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

In [2]:
```
data=pd.read_csv("D:\TITANIC.csv")
data
```

Out[2]:

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

891 rows × 12 columns

In [3]:
```
data.info()
data.shape
```

```
<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
```

Out[3]:
```
(891, 12)
```

In [4]:
```
data.head(11)
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	

In [5]:
```
data.tail()
```

Out[5]:

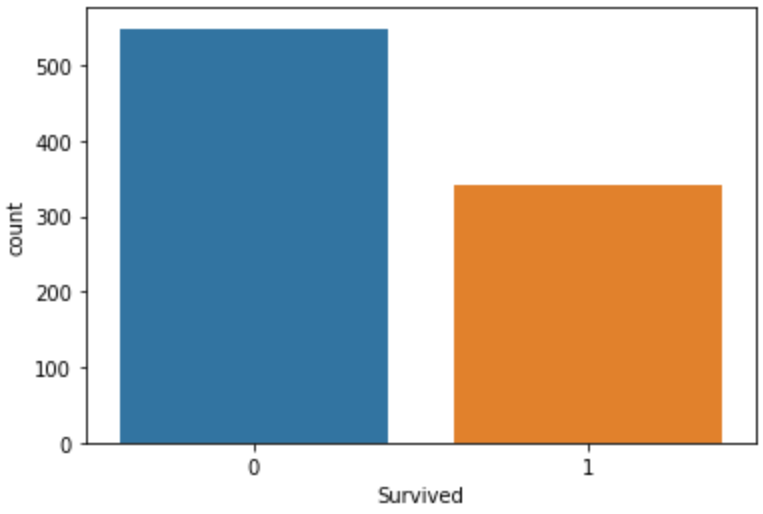
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607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

In [6]:
```
data.describe()
```

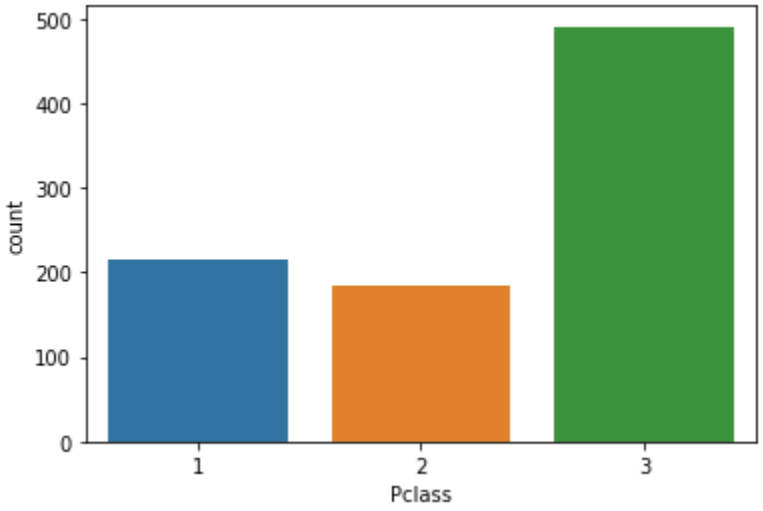
Out[6]:

	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

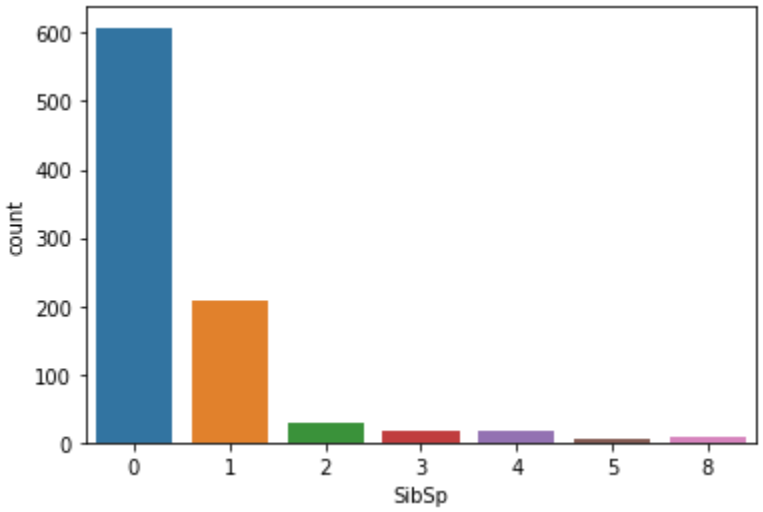
In [7]:
```
sns.countplot(x='Survived',data=data)
plt.show()
```



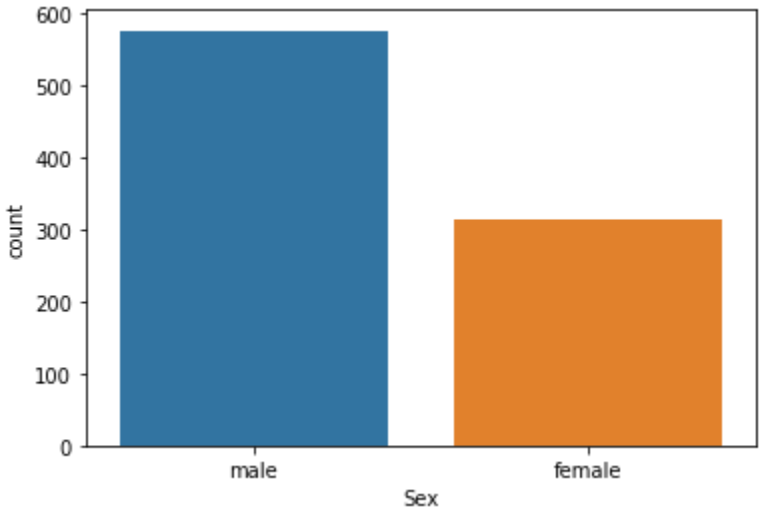
In [8]:
```
sns.countplot(x='Pclass',data=data)
plt.show()
