```
In [154]: import pandas as pd
          import numpy as np
          import seaborn as sns
          import matplotlib.pyplot as plt
          %matplotlib inline
```

```
In [155]: TR = pd.read_csv('train.csv')
```

```
In [156]: TR.shape

Out[156]: (891, 12)
```

```
In [157]: print('No or rows in data:', TR.shape[0])

          No or rows in data: 891
```

```
In [158]: print('No or columns in data:', TR.shape[1])

          No or columns in data: 12
```

In [159]:
```
TR.head(10)
```

Out[159]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
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	
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	
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	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	

In [160]:
```
pd.options.display.max_rows
```

Out[160]:
```
90
```

In [161]:
```
pd.options.display.max_columns
```

Out[161]:
```
40
```

In [162]:
```
pd.set_option('display.max_rows',90)
```

In [163]:
```
pd.set_option('display.max_columns',40)
```

In [164]:
```
pd.options.display.max_rows
```

Out[164]:
```
90
```

In [165]:
```
pd.options.display.max_columns
```

Out[165]:
```
40
```

In [166]:
```
TR.tail(12)
```

Out[166]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000
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 [167]:

```
TR.iloc[9:15, 10:12]
```

Out[167]:

	Cabin	Embarked
9	NaN	C
10	G6	S
11	C103	S
12	NaN	S
13	NaN	S
14	NaN	S

In [168]:

```
TR.iloc[[0,4,9],[2,4,7]]
```

Out[168]:

	Pclass	Sex	Parch
0	3	male	0
4	3	male	0
9	2	female	0

In [169]:

```
TR[['PassengerId', 'Name']]
```

Out[169]:

	PassengerId	Name
0	1	Braund, Mr. Owen Harris
1	2	Cumings, Mrs. John Bradley (Florence Briggs Th...
2	3	Heikkinen, Miss. Laina
3	4	Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	5	Allen, Mr. William Henry
...	...	...
886	887	Montvila, Rev. Juozas
887	888	Graham, Miss. Margaret Edith
888	889	Johnston, Miss. Catherine Helen "Carrie"
889	890	Behr, Mr. Karl Howell
890	891	Dooley, Mr. Patrick

891 rows × 2 columns

In [170]:

```
TR.dtypes
```

Out[170]:

```
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 [171]:

```
TR.count()
```

Out[171]:

```
PassengerId    891
Survived       891
Pclass         891
Name           891
Sex            891
Age            714
SibSp          891
Parch          891
Ticket         891
Fare           891
Cabin          204
Embarked       889
dtype: int64
```

In [172]:

```
TR.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 [173]:

```
TR.nunique()
```

Out[173]:

```
PassengerId    891
Survived         2
Pclass           3
Name           891
Sex              2
Age             88
SibSp            7
Parch            7
Ticket         681
Fare           248
Cabin          147
Embarked         3
dtype: int64
```

In [174]:

```
TR.describe(include='object')
```

Out[174]:

	Name	Sex	Ticket	Cabin	Embarked
count	891	891	891	204	889
unique	891	2	681	147	3
top	Torber, Mr. Ernst William	male	347082	G6	S
freq	1	577	7	4	644

In [175]:

```
TR.duplicated()
```

Out[175]:

```
0      False
1      False
2      False
3      False
4      False
       ...  
886    False
887    False
888    False
889    False
890    False
Length: 891, dtype: bool
```

In [176]:
```
TR.isnull()
```

Out[176]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [177]:
```
TR.isnull().sum()
```

Out[177]:
```
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 [178]:
```
TR['Age'].mean()
```

Out[178]:
```
29.69911764705882
```

In [179]:
```
TR['Survived'].mean()
```

Out[179]:
```
0.3838383838383838
```

In [180]:
```
TR['Cabin'].mode()
```

Out[180]:
```
0        B96 B98
1    C23 C25 C27
2             G6
dtype: object
```

In [181]:
```
TR['Age'].mode()
```

Out[181]:
```
0    24.0
dtype: float64
```

In [182]:
```
TR['Age'].median()
```

Out[182]:
```
28.0
```

In [183]:
```
TR['Embarked'].mode()
```

Out[183]:
```
0    S
dtype: object
```

In [184]:
```
TR['Age'].fillna(24, inplace=True)
```

In [185]:
```
TR['Cabin'].fillna(TR['Cabin'].mode(), inplace=True)
```

In [186]:
```
TR['Embarked'].fillna('S', inplace=True)
```

In [187]:
```
TR['Cabin'].fillna('G6', inplace=True)
```

In [188]:
```
TR.isnull().sum()
```

Out[188]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

In [189]:
```
TR.describe()
```

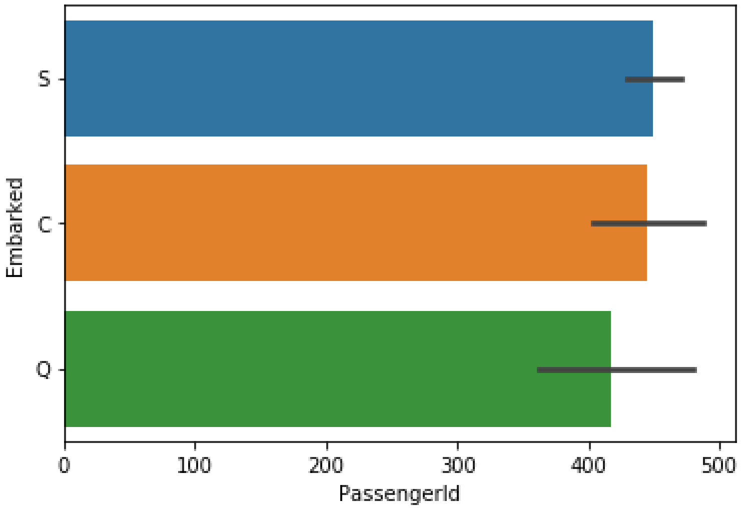
Out[189]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	28.566970	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	13.199572	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	22.000000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	24.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

Number of passenger embarked from each port.

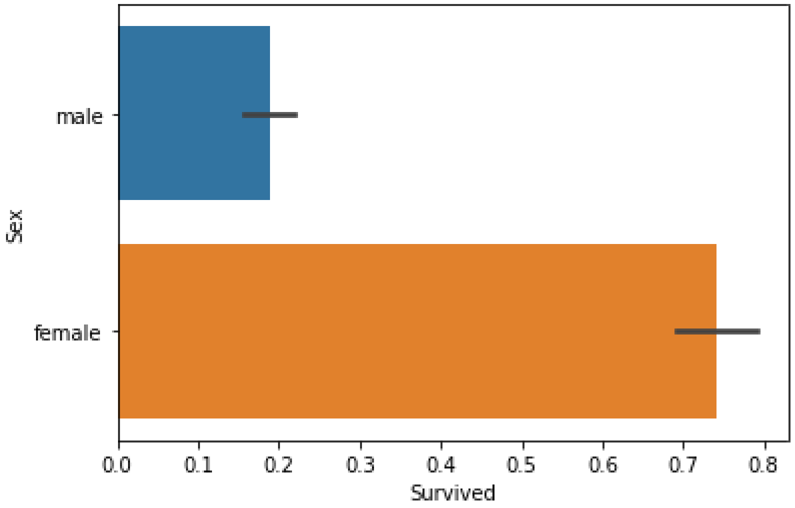
In [190]:
```
sns.barplot(x="PassengerId", y="Embarked", data=TR)
```

Out[190]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2052bfc4408>
```



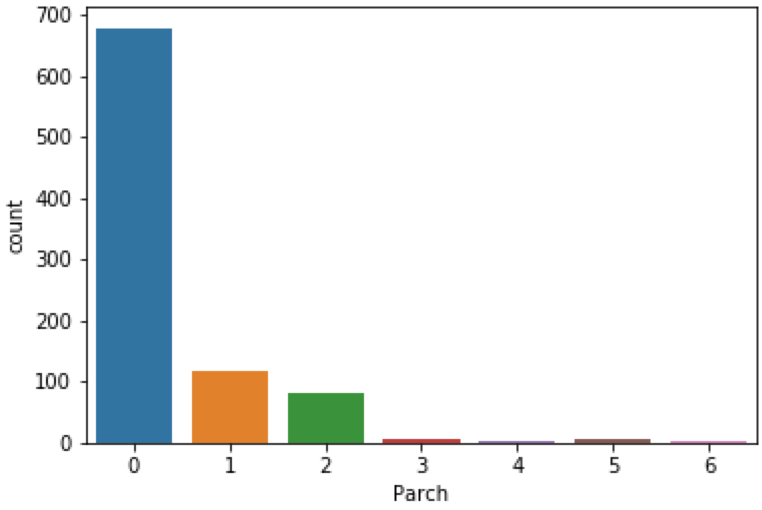
Below graphs shows gender preferred while surviving

```
In [191]: sns.barplot(x="Survived", y="Sex", data=TR)

Out[191]: <matplotlib.axes._subplots.AxesSubplot at 0x2052c0160c8>
```



Below graph shows passenger with or without children

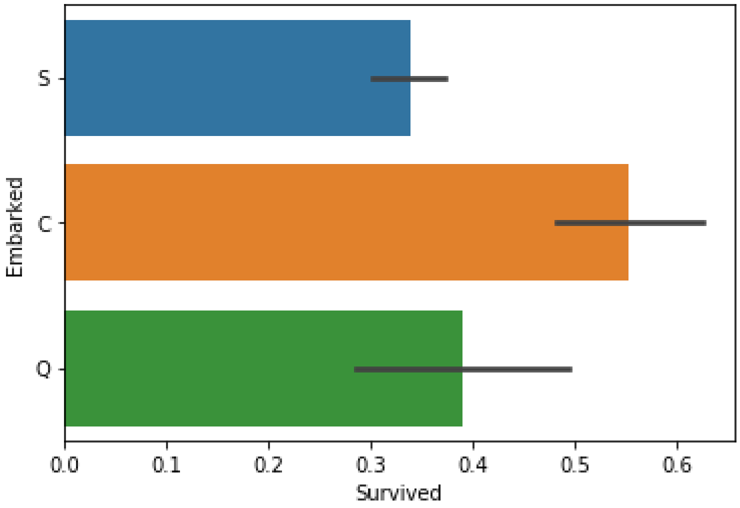
```
In [192]: sns.countplot('Parch', data=TR)

Out[192]: <matplotlib.axes._subplots.AxesSubplot at 0x2052bf06a88>
```



Below graphs shows number of people survived from each port

In [193]:
```
sns.barplot(x="Survived", y="Embarked", data=TR)
```

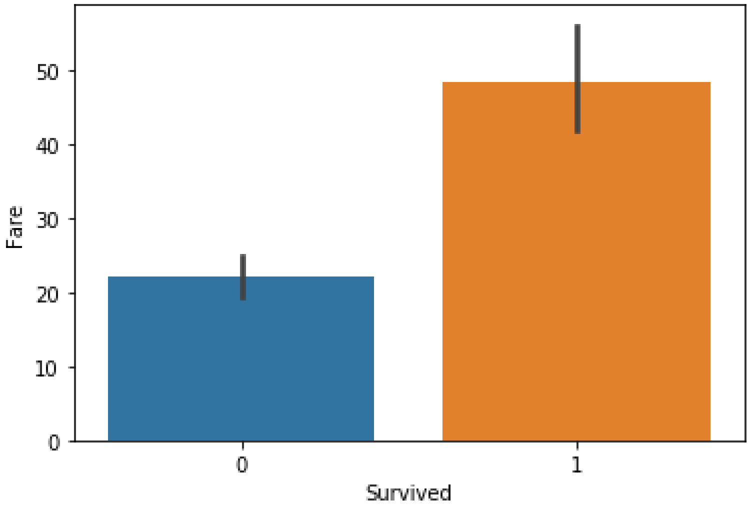
Out[193]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2052c0f0f48>
```



Below graphs depict average fare of passengers who survived or not. (0=Not Survived, 1=Survived)

In [194]:
```
sns.barplot(x="Survived", y="Fare", data=TR)
```

Out[194]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2052c145148>
```



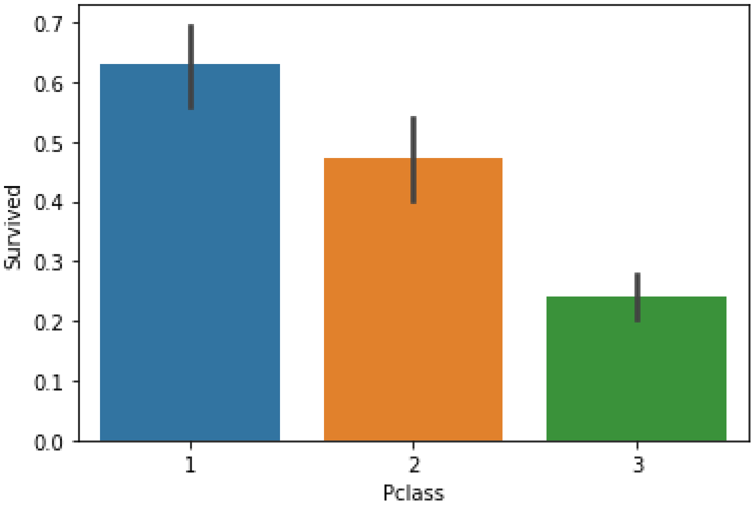
While rescuing highest priority was given to people from Pclass 1

In [195]:

```
sns.barplot(x="Pclass", y="Survived", data=TR)
```

Out[195]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2052c19a048>
```



In [201]:

```
print("Percentage of Pclass = 1 who survived:", TR["Survived"][TR["Pclass"] == 1].value_counts(normalize = True)[1]*100)

print("Percentage of Pclass = 2 who survived:", TR["Survived"][TR["Pclass"] == 2].value_counts(normalize = True)[1]*100)

print("Percentage of Pclass = 3 who survived:", TR["Survived"][TR["Pclass"] == 3].value_counts(normalize = True)[1]*100)
```

```
Percentage of Pclass = 1 who survived: 62.96296296296296
Percentage of Pclass = 2 who survived: 47.28260869565217
Percentage of Pclass = 3 who survived: 24.236252545824847
```

In [ ]: