In [43]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
```

In [44]:
```
#Load the dataset
#dataset used https://www.kaggle.com/competitions/titanic/data
data = pd.read_csv('train.csv')
```

In [45]:
```
data
```

Out[45]:

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

891 rows × 12 columns

In [46]:
```
data.head()
```

Out[46]:

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

In [47]:
```
data.head(10)
```

Out[47]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	I
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	I
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	I
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	I
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	I
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	I
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	I

In [48]:
```
data.tail()
```

Out[48]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN

In [49]:
```
data.tail(10)
```

Out[49]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500

In [50]:
```
# Data Preprocessing

#display information about data set
data.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
```

In [51]:
```
data.columns
```

Out[51]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [52]:
```
data.describe()
```

Out[52]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [53]:
```
data.describe(include='all')
```

Out[53]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891.000000
unique	NaN	NaN	NaN	891	2	NaN	NaN	NaN
top	NaN	NaN	NaN	Braund, Mr. Owen Harris	male	NaN	NaN	NaN
freq	NaN	NaN	NaN	1	577	NaN	NaN	NaN
mean	446.000000	0.383838	2.308642	NaN	NaN	29.699118	0.523008	0.381594
std	257.353842	0.486592	0.836071	NaN	NaN	14.526497	1.102743	0.806057
min	1.000000	0.000000	1.000000	NaN	NaN	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	NaN	NaN	20.125000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	NaN	NaN	28.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	NaN	NaN	38.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	NaN	NaN	80.000000	8.000000	6.000000

In [54]:
```
data.shape
```

Out[54]:
```
(891, 12)
```

In [55]:
```
data.dtypes
```

Out[55]:
```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

In [56]:
```
data.index
```

Out[56]:
```
RangeIndex(start=0, stop=891, step=1)
```

In [57]:
```
# Check The Missing Value in data using pandas isnull()
data.isnull()
```

Out[57]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Er
0	False	False	False	False	False	False	False	False	False	False	True	
1	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	True	
3	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	True	
...	...	...	...	...	...	...	...	...	...	...	...	
886	False	False	False	False	False	False	False	False	False	False	True	
887	False	False	False	False	False	False	False	False	False	False	False	
888	False	False	False	False	False	True	False	False	False	False	True	
889	False	False	False	False	False	False	False	False	False	False	False	
890	False	False	False	False	False	False	False	False	False	False	True	

891 rows × 12 columns

In [58]:
```
data.isnull().any()
```

Out[58]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

In [59]:
```
data.isnull().sum()
```

Out[59]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [60]:
```
data['Age'].fillna(data['Age'].mean(),inplace=True)
```

In [61]:
```
data.isnull().sum()
```

Out[61]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

Visualization

In [62]:
```
data['Name']
```

Out[62]:
```
0                                Braund, Mr. Owen Harris
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
2                                 Heikkinen, Miss. Laina
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
4                               Allen, Mr. William Henry
                             ...
886                                Montvila, Rev. Juozas
887                         Graham, Miss. Margaret Edith
888             Johnston, Miss. Catherine Helen "Carrie"
889                                Behr, Mr. Karl Howell
890                                  Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object
```

In [63]:
```
data['Sex'].value_counts()
```

Out[63]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [64]:
```
data['Ticket'].value_counts()
```

Out[64]:
```
347082      7
CA. 2343    7
1601        7
3101295     6
CA 2144     6
           ..
9234        1
19988       1
2693        1
PC 17612    1
370376      1
Name: Ticket, Length: 681, dtype: int64
```

In [65]:
```
data['Cabin'].value_counts()
```

Out[65]:
```
B96 B98        4
G6             4
C23 C25 C27    4
C22 C26        3
F33            3
              ..
E34            1
C7             1
C54            1
E36            1
C148           1
Name: Cabin, Length: 147, dtype: int64
```

In [66]:
```
data['Embarked'].value_counts()
```

Out[66]:
```
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [67]:
```
def fun1(value):
    if (value == "male"):
        return 1
    else:
        return 0
```

In [68]:
```
def fun2(value):
    if (value == 'S'):
        return 0
    elif (value == 'C'):
        return 1
    elif (value == 'Q'):
        return 2
    else:
        return 0
```

In [69]:
```
data["Sex"] = data["Sex"].apply(fun1)
```

In [70]:
```
data["Embarked"] = data["Embarked"].apply(fun2)
```

In [71]:
```
data.isnull().sum()
```

Out[71]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         0
dtype: int64
```