```



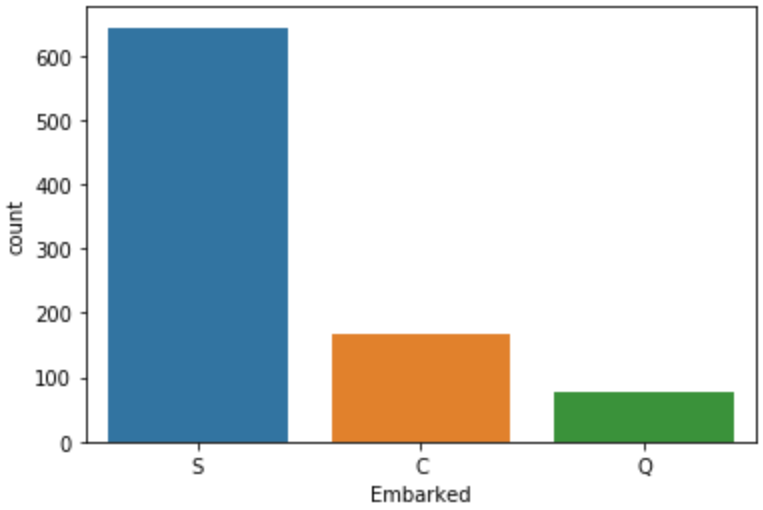
In [9]:
```
sns.countplot(x="SibSp",data=data)
plt.show()
```



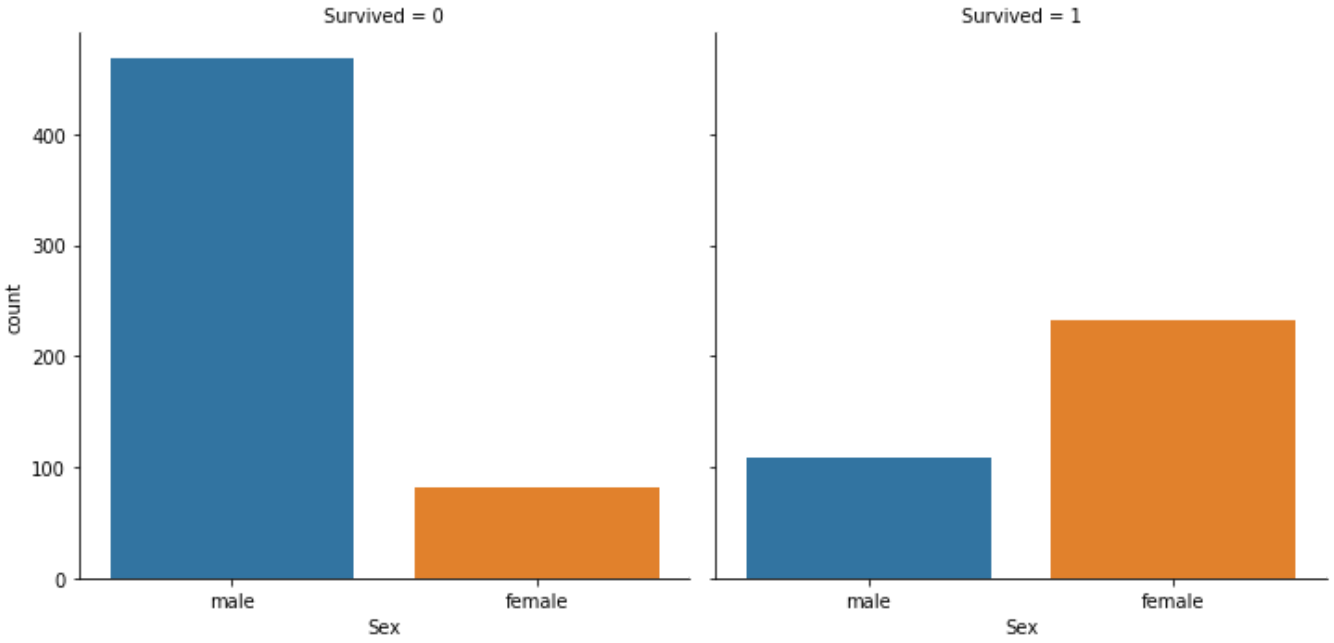
In [10]:
```
sns.countplot(x='Sex',data=data)
plt.show()
```



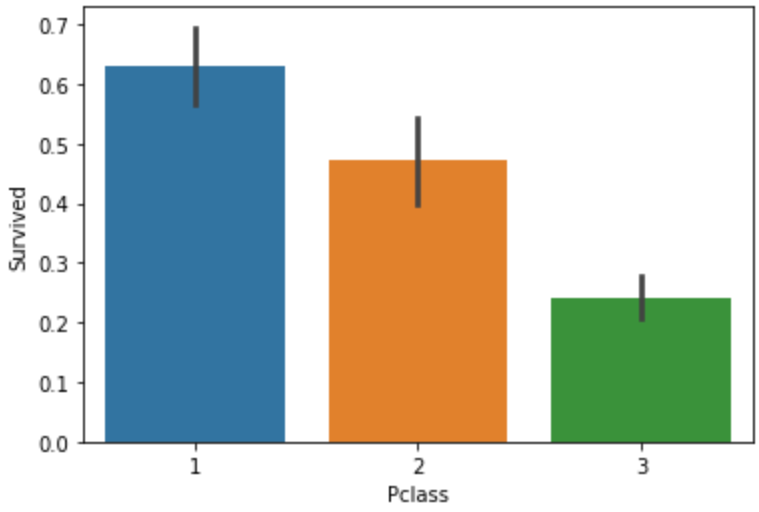
In [11]:
```
sns.countplot(x='Embarked',data=data)
plt.show()
```



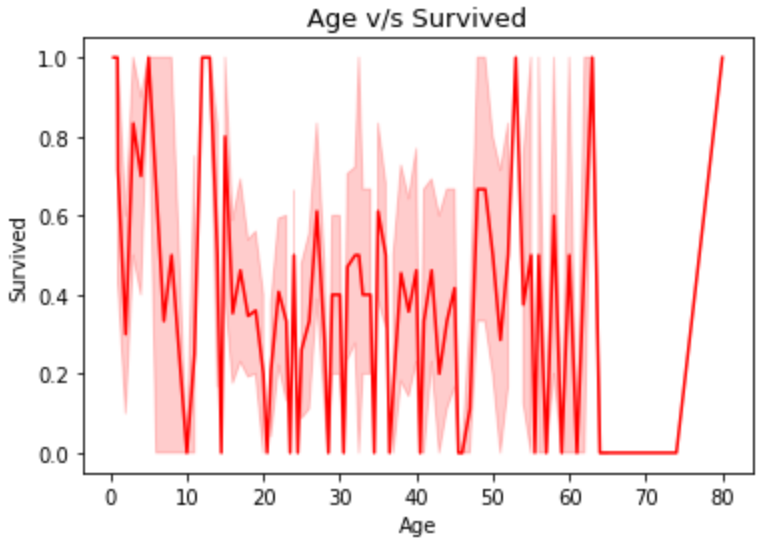
In [12]:
```
sns.catplot(x='Sex',col='Survived',kind='count',data=data)
plt.show()
```



In [13]:
```
sns.barplot(x='Pclass',y='Survived',data=data)
plt.show()
```

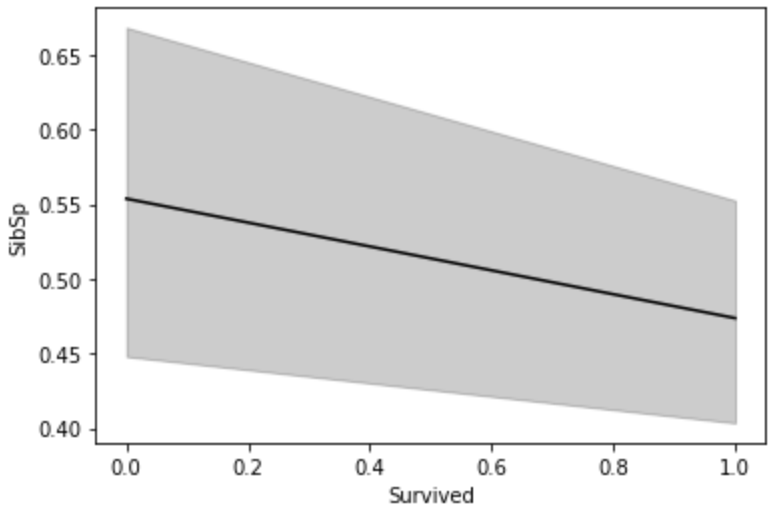


In [14]:
```
sns.lineplot(x='Age',y="Survived",data=data,color='r')
plt.title("Age v/s Survived",fontsize=13)
plt.show()
```

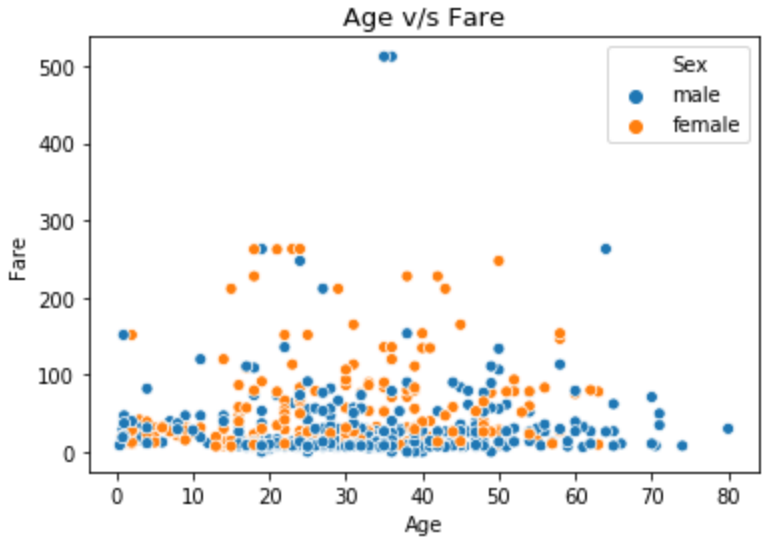


In [15]:
```
sns.lineplot(x='Survived',y='SibSp',data=data,color='k')
```

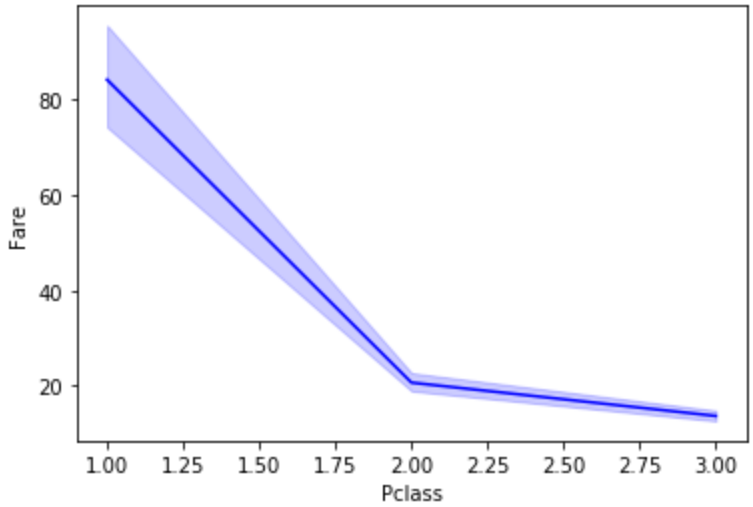
Out[15]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x29562f25788>
```



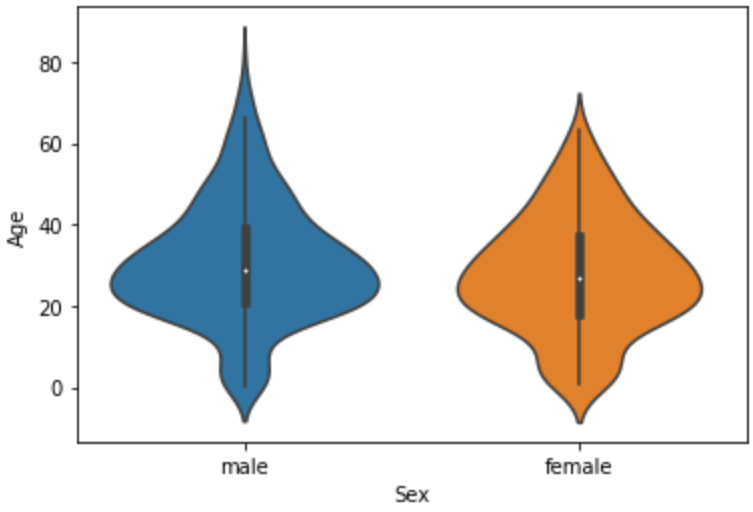
In [16]:
```
sns.scatterplot(x='Age',y='Fare',data=data,hue='Sex')
plt.title('Age v/s Fare',fontsize=13)
plt.show()
```



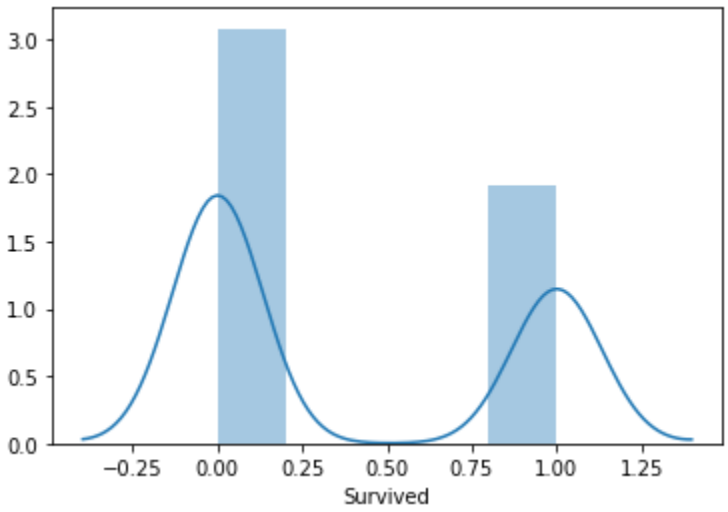
In [17]:
```
sns.lineplot(x='Pclass',y='Fare',data=data,color='b')
plt.show()
```



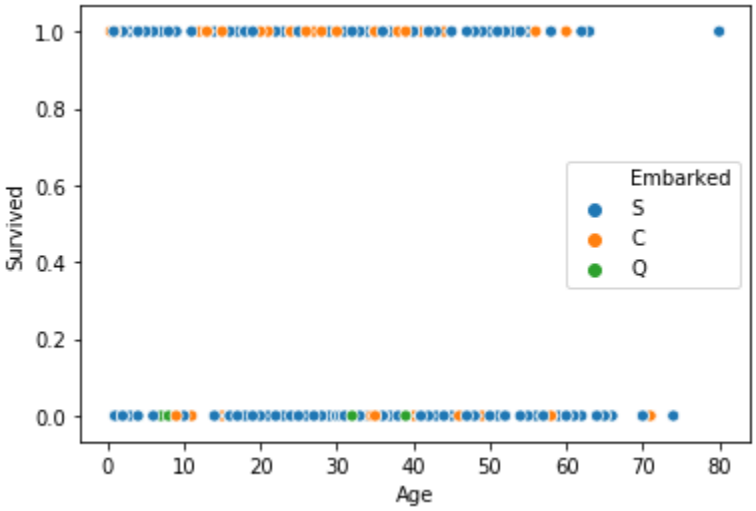
In [18]:
```
sns.violinplot(x='Sex',y='Age',data=data,split=True)
plt.show()
```



In [19]:
```
sns.distplot(data['Survived'])
plt.show()
```



In [20]:
```
sns.scatterplot(x='Age',y='Survived',data=data,hue='Embarked')
plt.show()
```

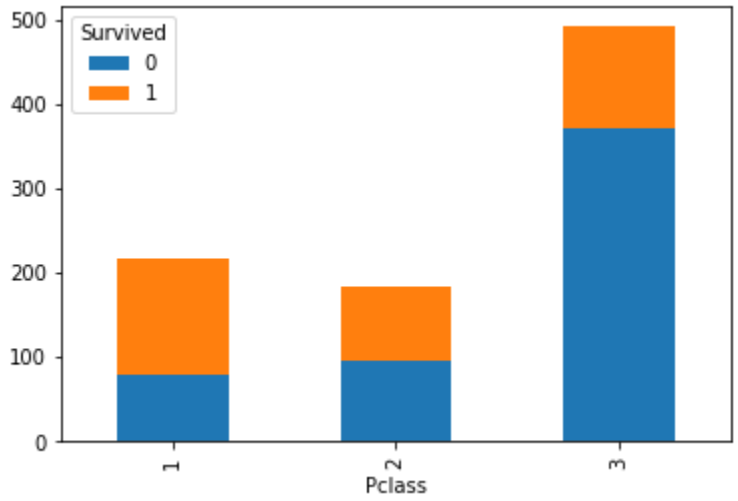


In [21]:
```
pd.crosstab([data.Sex,data.Survived],data.Pclass)
```

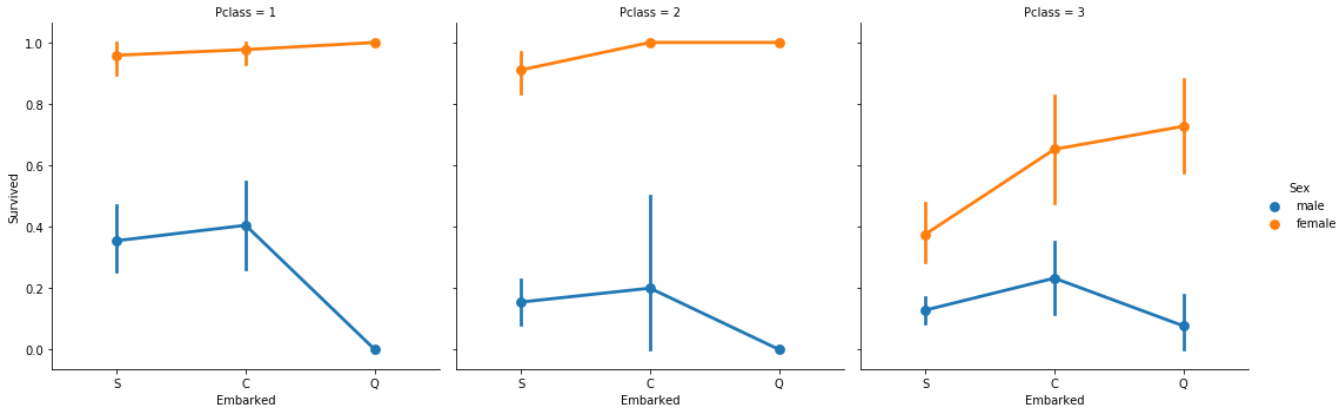
Out[21]:

Pclass		1	2	3
Sex	**Survived**			
female	**0**	3	6	72
	1	91	70	72
male	**0**	77	91	300
	1	45	17	47

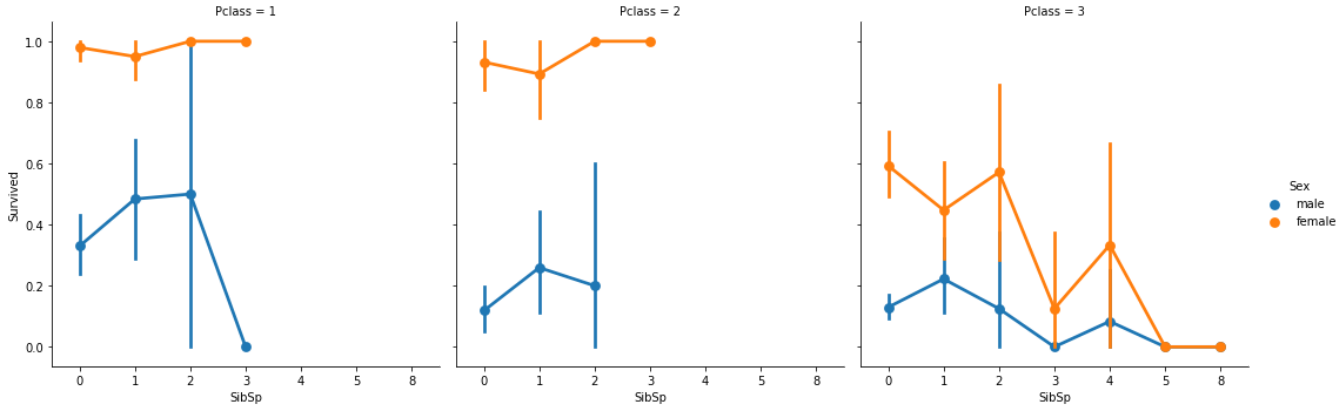
In [22]:
```
pd.crosstab(data["Pclass"],data['Survived']).plot(kind='bar',stacked=True)
plt.show()
```



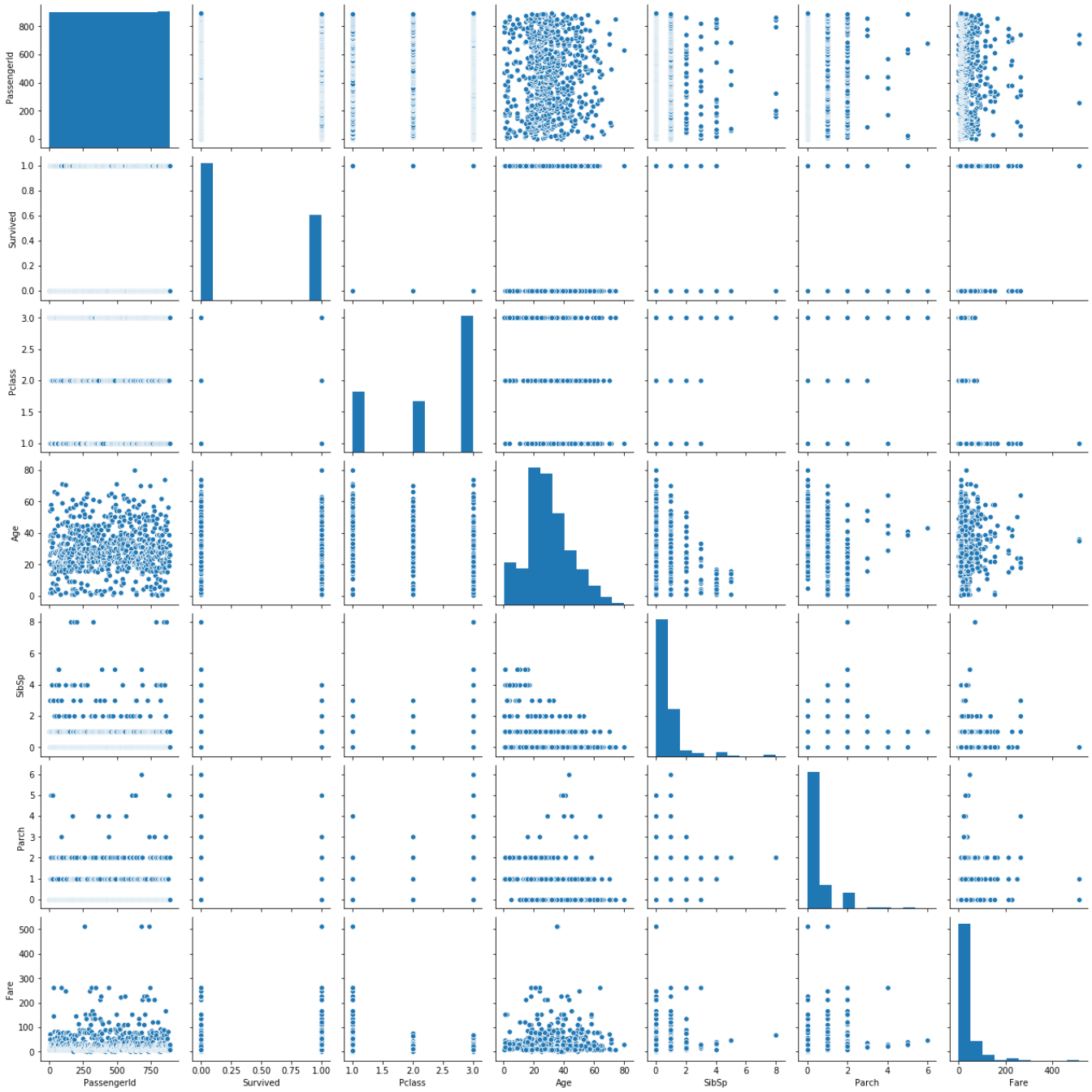
In [23]:
```
sns.catplot('Embarked','Survived',col='Pclass',hue='Sex',data=data,kind='point'
)
plt.show()
```



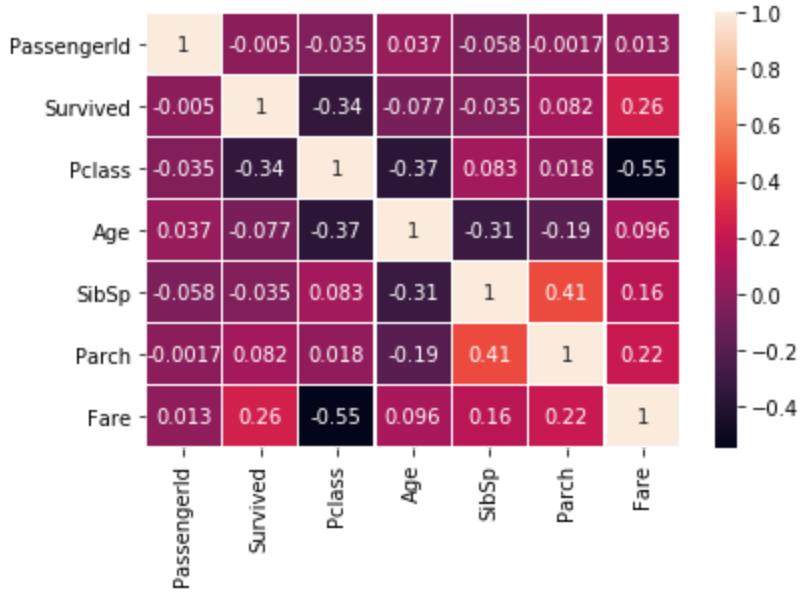
In [24]:
```
sns.catplot('SibSp','Survived',col='Pclass',hue='Sex',data=data,kind='point')
plt.show()
```



In [25]:
```
sns.pairplot(data)
plt.show()
```



In [26]:
```
sns.heatmap(data.corr(),annot=True,linewidth=0.5)
plt.show()
```



In [27]:
```
data.isnull()
```

Out[27]:

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

891 rows × 12 columns

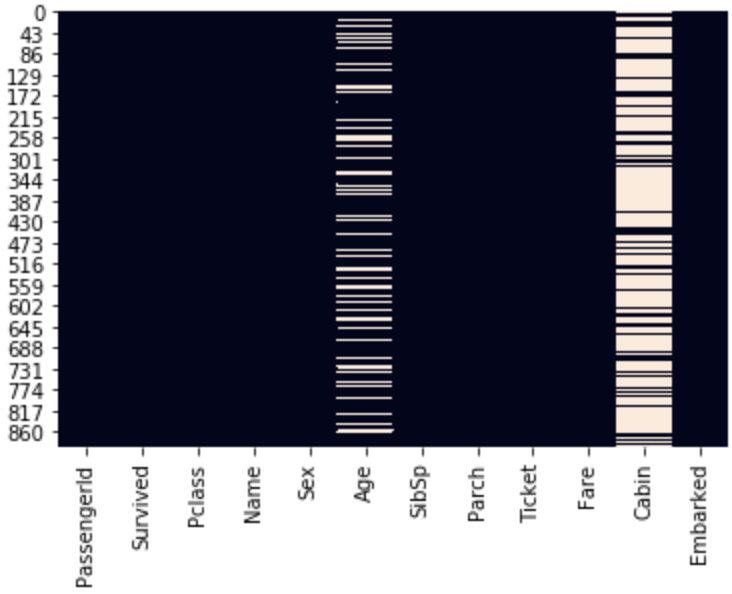
In [28]:
```
data.isnull().sum()
```

Out[28]:
```
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
```

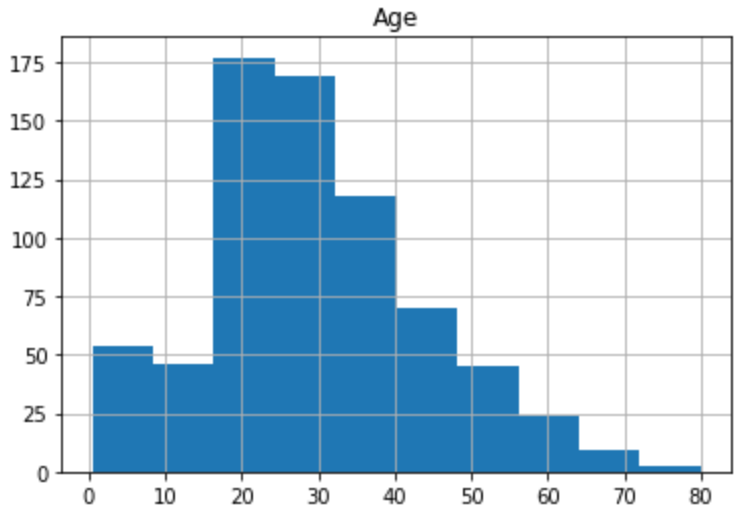
In [29]:
```
data.isnull().sum().sum()
```

Out[29]:
```
866
```

In [30]:
```
sns.heatmap(data.isnull(),cbar=False)
plt.show()
```



In [31]:
```
data.hist(column=['Age'])
plt.show()
```



In [32]:
```
data['Age']=data['Age'].fillna(value=data['Age'].median())
```

In [33]:
```
data['Embarked'].fillna('S')
```

Out[33]:
```
0      S
1      C
2      S
3      S
4      S
      ..
886    S
887    S
888    S
889    C
890    Q
Name: Embarked, Length: 891, dtype: object
```

In [34]:
```
data
```

Out[34]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	

891 rows × 12 columns

In [35]:
```
data
```

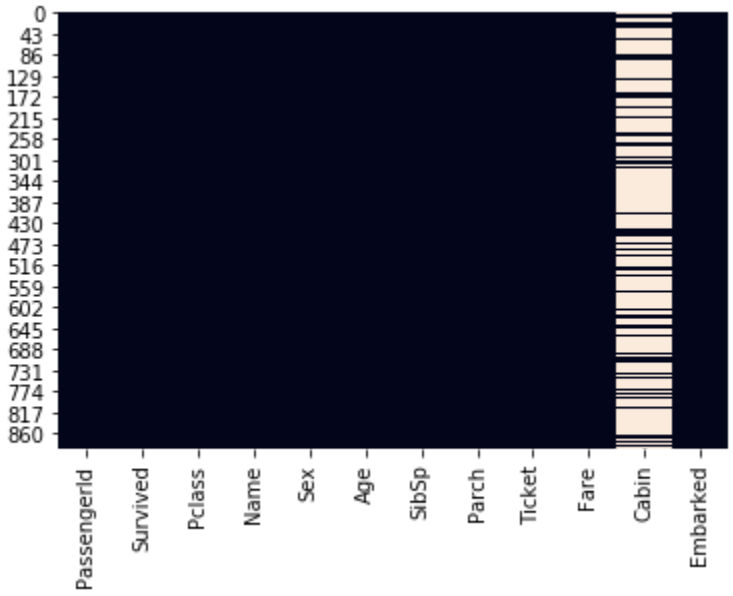
Out[35]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	

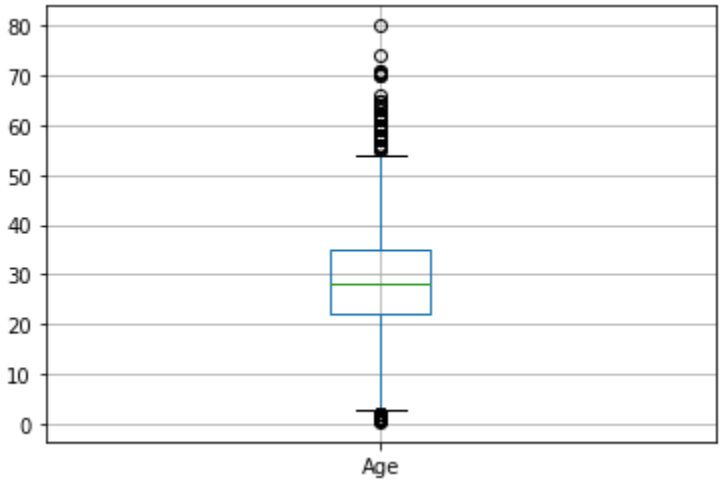
891 rows × 12 columns

In [36]:
```
sns.heatmap(data.isnull(),cbar=False)
```

Out[36]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x29565f73d48>
```



In [37]:
```
data.boxplot(column=['Age'])
plt.show()
```



In [38]:
```
data['Age'].describe()
```

Out[38]:
```
count    891.000000
mean      29.361582
std       13.019697
min        0.420000
25%       22.000000
50%       28.000000
75%       35.000000
max       80.000000
Name: Age, dtype: float64
```

In [39]:
```
q3=data["Age"].quantile(0.75)
q3
```

Out[39]:
```
35.0
```

In [40]:
```
q1=data["Age"].quantile(0.25)
q1
```

Out[40]:
```
22.0
```

In [41]:
```
IQR=data["Age"].quantile(0.75)-data['Age'].quantile(0.25)
print(IQR)
```

```
13.0
```

In [42]:
```
upper_outlierlimit=data['Age'].quantile(0.75)+1.5*IQR
lower_outlierlimit=data['Age'].quantile(0.25)-1.5*IQR
print(upper_outlierlimit)
print(lower_outlierlimit)
```

```
54.5
2.5
```

In [43]:
```
outliervalues=data[(data['Age']>=upper_outlierlimit)|(data['Age']<=lower_outlierlimit)]
outliervalues
```

Out[43]:

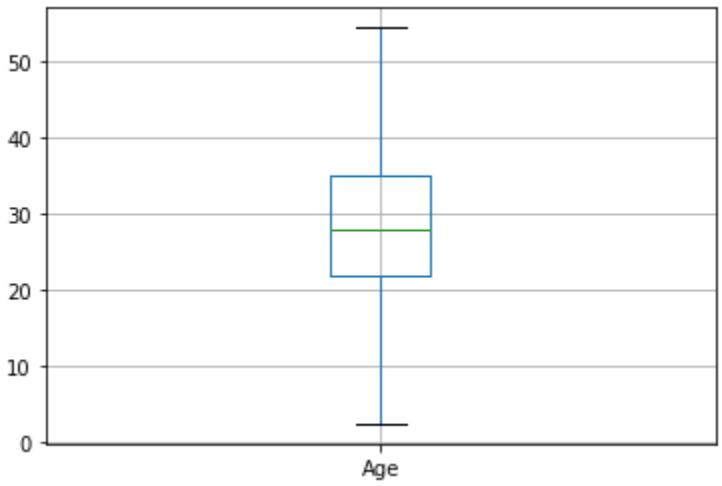
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Er
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.00	3	1	349909	21.0750	NaN	
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.00	0	0	113783	26.5500	C103	
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.00	0	0	248706	16.0000	NaN	
16	17	0	3	Rice, Master. Eugene	male	2.00	4	1	382652	29.1250	NaN	
33	34	0	2	Wheadon, Mr. Edward H	male	66.00	0	0	C.A. 24579	10.5000	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
827	828	1	2	Mallet, Master. Andre	male	1.00	0	2	S.C./PARIS 2079	37.0042	NaN	
829	830	1	1	Stone, Mrs. George Nelson (Martha Evelyn)	female	62.00	0	0	113572	80.0000	B28	
831	832	1	2	Richards, Master. George Sibley	male	0.83	1	1	29106	18.7500	NaN	
851	852	0	3	Svensson, Mr. Johan	male	74.00	0	0	347060	7.7750	NaN	
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.00	0	1	11767	83.1583	C50	

66 rows × 12 columns

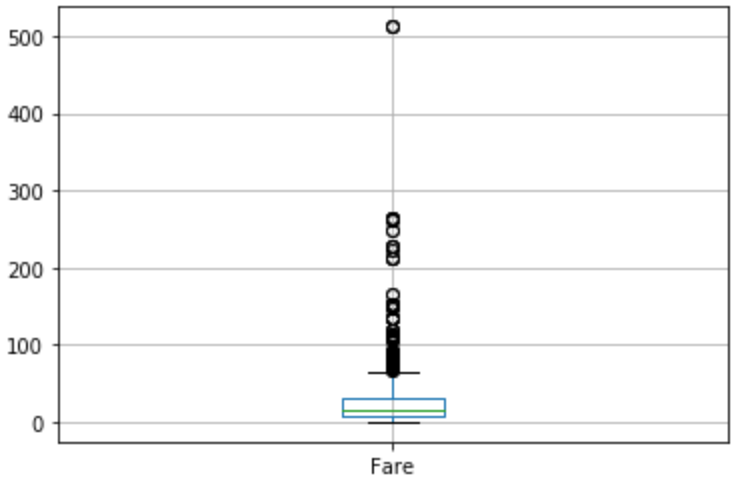
In [44]:
```
data['Age']=np.where(data['Age']>=54.5,54.5,data['Age'])
data['Age']=np.where(data['Age']<=2.5,2.5,data['Age'])
print(data['Age'],data['Age'])
```

```
0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888    28.0
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64 0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888    28.0
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64
```

In [45]:
```
data.boxplot(column=['Age'])              # NEW BOXPLOT FOR 'AGE' AFTER REMOVING OUTLIERS
plt.show()
```



In [46]:
```
data.boxplot(column=['Fare'])
plt.show()
```



In [47]:
```
data['Fare'].describe()
```

Out[47]:
```
count    891.000000
mean      32.204208
std       49.693429
min        0.000000
25%        7.910400
50%       14.454200
75%       31.000000
max      512.329200
Name: Fare, dtype: float64
```

In [48]:
```
q3=data["Fare"].quantile(0.75)
q3
```

Out[48]:
```
31.0
```

In [49]:
```
q1=data["Fare"].quantile(0.25)
q1
```

Out[49]:
```
7.9104
```

In [50]:
```
IQR=data["Fare"].quantile(0.75)-data['Fare'].quantile(0.25)
print(IQR)
```

```
23.0896
```

In [51]:
```
upper_outlierlimit=data['Fare'].quantile(0.75)+1.5*IQR
lower_outlierlimit=data['Fare'].quantile(0.25)-1.5*IQR
print(upper_outlierlimit)
print(lower_outlierlimit)
```

```
65.6344
-26.724
```

In [52]:
```
outliervalues=data[(data['Fare']>=upper_outlierlimit)|(data['Fare']<=lower_outl
ierlimit)]
outliervalues
```

Out[52]:

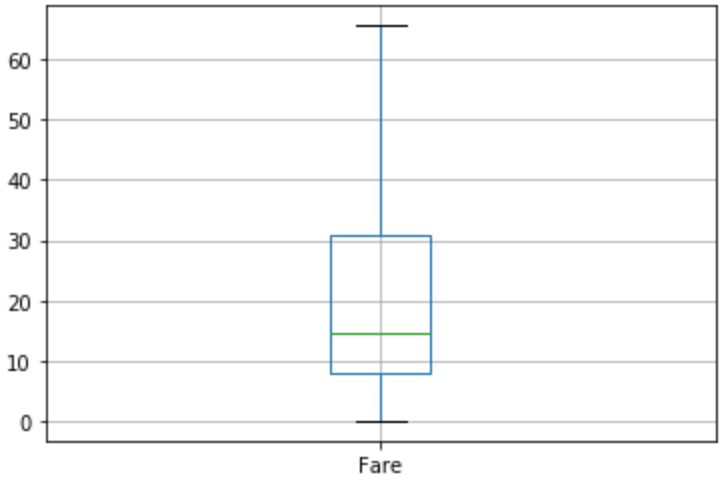
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.0000	C23 C25 C27	
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	28.0	1	0	PC 17569	146.5208	B78	
34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.0	1	0	PC 17604	82.1708	NaN	
52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.0	1	0	PC 17572	76.7292	D33	
...	...	...	...	...	...	...	...	...	...	...	...	
846	847	0	3	Sage, Mr. Douglas Bullen	male	28.0	8	2	CA. 2343	69.5500	NaN	
849	850	1	1	Goldenberg, Mrs. Samuel L (Edwiga Grabowska)	female	28.0	1	0	17453	89.1042	C92	
856	857	1	1	Wick, Mrs. George Dennick (Mary Hitchcock)	female	45.0	1	1	36928	164.8667	NaN	
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	28.0	8	2	CA. 2343	69.5500	NaN	
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	54.5	0	1	11767	83.1583	C50	

116 rows × 12 columns

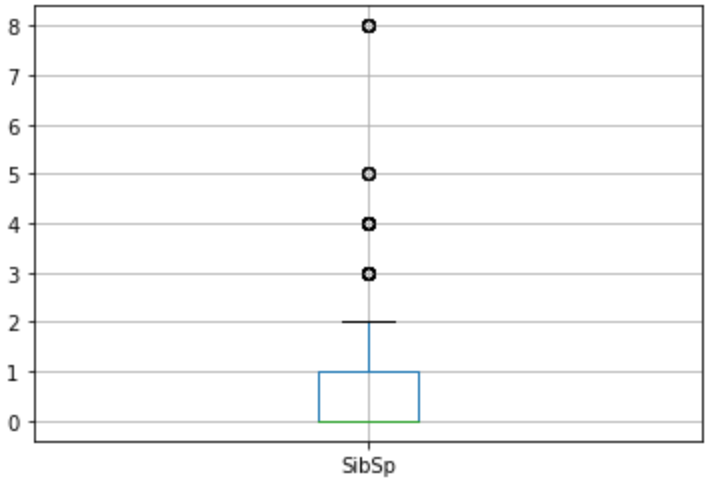
In [53]:
```
data['Fare']=np.where(data['Fare']>=65.6344,65.6344,data['Fare'])
print(data['Fare'])
```

```
0       7.2500
1      65.6344
2       7.9250
3      53.1000
4       8.0500
        ...
886    13.0000
887    30.0000
888    23.4500
889    30.0000
890     7.7500
Name: Fare, Length: 891, dtype: float64
```

In [54]:
```
data.boxplot(column=['Fare'])     # NEW BOXPLOT FOR 'FARE' AFTER REMOVING OUTLIERS
plt.show()
```



In [55]:
```
data.boxplot(column=['SibSp'])
plt.show()
```



In [56]:
```
data['SibSp'].describe()
```

Out[56]:
```
count    891.000000
mean       0.523008
std        1.102743
min        0.000000
25%        0.000000
50%        0.000000
75%        1.000000
max        8.000000
Name: SibSp, dtype: float64
```

In [57]:
```
q3=data["SibSp"].quantile(0.75)
q3
```

Out[57]:
```
1.0
```

In [58]:
```
q1=data["SibSp"].quantile(0.25)
q1
```

Out[58]:
```
0.0
```

In [59]:
```
IQR=data["SibSp"].quantile(0.75)-data['SibSp'].quantile(0.25)
print(IQR)
```

```
1.0
```

In [60]:
```
upper_outlierlimit=data['SibSp'].quantile(0.75)+1.5*IQR
lower_outlierlimit=data['SibSp'].quantile(0.25)-1.5*IQR
print(upper_outlierlimit)
print(lower_outlierlimit)
```

```
2.5
-1.5
```

In [61]:
```
outliervalues=data[(data['SibSp']>=upper_outlierlimit)|(data['SibSp']<=lower_ou
tlierlimit)]
outliervalues
```

Out[61]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.5	3	1	349909	21.0750	NaN	
16	17	0	3	Rice, Master. Eugene	male	2.5	4	1	382652	29.1250	NaN	
24	25	0	3	Palsson, Miss. Torborg Danira	female	8.0	3	1	349909	21.0750	NaN	
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	65.6344	C23 C25 C27	
50	51	0	3	Panula, Master. Juha Niilo	male	7.0	4	1	3101295	39.6875	NaN	
59	60	0	3	Goodwin, Master. William Frederick	male	11.0	5	2	CA 2144	46.9000	NaN	
63	64	0	3	Skoog, Master. Harald	male	4.0	3	2	347088	27.9000	NaN	
68	69	1	3	Andersson, Miss. Erna Alexandra	female	17.0	4	2	3101281	7.9250	NaN	
71	72	0	3	Goodwin, Miss. Lillian Amy	female	16.0	5	2	CA 2144	46.9000	NaN	
85	86	1	3	Backstrom, Mrs. Karl Alfred (Maria Mathilda Gu...	female	33.0	3	0	3101278	15.8500	NaN	
88	89	1	1	Fortune, Miss. Mabel Helen	female	23.0	3	2	19950	65.6344	C23 C25 C27	
119	120	0	3	Andersson, Miss. Ellis Anna Maria	female	2.5	4	2	347082	31.2750	NaN	
159	160	0	3	Sage, Master. Thomas Henry	male	28.0	8	2	CA. 2343	65.6344	NaN	
164	165	0	3	Panula, Master. Eino Viljami	male	2.5	4	1	3101295	39.6875	NaN	
171	172	0	3	Rice, Master. Arthur	male	4.0	4	1	382652	29.1250	NaN	

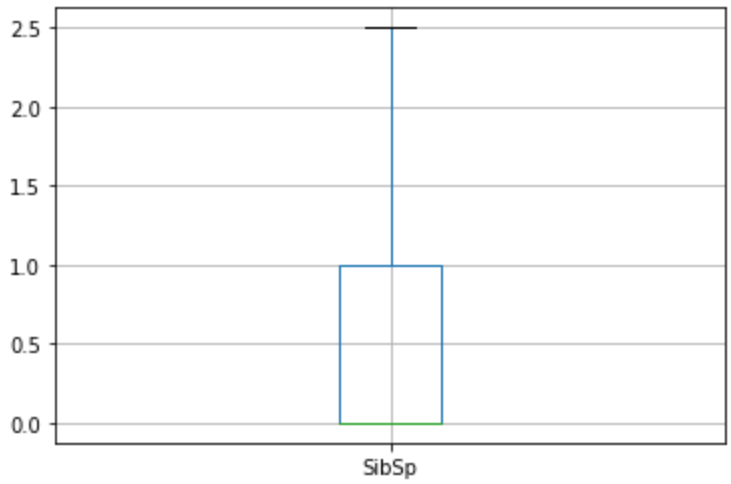
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
176	177	0	3	Lefebre, Master. Henry Forbes	male	28.0	3	1	4133	25.4667	NaN	
180	181	0	3	Sage, Miss. Constance Gladys	female	28.0	8	2	CA. 2343	65.6344	NaN	
182	183	0	3	Asplund, Master. Clarence Gustaf Hugo	male	9.0	4	2	347077	31.3875	NaN	
201	202	0	3	Sage, Mr. Frederick	male	28.0	8	2	CA. 2343	65.6344	NaN	
229	230	0	3	Lefebre, Miss. Mathilde	female	28.0	3	1	4133	25.4667	NaN	
233	234	1	3	Asplund, Miss. Lillian Gertrud	female	5.0	4	2	347077	31.3875	NaN	
261	262	1	3	Asplund, Master. Edvin Rojj Felix	male	3.0	4	2	347077	31.3875	NaN	
266	267	0	3	Panula, Mr. Ernesti Arvid	male	16.0	4	1	3101295	39.6875	NaN	
278	279	0	3	Rice, Master. Eric	male	7.0	4	1	382652	29.1250	NaN	
324	325	0	3	Sage, Mr. George John Jr	male	28.0	8	2	CA. 2343	65.6344	NaN	
341	342	1	1	Fortune, Miss. Alice Elizabeth	female	24.0	3	2	19950	65.6344	C23 C25 C27	
374	375	0	3	Palsson, Miss. Stina Viola	female	3.0	3	1	349909	21.0750	NaN	
386	387	0	3	Goodwin, Master. Sidney Leonard	male	2.5	5	2	CA 2144	46.9000	NaN	
409	410	0	3	Lefebre, Miss. Ida	female	28.0	3	1	4133	25.4667	NaN	
480	481	0	3	Goodwin, Master. Harold Victor	male	9.0	5	2	CA 2144	46.9000	NaN	
485	486	0	3	Lefebre, Miss. Jeannie	female	28.0	3	1	4133	25.4667	NaN	
541	542	0	3	Andersson, Miss. Ingeborg Constanzia	female	9.0	4	2	347082	31.2750	NaN	

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
542	543	0	3	Andersson, Miss. Sigrid Elisabeth	female	11.0	4	2	347082	31.2750	NaN	
634	635	0	3	Skoog, Miss. Mabel	female	9.0	3	2	347088	27.9000	NaN	
642	643	0	3	Skoog, Miss. Margit Elizabeth	female	2.5	3	2	347088	27.9000	NaN	
683	684	0	3	Goodwin, Mr. Charles Edward	male	14.0	5	2	CA 2144	46.9000	NaN	
686	687	0	3	Panula, Mr. Jaako Arnold	male	14.0	4	1	3101295	39.6875	NaN	
726	727	1	2	Renouf, Mrs. Peter Henry (Lillian Jefferys)	female	30.0	3	0	31027	21.0000	NaN	
787	788	0	3	Rice, Master. George Hugh	male	8.0	4	1	382652	29.1250	NaN	
792	793	0	3	Sage, Miss. Stella Anna	female	28.0	8	2	CA. 2343	65.6344	NaN	
813	814	0	3	Andersson, Miss. Ebba Iris Alfrida	female	6.0	4	2	347082	31.2750	NaN	
819	820	0	3	Skoog, Master. Karl Thorsten	male	10.0	3	2	347088	27.9000	NaN	
824	825	0	3	Panula, Master. Urho Abraham	male	2.5	4	1	3101295	39.6875	NaN	
846	847	0	3	Sage, Mr. Douglas Bullen	male	28.0	8	2	CA. 2343	65.6344	NaN	
850	851	0	3	Andersson, Master. Sigvard Harald Elias	male	4.0	4	2	347082	31.2750	NaN	
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	28.0	8	2	CA. 2343	65.6344	NaN	

In [62]:
```
data['SibSp']=np.where(data['SibSp']>=2.5,2.5,data['SibSp'])
print(data['SibSp'])
```

```
0      1.0
1      1.0
2      0.0
3      1.0
4      0.0
      ...
886    0.0
887    0.0
888    1.0
889    0.0
890    0.0
Name: SibSp, Length: 891, dtype: float64
```

In [63]:
```
data.boxplot(column=['SibSp'])                    # NEW BOXPLOT FOR 'Sibsp' A
FTER REMOVING OUTLIERS
plt.show()
```



In [64]:
```
data
```

Out[64]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1.0	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1.0	0	PC 17599	65.6344	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0.0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1.0	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0.0	0	373450	8.0500	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0.0	0	211536	13.0000	NaN	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0.0	0	112053	30.0000	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1.0	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0.0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0.0	0	370376	7.7500	NaN	

891 rows × 12 columns

In [65]:
```
obj= data.dtypes == np.object
obj
```

Out[65]:
```
PassengerId    False
Survived       False
Pclass         False
Name            True
Sex             True
Age            False
SibSp          False
Parch          False
Ticket          True
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

In [66]:
```
data.columns[obj]
```

Out[66]:
```
Index(['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked'], dtype='object')
```

In [67]:
```
data
```

Out[67]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emb
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1.0	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1.0	0	PC 17599	65.6344	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0.0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1.0	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0.0	0	373450	8.0500	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0.0	0	211536	13.0000	NaN	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0.0	0	112053	30.0000	B42	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1.0	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0.0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0.0	0	370376	7.7500	NaN	

891 rows × 12 columns

In [68]:
```
data.drop('Cabin',axis=1,inplace=True)
```

In [69]:
```
data
```

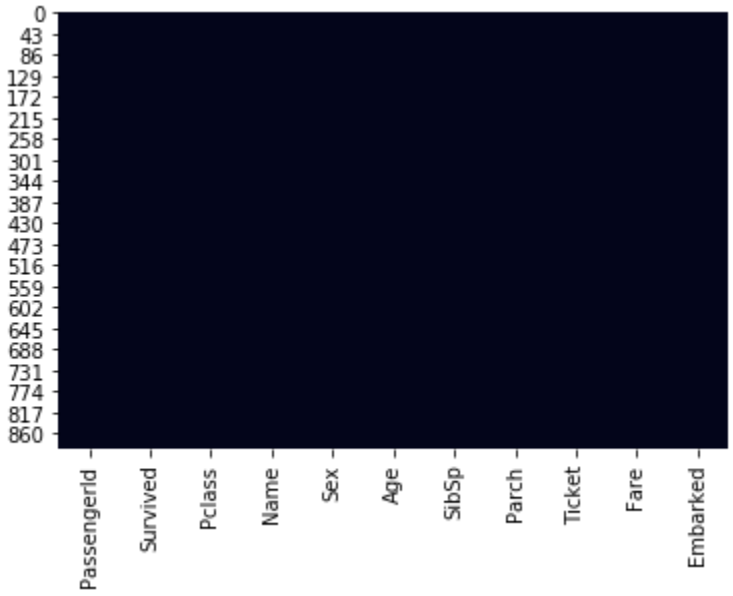
Out[69]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1.0	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1.0	0	PC 17599	65.6344	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0.0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1.0	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0.0	0	373450	8.0500	S
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0.0	0	211536	13.0000	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0.0	0	112053	30.0000	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1.0	2	W./C. 6607	23.4500	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0.0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0.0	0	370376	7.7500	Q

891 rows × 11 columns

In [70]:
```
sns.heatmap(data.isnull(),cbar=False)       #NO NULL VALUES IN THE DATA SET
```

Out[70]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x29560ed4d08>
```



In [71]:
```
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    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    float64
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Embarked     889 non-null    object 
dtypes: float64(3), int64(4), object(4)
memory usage: 76.7+ KB
```

In [72]:
```
sex=pd.get_dummies(data['Sex'],drop_first=True)
embark=pd.get_dummies(data['Embarked'],drop_first=True)
```

In [73]:
```
sex.head()
```

Out[73]:

	male
0	1
1	0
2	0
3	0
4	1

In [74]:
```
embark.head()
```

Out[74]:

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

In [75]:
```
data.drop(['Sex','Name','Ticket','Embarked'],axis=1,inplace=True)
```

In [76]:
```
data.head()
```

Out[76]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
0	1	0	3	22.0	1.0	0	7.2500
1	2	1	1	38.0	1.0	0	65.6344
2	3	1	3	26.0	0.0	0	7.9250
3	4	1	1	35.0	1.0	0	53.1000
4	5	0	3	35.0	0.0	0	8.0500

In [77]:
```
data=pd.concat([data,sex,embark],axis=1)
```

In [78]:
```
data
```

Out[78]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1.0	0	7.2500	1	0	1
1	2	1	1	38.0	1.0	0	65.6344	0	0	0
2	3	1	3	26.0	0.0	0	7.9250	0	0	1
3	4	1	1	35.0	1.0	0	53.1000	0	0	1
4	5	0	3	35.0	0.0	0	8.0500	1	0	1
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	27.0	0.0	0	13.0000	1	0	1
887	888	1	1	19.0	0.0	0	30.0000	0	0	1
888	889	0	3	28.0	1.0	2	23.4500	0	0	1
889	890	1	1	26.0	0.0	0	30.0000	1	0	0
890	891	0	3	32.0	0.0	0	7.7500	1	1	0

891 rows × 10 columns

In [79]:
```
data.head()
```

Out[79]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1.0	0	7.2500	1	0	1
1	2	1	1	38.0	1.0	0	65.6344	0	0	0
2	3	1	3	26.0	0.0	0	7.9250	0	0	1
3	4	1	1	35.0	1.0	0	53.1000	0	0	1
4	5	0	3	35.0	0.0	0	8.0500	1	0	1