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

```python
In [ ]:   os.chdir("C:\\Users\\admin\\Desktop\\python")
```

```python
In [23]:  df = pd.read_csv("train.csv")
          df   ## Display the DataFrame
```

Out[23]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

```python
In [24]:  df.head()
```

Out[24]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```python
In [25]:  print(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
None
```
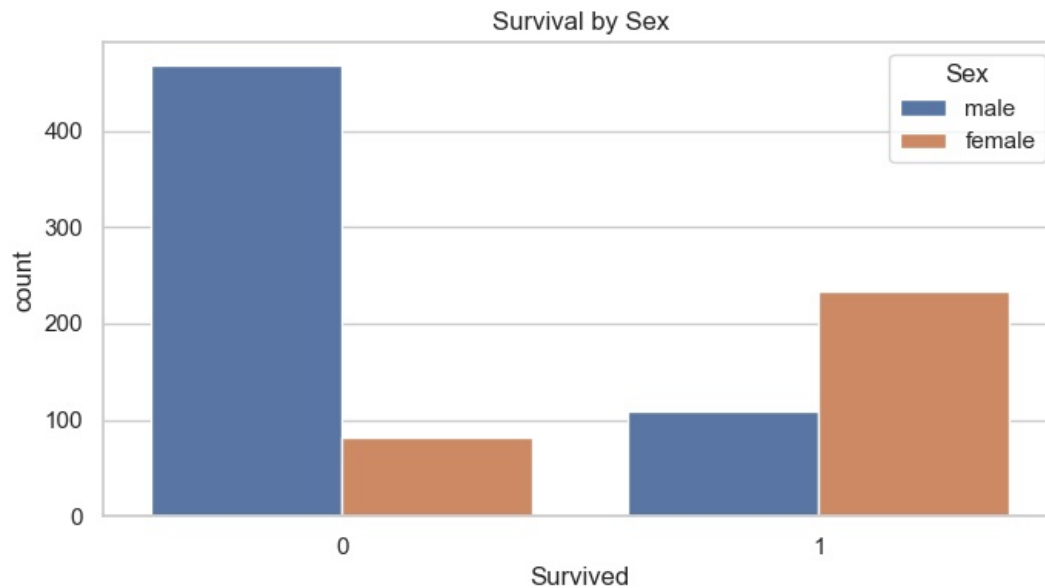
```python
In [26]:  print(df.columns)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```python
In [ ]:   df['Age'].fillna(df['Age'].median(), inplace=True)
          df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
```

```python
In [27]:  import warnings
```
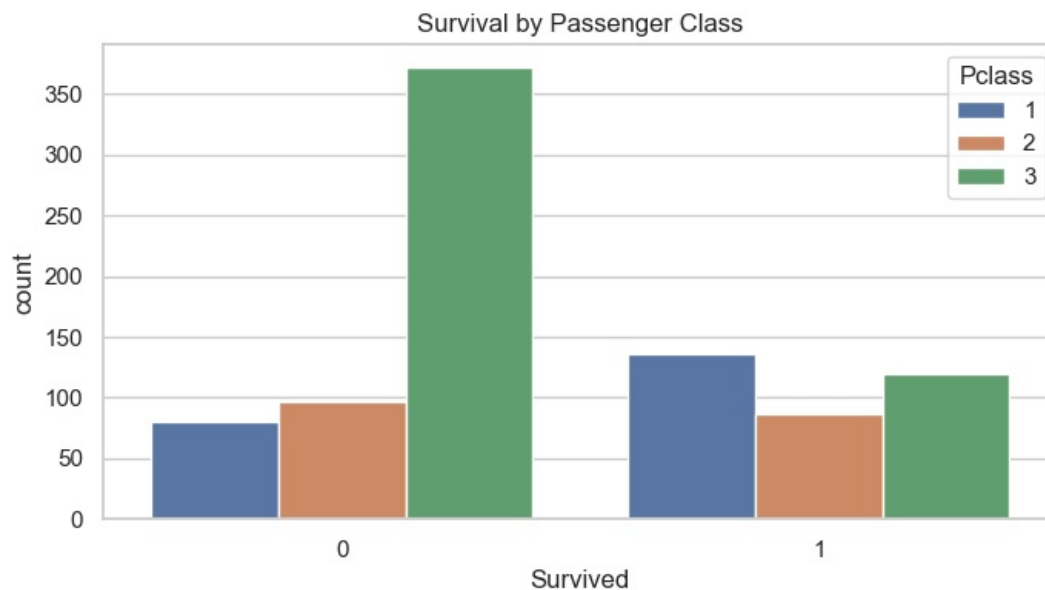
```
# Suppress FutureWarnings
warnings.simplefilter(action='ignore', category=FutureWarning)

# Survival vs. sex
# This is  Seaborn code
sns.set(style="whitegrid")
plt.figure(figsize=(8, 4))
sns.countplot(x="Survived", hue="Sex", data=df)
plt.title("Survival by Sex")
plt.show()
```
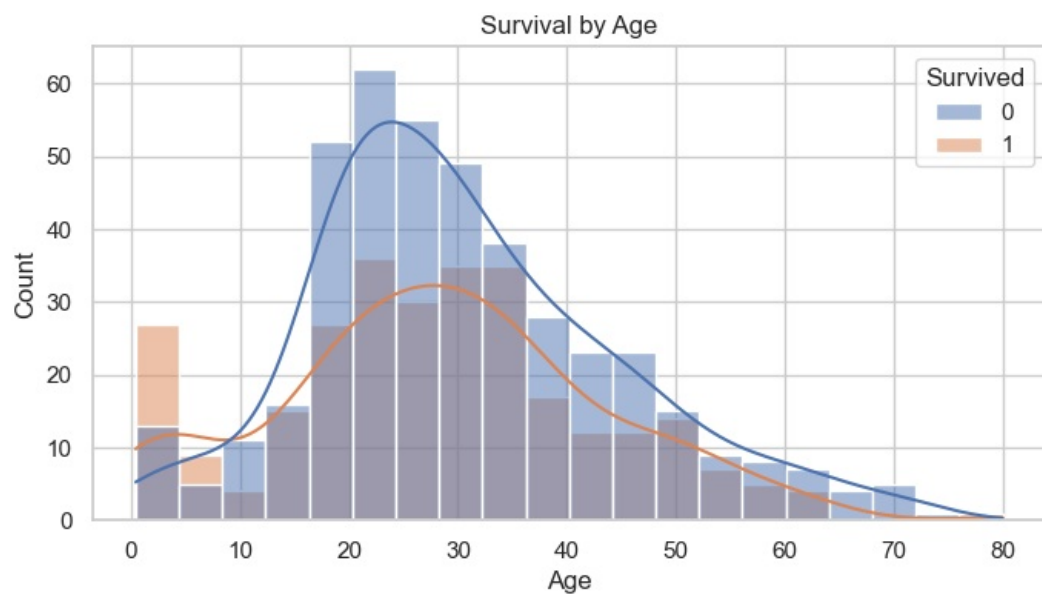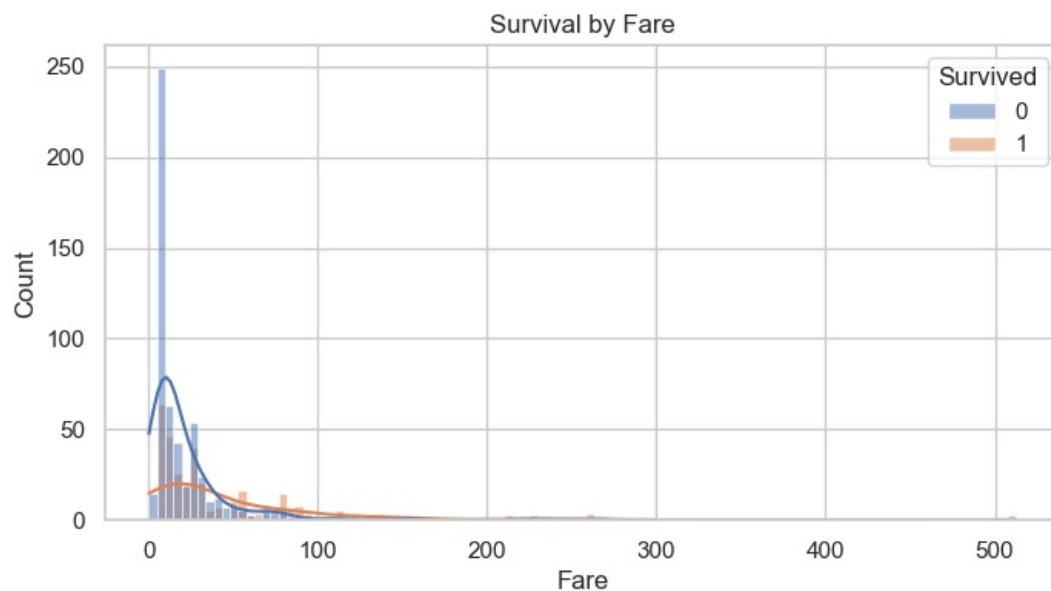

In [28]:
```
# Survival vs. passenger class
plt.figure(figsize=(8, 4))
sns.countplot(x="Survived", hue="Pclass", data=df)
plt.title("Survival by Passenger Class")
plt.show()
```


In [29]:
```
# Survival vs. age
plt.figure(figsize=(8, 4))
sns.histplot(data=df, x='Age', kde=True, hue='Survived', common_norm=False)
plt.title("Survival by Age")
plt.show()
```

Survival by Age

```python
# Survival vs. fare
plt.figure(figsize=(8, 4))
sns.histplot(data=df, x='Fare', kde=True, hue='Survived', common_norm=False)
plt.title("Survival by Fare")
plt.show()
```
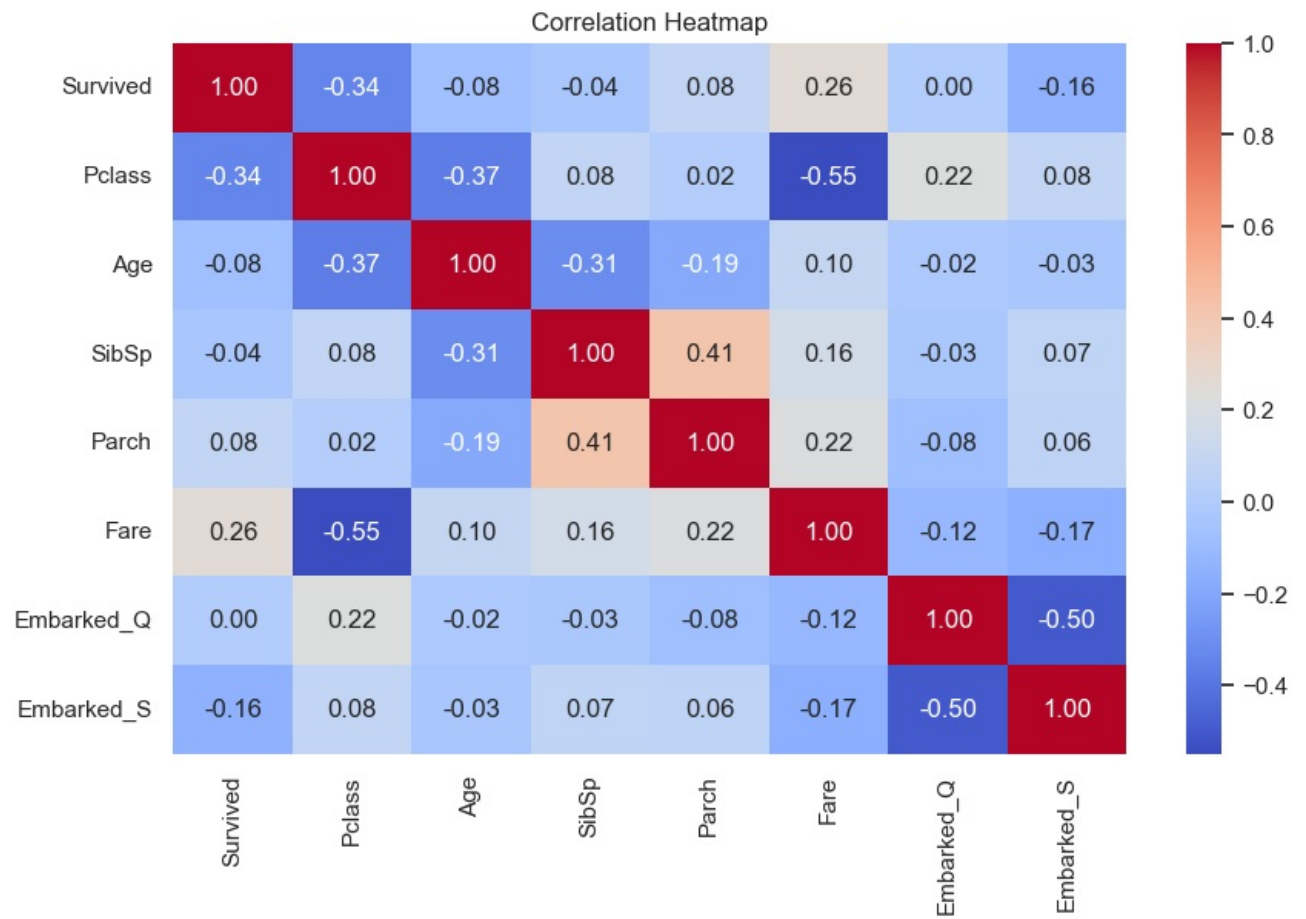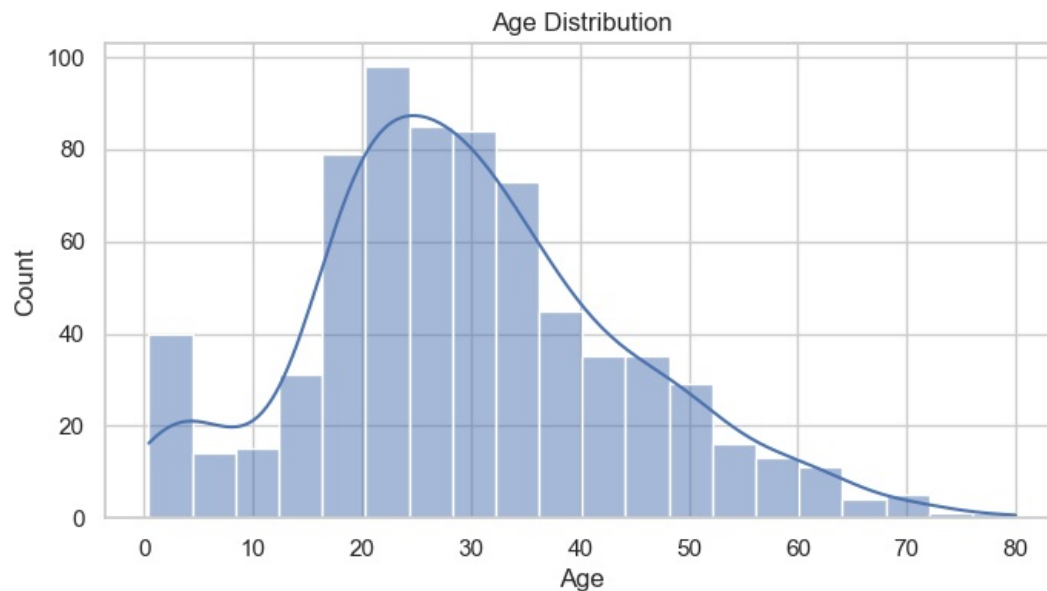

Survival by Fare

```python
# One-hot encode the 'Embarked' column
df = pd.get_dummies(df, columns=['Embarked'], drop_first=True)

# Now, run the correlation matrix
plt.figure(figsize=(10, 6))
correlation_matrix = df[['Survived', 'Pclass', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked_Q', 'Embarked_S']].co
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```
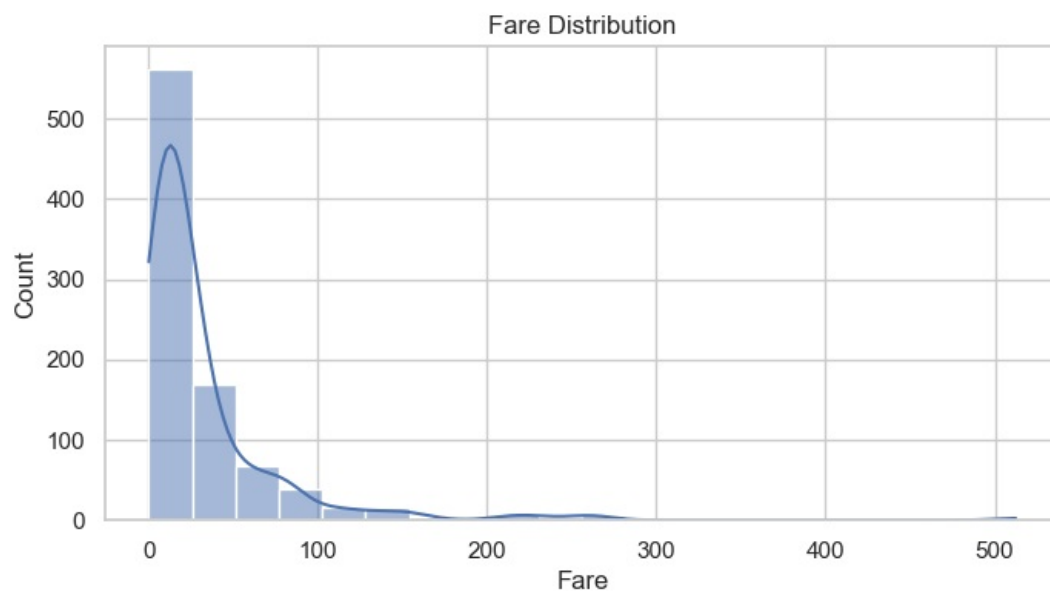
Correlation Heatmap

	Survived	Pclass	Age	SibSp	Parch	Fare	Embarked_Q	Embarked_S
Survived	1.00	-0.34	-0.08	-0.04	0.08	0.26	0.00	-0.16
Pclass	-0.34	1.00	-0.37	0.08	0.02	-0.55	0.22	0.08
Age	-0.08	-0.37	1.00	-0.31	-0.19	0.10	-0.02	-0.03
SibSp	-0.04	0.08	-0.31	1.00	0.41	0.16	-0.03	0.07
Parch	0.08	0.02	-0.19	0.41	1.00	0.22	-0.08	0.06
Fare	0.26	-0.55	0.10	0.16	0.22	1.00	-0.12	-0.17
Embarked_Q	0.00	0.22	-0.02	-0.03	-0.08	-0.12	1.00	-0.50
Embarked_S	-0.16	0.08	-0.03	0.07	0.06	-0.17	-0.50	1.00

In [32]:
```python
# Explore the distribution of passengers' ages
plt.figure(figsize=(8, 4))
sns.histplot(data=df, x='Age', kde=True, bins=20)
plt.title("Age Distribution")
plt.show()
```

Age Distribution

In [33]:
```python
# Explore the distribution of passengers' fares
plt.figure(figsize=(8, 4))
sns.histplot(data=df, x='Fare', kde=True, bins=20)
plt.title("Fare Distribution")
plt.show()
```

Fare Distribution

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js