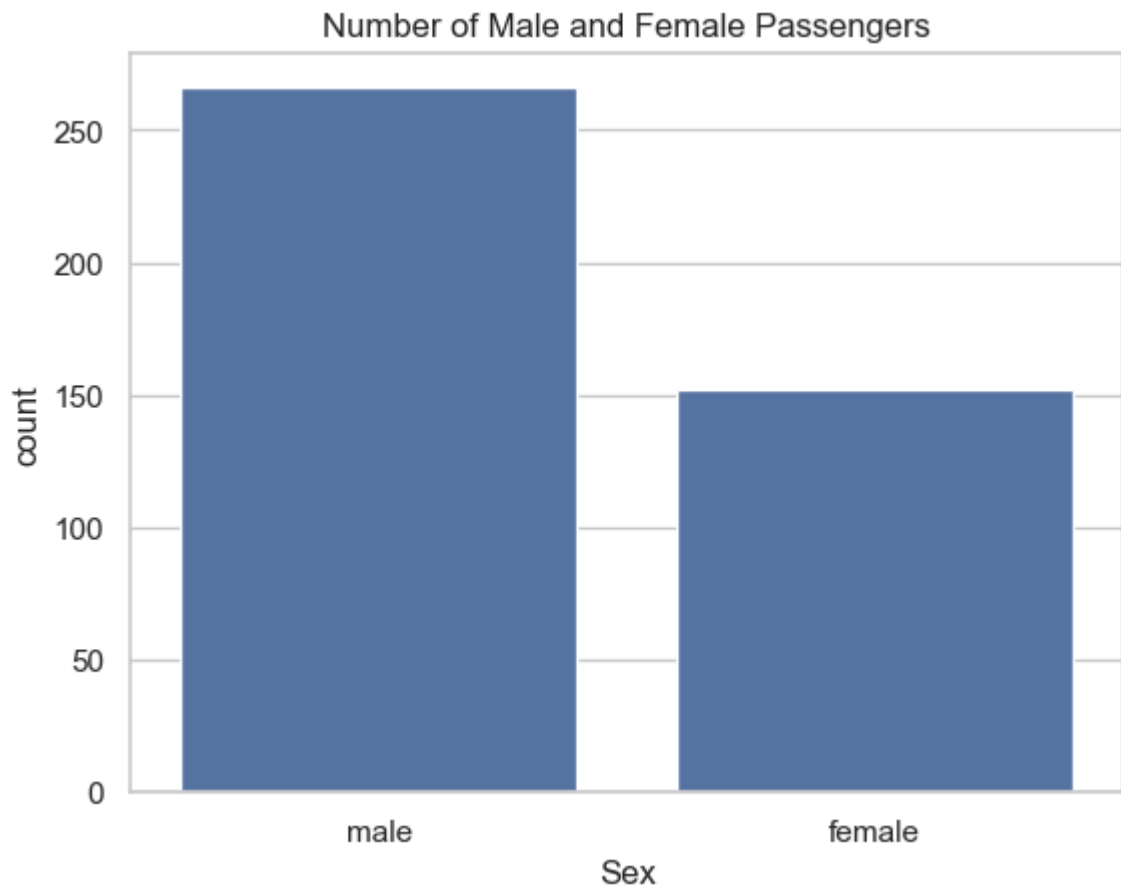
# Fill 'Embarked' with most common value (mode)
```

```
most_common_embarked = df['Embarked'].mode()[0]  
df['Embarked'] = df['Embarked'].fillna(most_common_embarked)
```

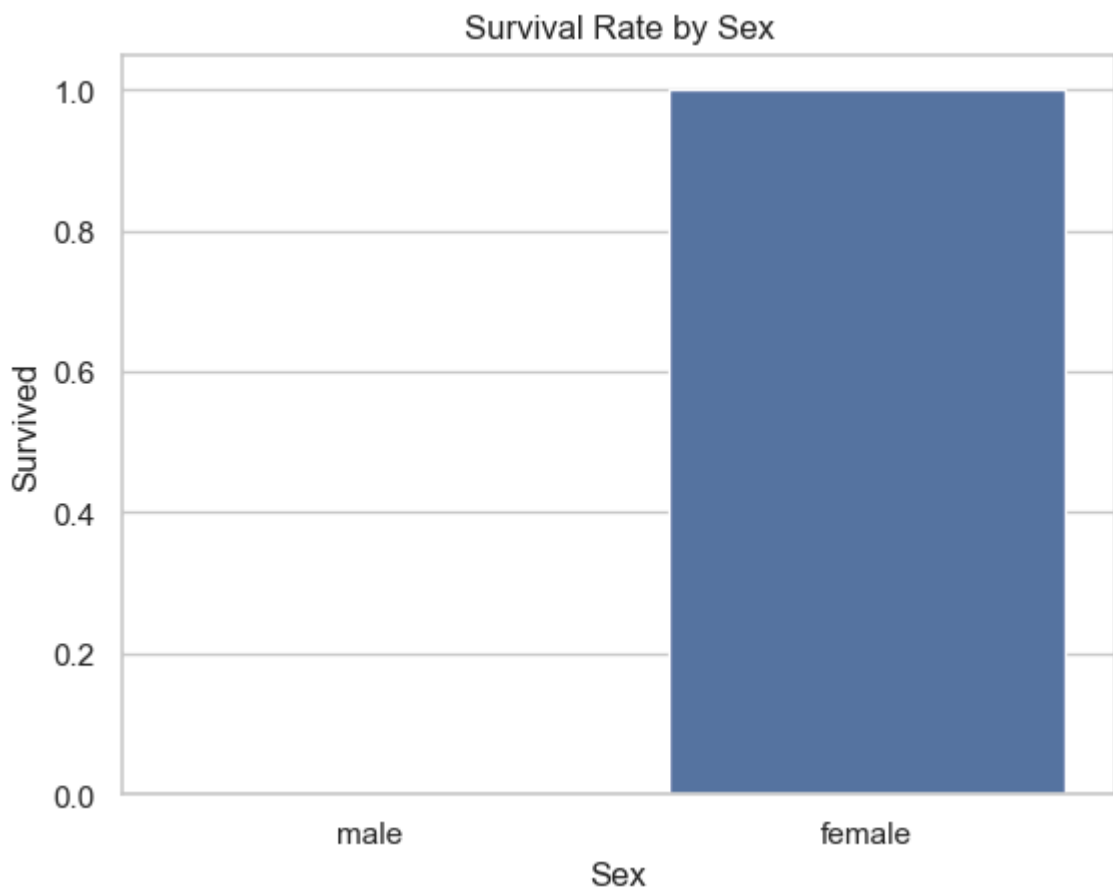
```
In [10]: sns.histplot(df['Age'], bins=30, kde=True)  
plt.title('Age Distribution')  
plt.show()
```



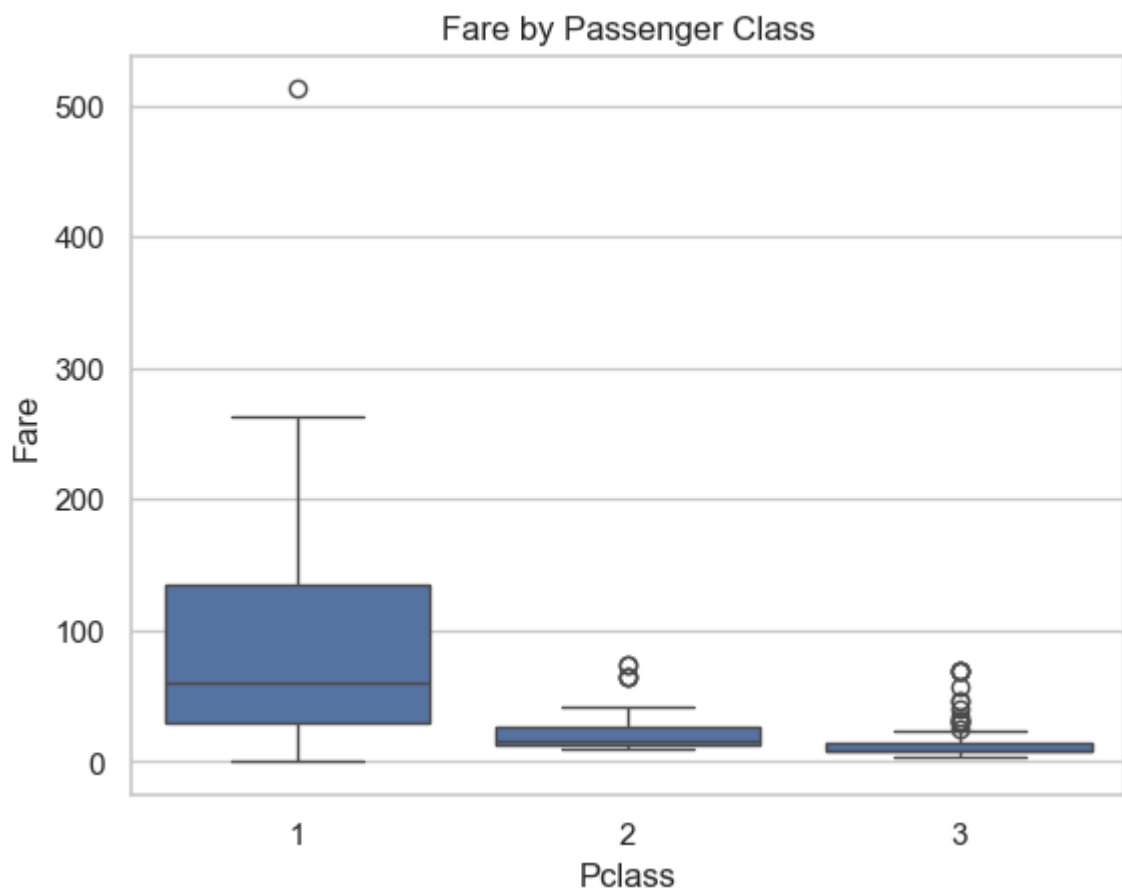
```
In [11]: sns.countplot(x='Sex', data=df)  
plt.title('Number of Male and Female Passengers')  
plt.show()
```



```
