```python
#Titanic survival data analysis by Hema Sundhar

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

import pandas as pd

# Specify the path to your CSV file
csv_file_path = r"C:\Users\Asus\Downloads\Titanic-Dataset.csv"

# Read the CSV file into a DataFrame
df = pd.read_csv(csv_file_path)

#overview of dataset
print(df.head())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3


                                                Name     Sex   Age
SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0
1
1   Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0
1
2                             Heikkinen, Miss. Laina  female  26.0
0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0
1
4                           Allen, Mr. William Henry    male  35.0
0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

```python
#overview of dataset
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

#survival count
#data visualisation
plt.figure(figsize=(6, 4))
sns.countplot(x='Survived', data=df)
plt.title('Survival Count')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()
```
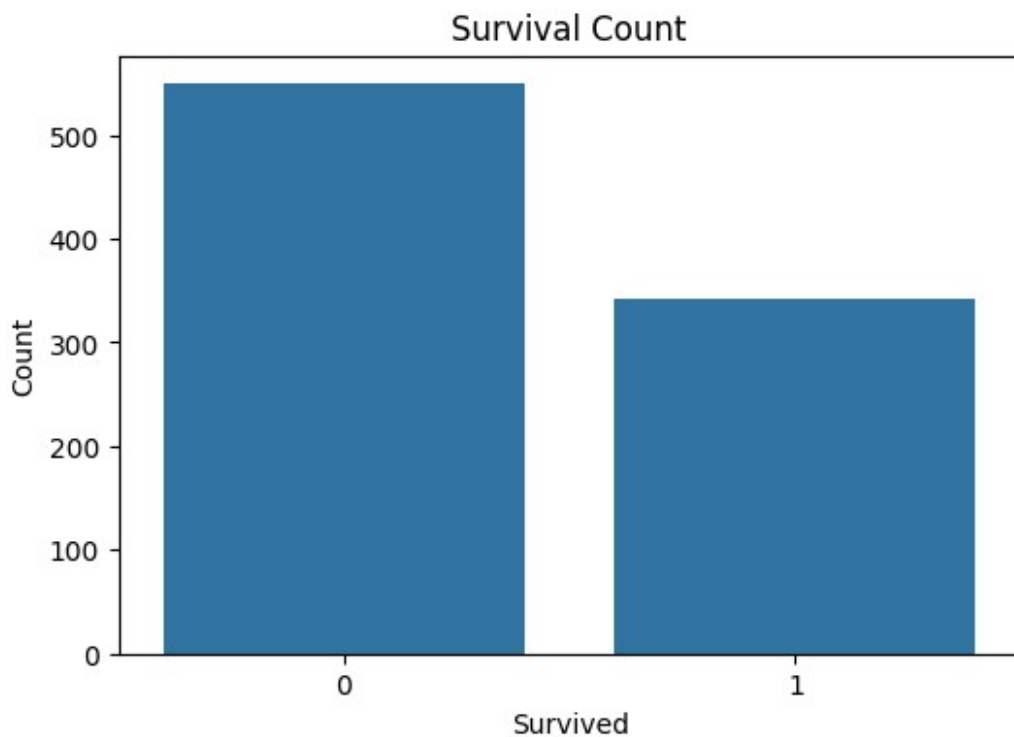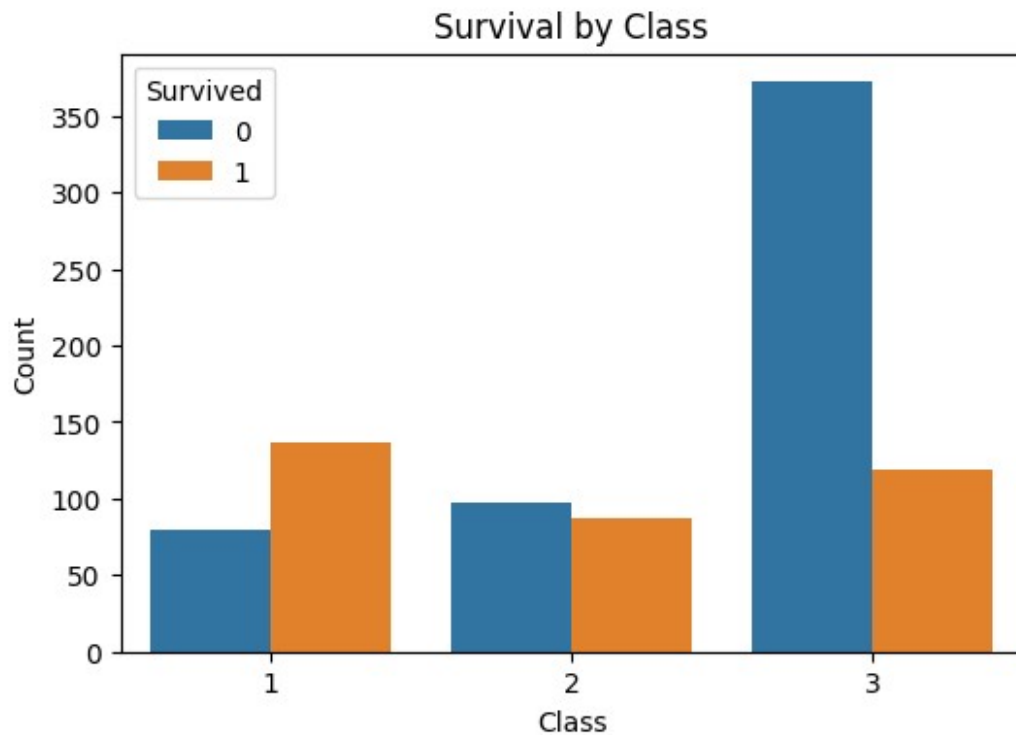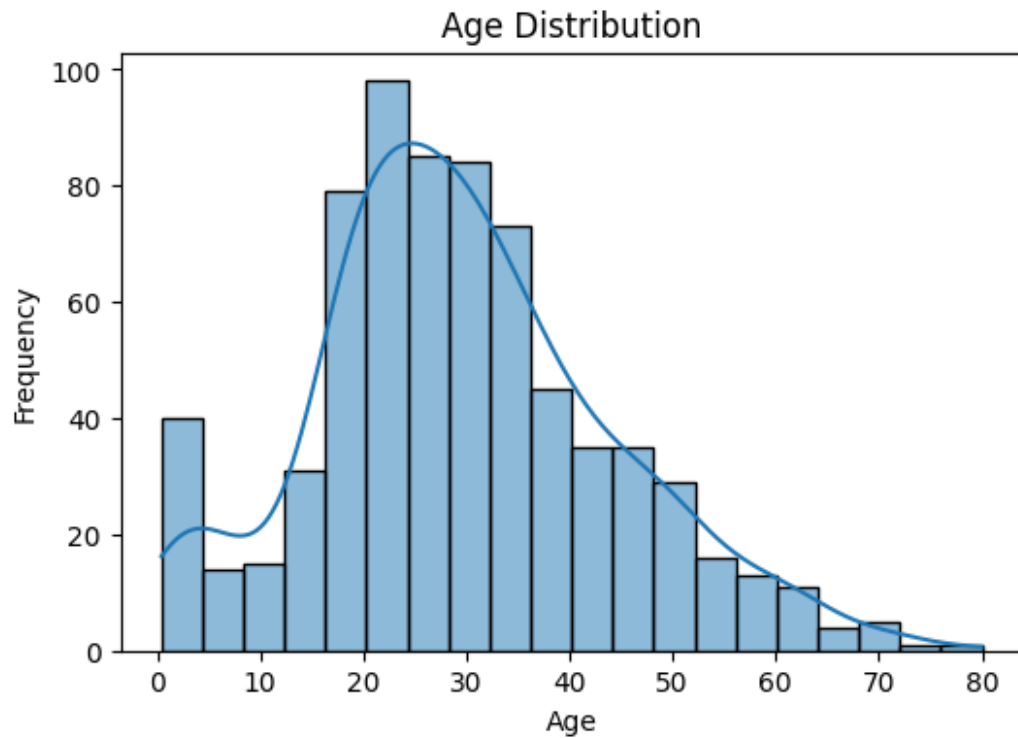
```
#survival by pclass

plt.figure(figsize=(6, 4))
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title('Survival by Class')
plt.xlabel('Class')
plt.ylabel('Count')
plt.show()
```
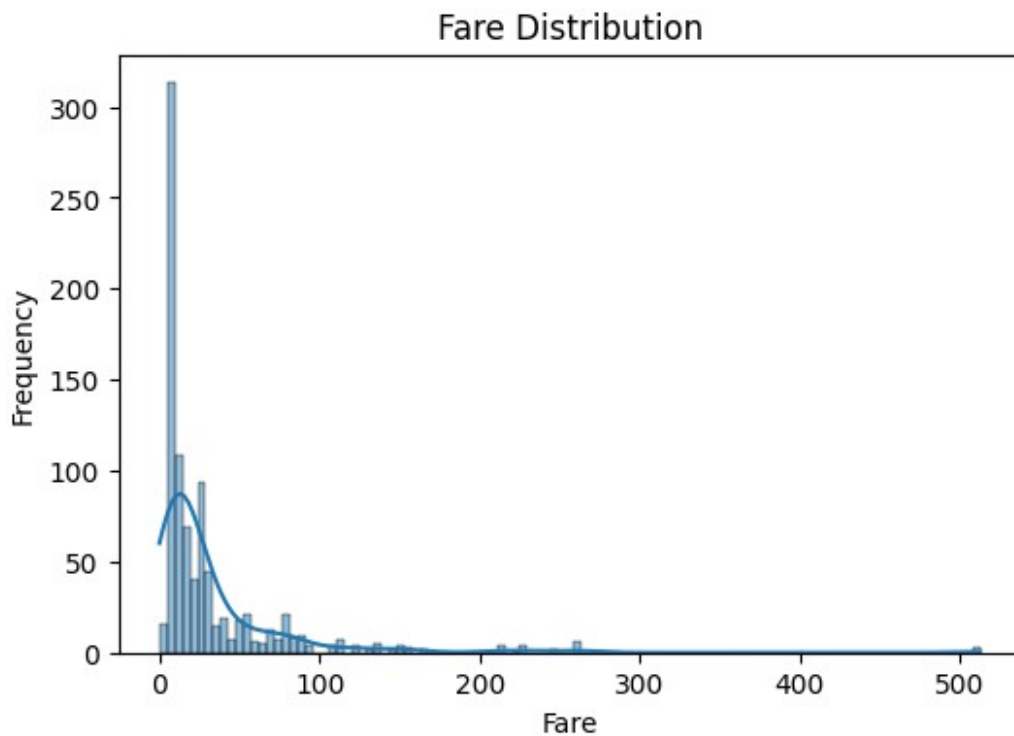


```
# Age distribution
plt.figure(figsize=(6, 4))
sns.histplot(df['Age'].dropna(), kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

Age Distribution

```python
# Fare distribution
plt.figure(figsize=(6, 4))
sns.histplot(df['Fare'], kde=True)
plt.title('Fare Distribution')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```

Fare Distribution


```python
# Drop or exclude non-numeric columns before computing correlations
numeric_data = df.select_dtypes(include=['float64', 'int64'])
correlation_matrix = numeric_data.corr()

# Heatmap of correlations
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

Correlation Heatmap