In [73]:
```
data.columns
```

Out[73]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [74]:
```
data = data.drop("Cabin", axis=1)
```

In [75]:
```
data.columns
```

Out[75]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Embarked'],
      dtype='object')
```

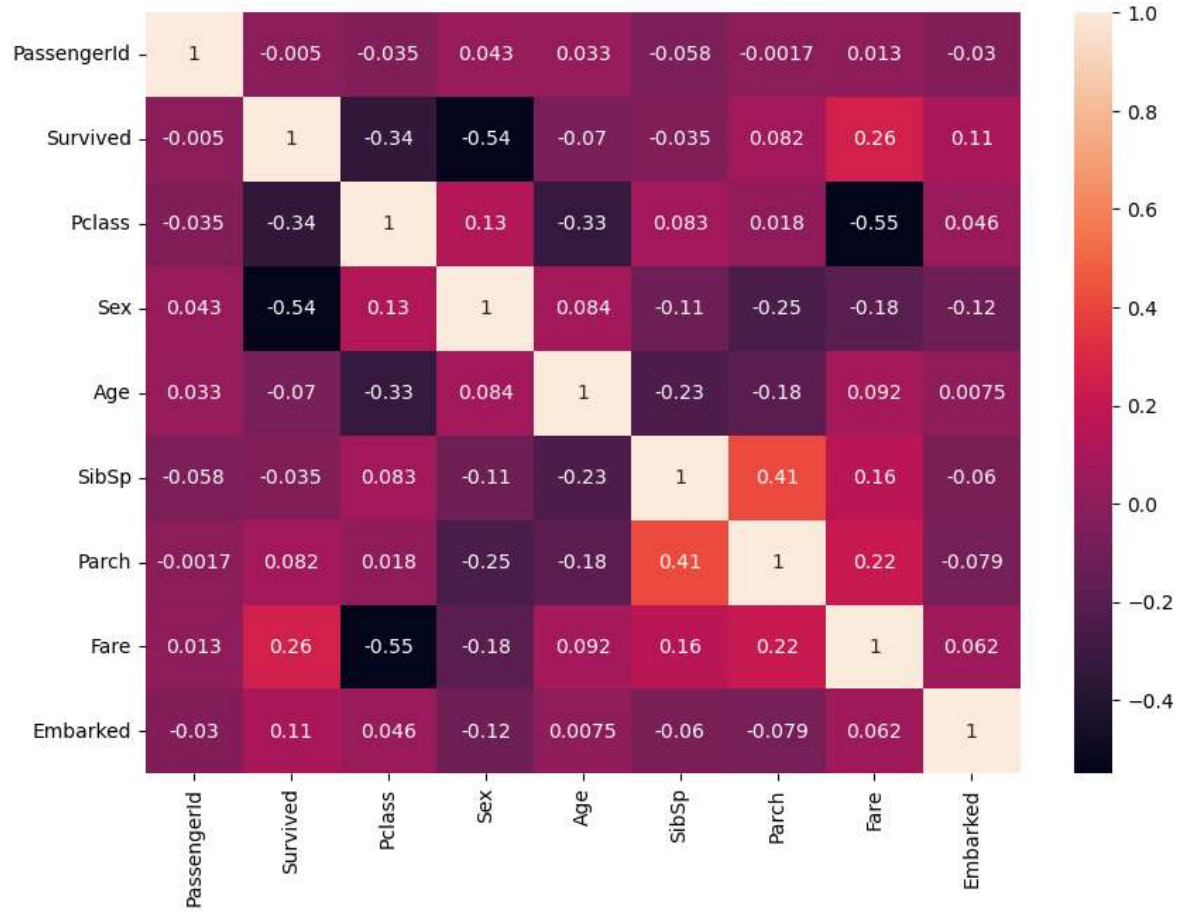
In [77]:
```
data.isnull().sum()
```

Out[77]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [78]:
```
data.shape
```

Out[78]:
```
(891, 11)
```

In [80]:
```
plt.figure(figsize=(10,7))
sns.heatmap(data.corr(), annot=True)
plt.show()
```



In [81]:
```
#From the above corelation matrix, it is clear that 'Fare' and 'Survived' have
#Meaning higher the cost of the ticket, higher is the chance of survival.
```

In [82]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 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    int64  
 5   Age          891 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  Embarked     891 non-null    int64  
dtypes: float64(2), int64(7), object(2)
memory usage: 76.7+ KB
```

In [90]:
```
data.head(50)
```

Out[90]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	1	22.000000	1	0	A/5 21171	7.
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.000000	1	0	PC 17599	71.
2	3	1	3	Heikkinen, Miss. Laina	0	26.000000	0	0	STON/O2. 3101282	7.
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.000000	1	0	113803	53.
4	5	0	3	Allen, Mr. William Henry	1	35.000000	0	0	373450	8.
5	6	0	3	Moran, Mr. James	1	29.699118	0	0	330877	8.
6	7	0	1	McCarthy, Mr. Timothy J	1	54.000000	0	0	17463	51.
7	8	0	3	Palsson, Master. Gosta Leonard	1	2.000000	3	1	349909	21.
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	0	27.000000	0	2	347742	11.
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	0	14.000000	1	0	237736	30.
10	11	1	3	Sandstrom, Miss. Marguerite Rut	0	4.000000	1	1	PP 9549	16.
11	12	1	1	Bonnell, Miss. Elizabeth	0	58.000000	0	0	113783	26.
12	13	0	3	Saundercock, Mr. William Henry	1	20.000000	0	0	A/5. 2151	8.
13	14	0	3	Andersson, Mr. Anders Johan	1	39.000000	1	5	347082	31.
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	0	14.000000	0	0	350406	7.
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	0	55.000000	0	0	248706	16.

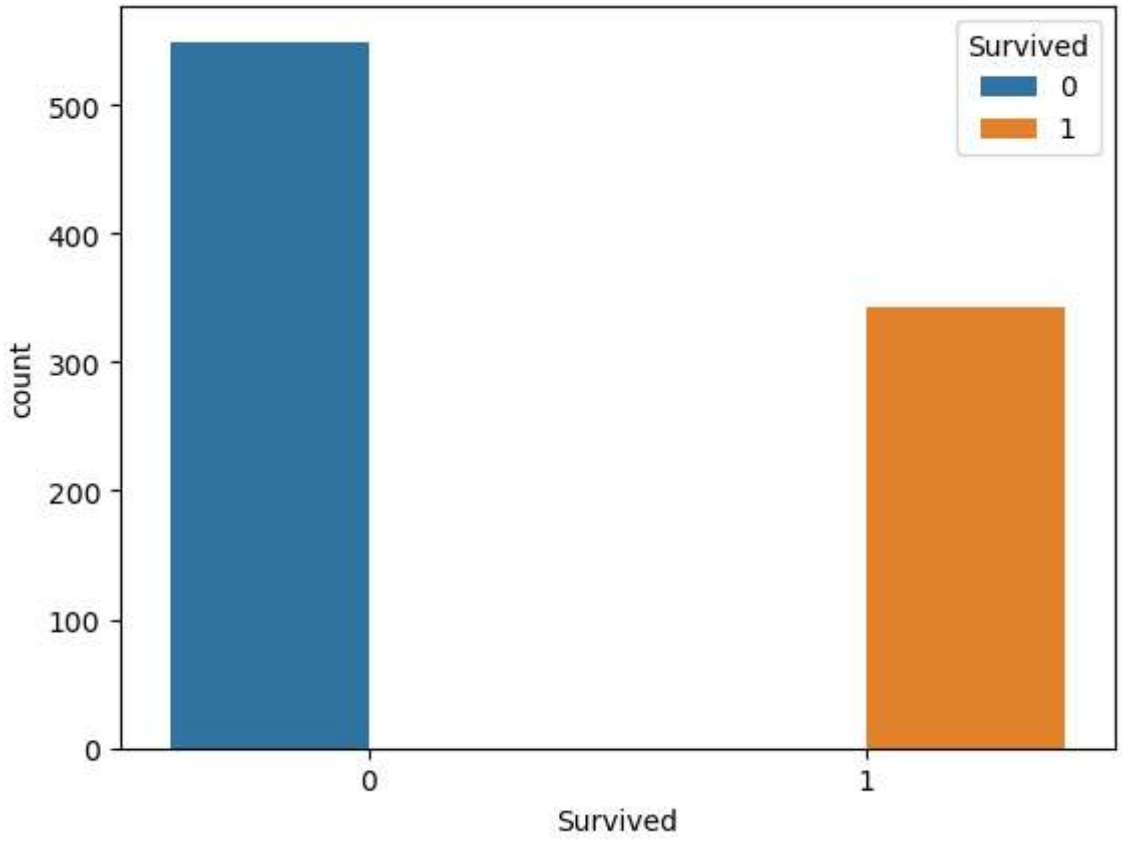
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
16	17	0	3	Rice, Master. Eugene	1	2.000000	4	1	382652	29.
17	18	1	2	Williams, Mr. Charles Eugene	1	29.699118	0	0	244373	13.
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	0	31.000000	1	0	345763	18.
19	20	1	3	Masselmani, Mrs. Fatima	0	29.699118	0	0	2649	7.
20	21	0	2	Fynney, Mr. Joseph J	1	35.000000	0	0	239865	26.
21	22	1	2	Beesley, Mr. Lawrence	1	34.000000	0	0	248698	13.
22	23	1	3	McGowan, Miss. Anna "Annie"	0	15.000000	0	0	330923	8.
23	24	1	1	Sloper, Mr. William Thompson	1	28.000000	0	0	113788	35.
24	25	0	3	Palsson, Miss. Torborg Danira	0	8.000000	3	1	349909	21.
25	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	0	38.000000	1	5	347077	31.
26	27	0	3	Emir, Mr. Farred Chehab	1	29.699118	0	0	2631	7.
27	28	0	1	Fortune, Mr. Charles Alexander	1	19.000000	3	2	19950	263.
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	0	29.699118	0	0	330959	7.
29	30	0	3	Todoroff, Mr. Lalio	1	29.699118	0	0	349216	7.
30	31	0	1	Uruchurtu, Don. Manuel E	1	40.000000	0	0	PC 17601	27.
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	0	29.699118	1	0	PC 17569	146.
32	33	1	3	Glynn, Miss. Mary Agatha	0	29.699118	0	0	335677	7.

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
33	34	0	2	Wheadon, Mr. Edward H	1	66.000000	0	0	C.A. 24579	10.
34	35	0	1	Meyer, Mr. Edgar Joseph	1	28.000000	1	0	PC 17604	82.
35	36	0	1	Holverson, Mr. Alexander Oskar	1	42.000000	1	0	113789	52.
36	37	1	3	Mamee, Mr. Hanna	1	29.699118	0	0	2677	7.
37	38	0	3	Cann, Mr. Ernest Charles	1	21.000000	0	0	A./5. 2152	8.
38	39	0	3	Vander Planke, Miss. Augusta Maria	0	18.000000	2	0	345764	18.
39	40	1	3	Nicola-Yarred, Miss. Jamila	0	14.000000	1	0	2651	11.
40	41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	0	40.000000	1	0	7546	9.
41	42	0	2	Turpin, Mrs. William John Robert (Dorothy Ann ...	0	27.000000	1	0	11668	21.
42	43	0	3	Kraeff, Mr. Theodor	1	29.699118	0	0	349253	7.
43	44	1	2	Laroche, Miss. Simonne Marie Anne Andree	0	3.000000	1	2	SC/Paris 2123	41.
44	45	1	3	Devaney, Miss. Margaret Delia	0	19.000000	0	0	330958	7.
45	46	0	3	Rogers, Mr. William John	1	29.699118	0	0	S.C./A.4. 23567	8.
46	47	0	3	Lennon, Mr. Denis	1	29.699118	1	0	370371	15.
47	48	1	3	O'Driscoll, Miss. Bridget	0	29.699118	0	0	14311	7.
48	49	0	3	Samaan, Mr. Youssef	1	29.699118	2	0	2662	21.

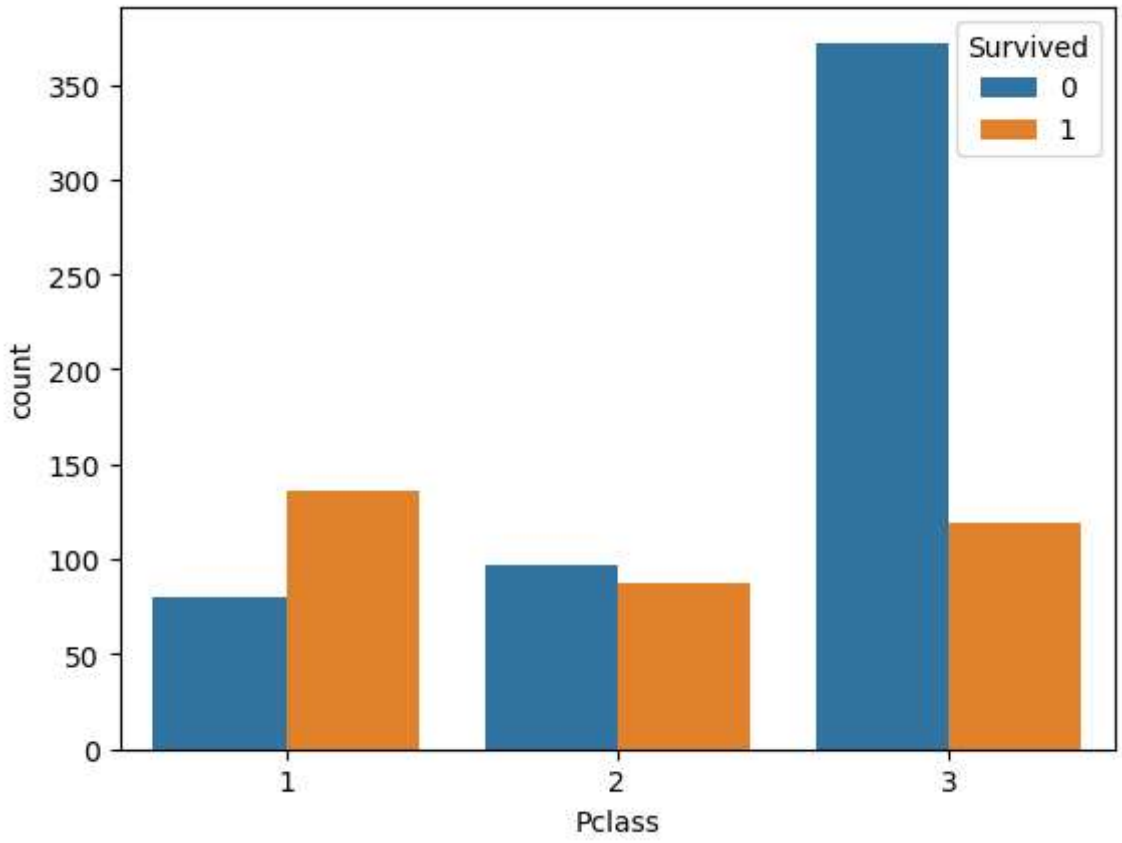
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
49	50	0	3	Arnold-Franchi, Mrs. Josef (Josefine Franchi)	0	18.000000	1	0	349237	17.

Survived is the label

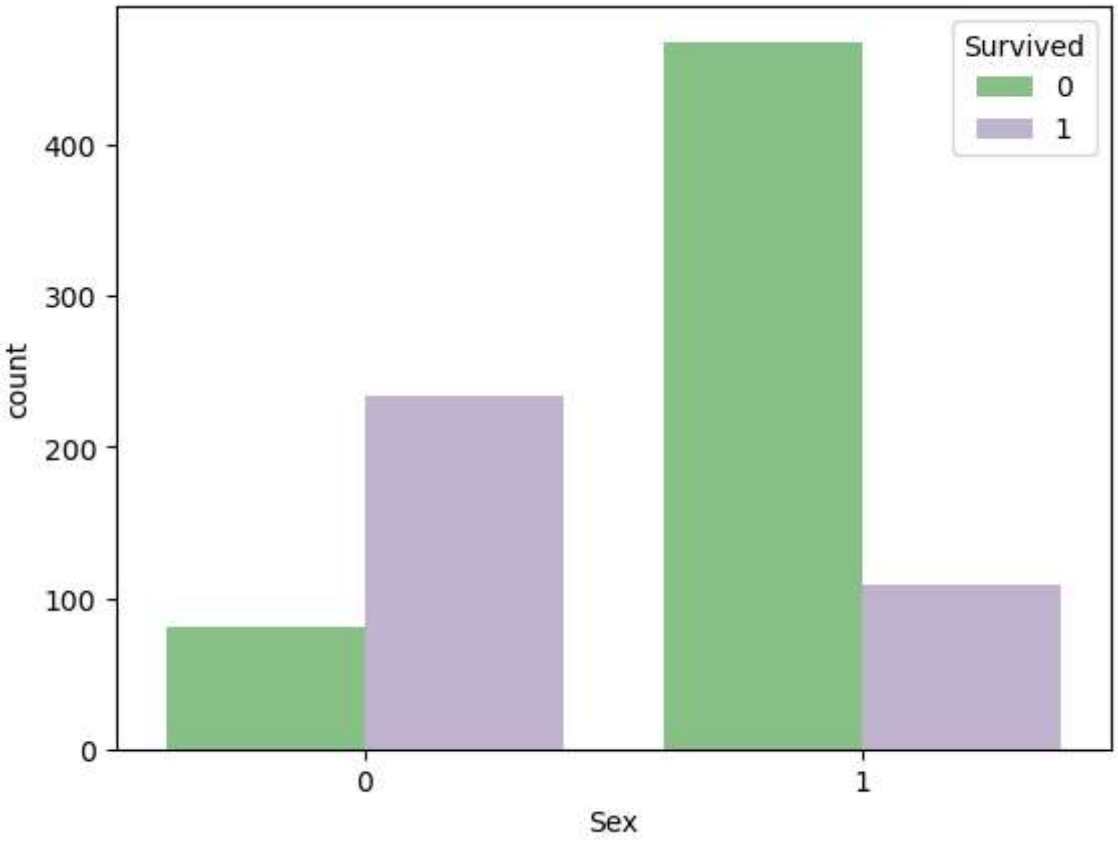
In [92]:
```
sns.countplot(data=data, x="Survived", hue='Survived')
plt.show()
```



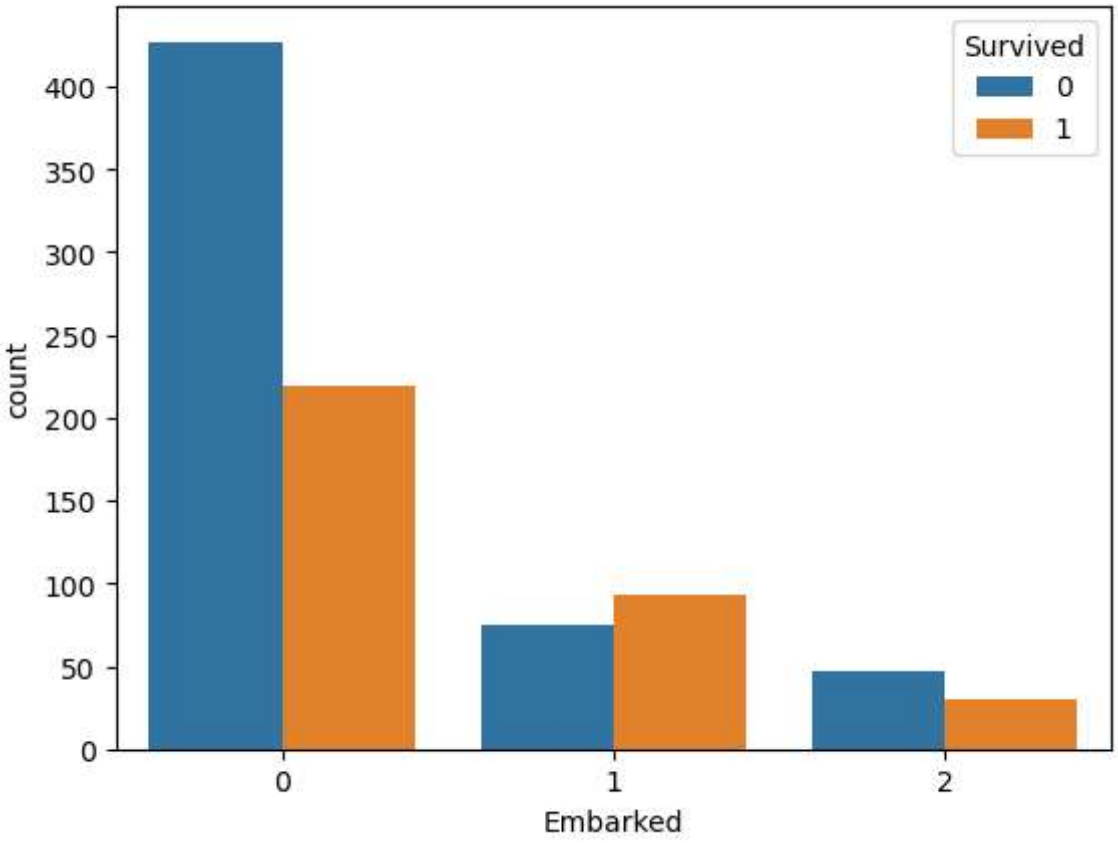
In [96]:
```
# Assuming you have a DataFrame named `data` with columns "Pclass" and "Surviv

# Create a count plot with "Pclass" on the x-axis and "Survived" as the hue
sns.countplot(x=data["Pclass"], hue=data["Survived"])

# Display the plot
plt.show()
```



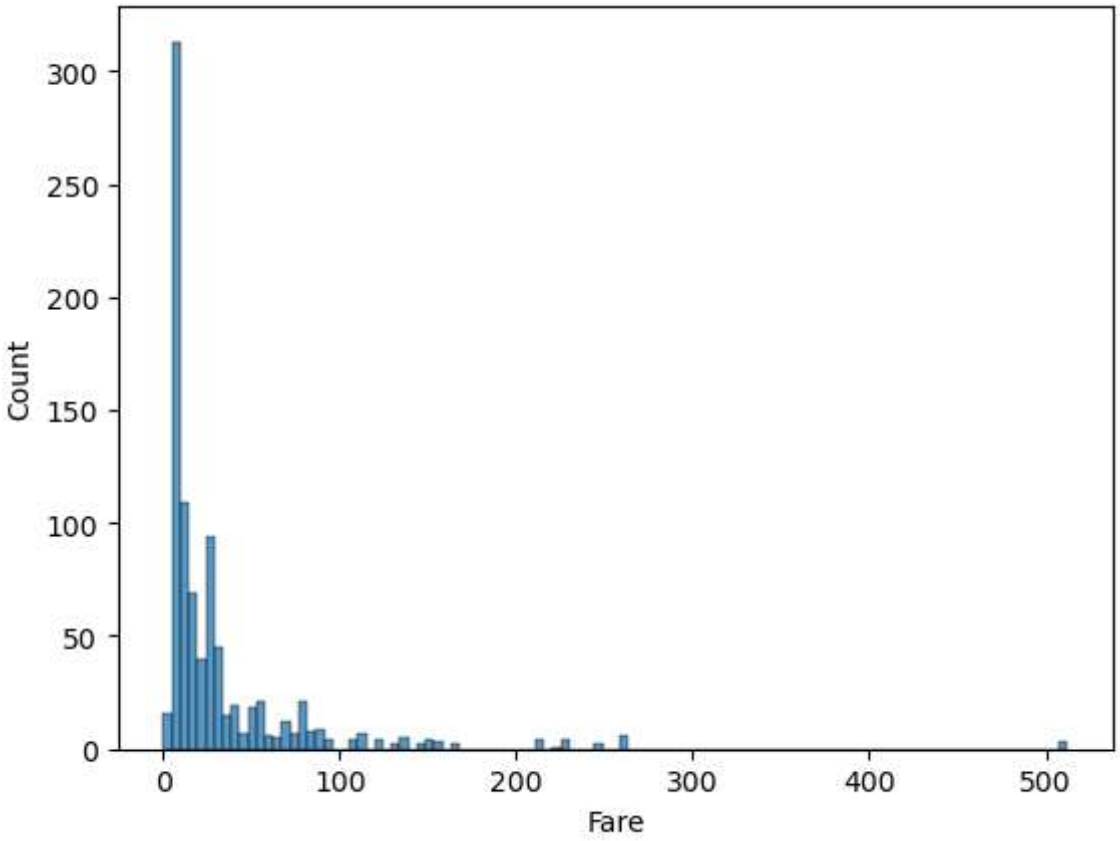
In [98]:
```
sns.countplot(x=data["Sex"], hue=data["Survived"], palette="Accent")
plt.show()
```



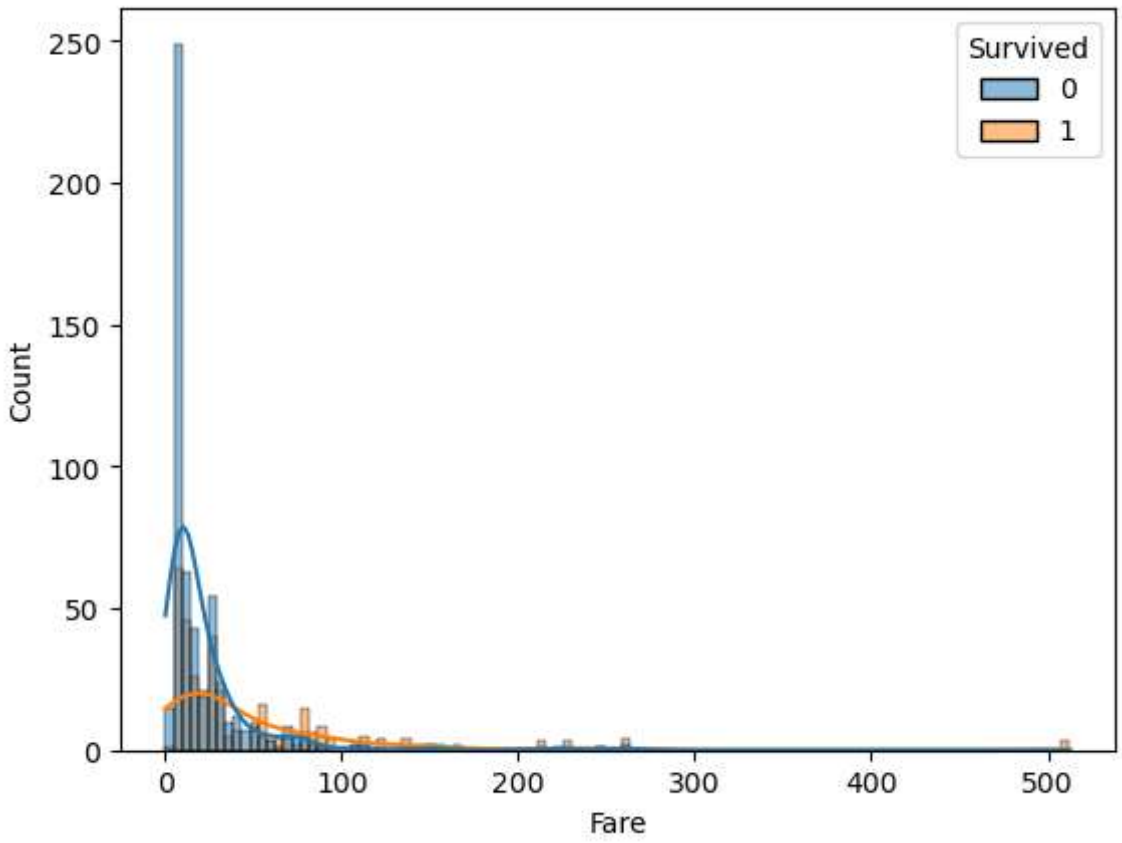
In [99]:
```
sns.countplot(x=data["Embarked"], hue=data["Survived"])
plt.show()
```



In [100]:
```
sns.histplot(data["Fare"])
plt.show()
```



In [101]:
```
sns.histplot(data = data, x = 'Fare', hue = 'Survived',kde = True);
```



In [102]:
```
#The corelation can be confirmed from the above histplot.
#The number of people having cheap tickets was extremely high. However only a
#On the other hand, most of the handful people with expensive tickets survivea
```

In [ ]:
```

```