In [12]: sns.barplot(x='Sex', y='Survived', data=df)
plt.title('Survival Rate by Sex')
plt.show()
```

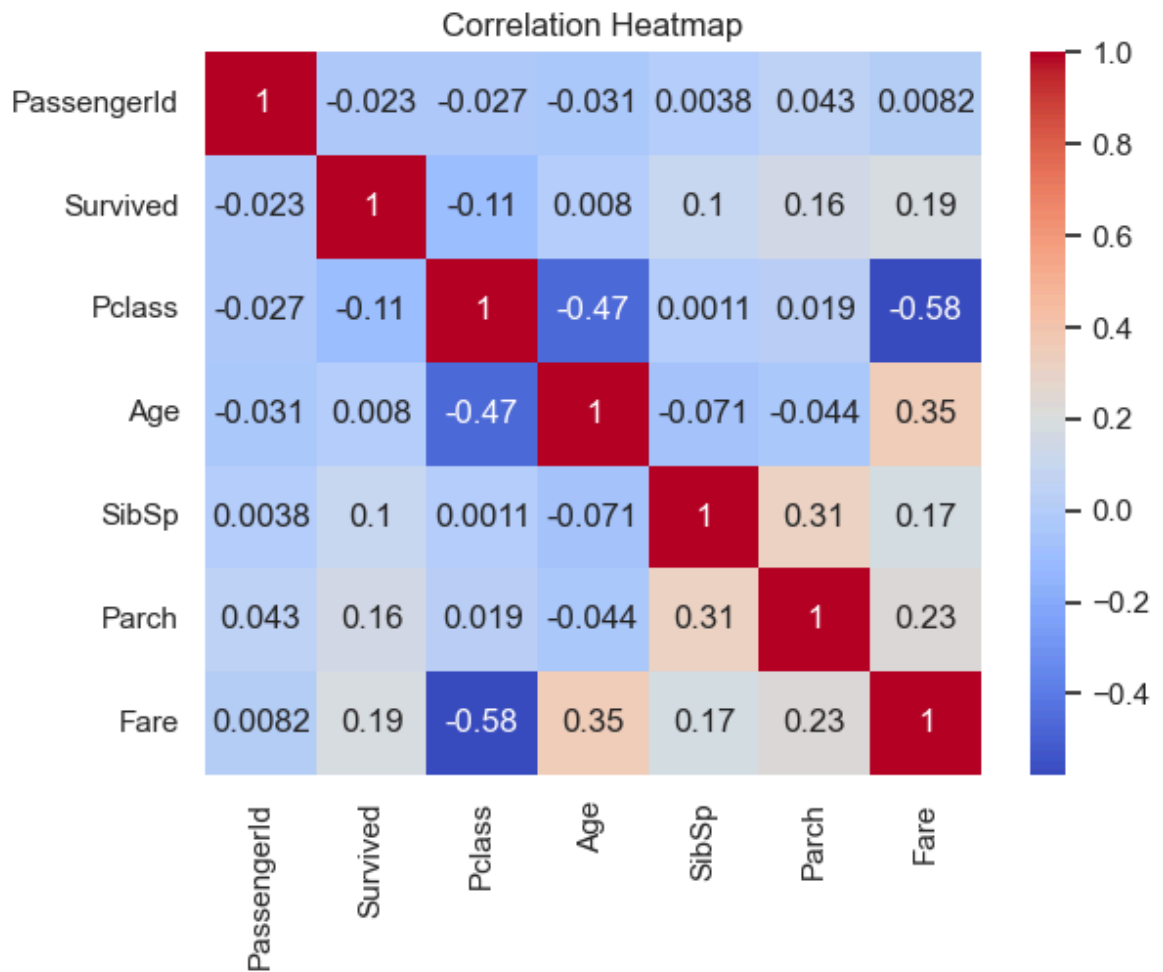


```
In [13]: sns.boxplot(x='Pclass', y='Fare', data=df)
plt.title('Fare by Passenger Class')
plt.show()
```

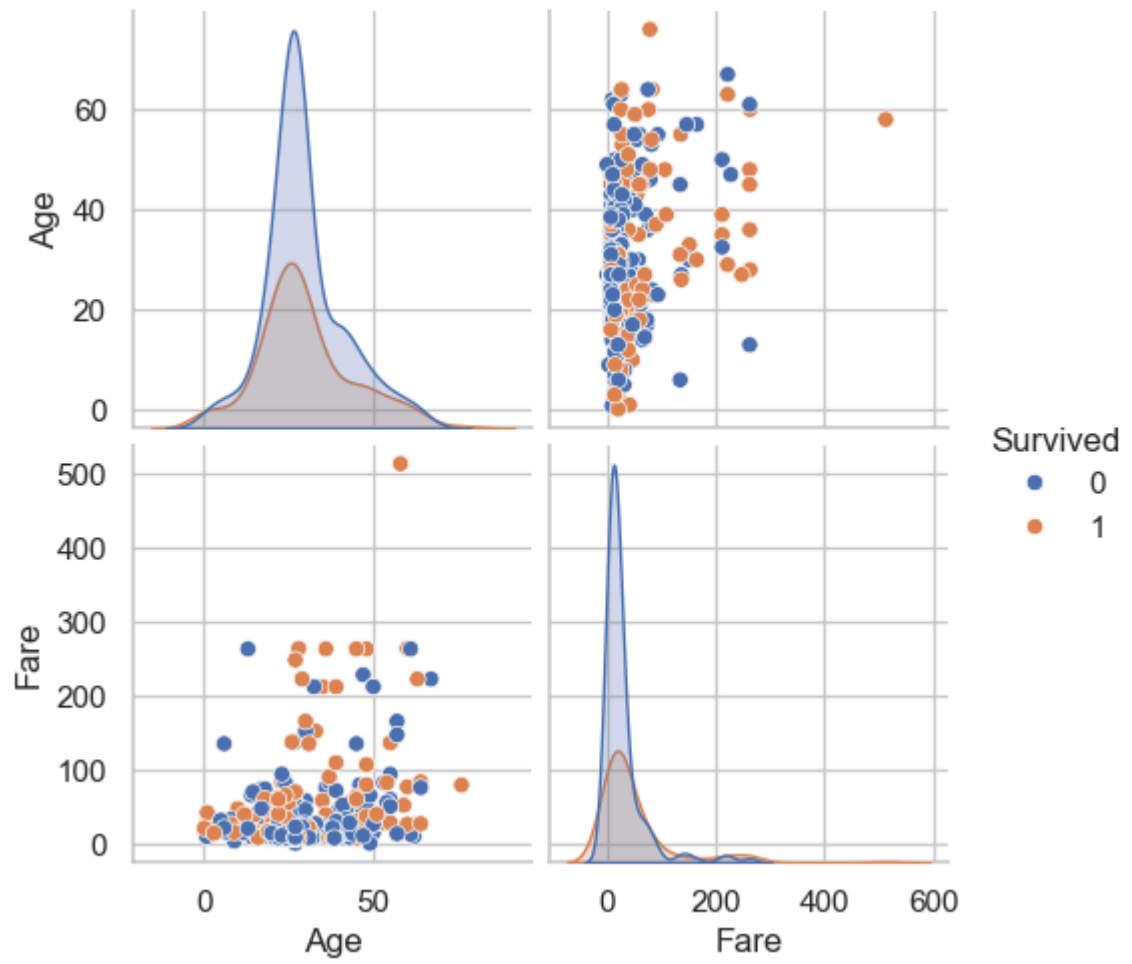


```
In [14]: # Select only numeric columns
numeric_df = df.select_dtypes(include=['number'])

# Create the heatmap
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



```
In [15]: sns.pairplot(df[['Age', 'Fare', 'Survived']], hue='Survived')  
plt.show()
```



In [ ]: