

```
In [1]: import pandas as pd
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv("/home/placement/Downloads/Titanic Dataset.csv")
```

```
In [2]: data.describe()
```

Out[2]:

	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 [3]: 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 [4]: data.isna().sum()

```
Out[4]: 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 [5]: data.head(15)
```

```
Out[5]:
```

	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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
12	13	0	3	Saunderscock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S

```
In [6]: data['Pclass'].unique()
```

```
Out[6]: array([3, 1, 2])
```

```
In [7]: data['Survived'].unique()
```

```
Out[7]: array([0, 1])
```

```
In [8]: data['SibSp'].unique()
```

```
Out[8]: array([1, 0, 3, 4, 2, 5, 8])
```

```
In [9]: data['Parch'].unique()
```

```
Out[9]: array([0, 1, 2, 5, 3, 4, 6])
```

```
In [10]: data['Age'].unique()
```

```
Out[10]: array([22. , 38. , 26. , 35. , nan, 54. , 2. , 27. , 14. ,  
         4. , 58. , 20. , 39. , 55. , 31. , 34. , 15. , 28. ,  
         8. , 19. , 40. , 66. , 42. , 21. , 18. , 3. , 7. ,  
        49. , 29. , 65. , 28.5 , 5. , 11. , 45. , 17. , 32. ,  
        16. , 25. , 0.83, 30. , 33. , 23. , 24. , 46. , 59. ,  
        71. , 37. , 47. , 14.5 , 70.5 , 32.5 , 12. , 9. , 36.5 ,  
        51. , 55.5 , 40.5 , 44. , 1. , 61. , 56. , 50. , 36. ,  
        45.5 , 20.5 , 62. , 41. , 52. , 63. , 23.5 , 0.92, 43. ,  
        60. , 10. , 64. , 13. , 48. , 0.75, 53. , 57. , 80. ,  
        70. , 24.5 , 6. , 0.67, 30.5 , 0.42, 34.5 , 74. ])
```

```
In [11]: data1=data.drop(['PassengerId', 'Name', 'Cabin', 'SibSp', 'Parch', 'Ticket'],axis=1)
data1
```

Out[11]:

	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	3	male	22.0	7.2500	S
1	1	1	female	38.0	71.2833	C
2	1	3	female	26.0	7.9250	S
3	1	1	female	35.0	53.1000	S
4	0	3	male	35.0	8.0500	S
...
886	0	2	male	27.0	13.0000	S
887	1	1	female	19.0	30.0000	S
888	0	3	female	NaN	23.4500	S
889	1	1	male	26.0	30.0000	C
890	0	3	male	32.0	7.7500	Q

891 rows × 6 columns

```
In [12]: data1['Sex']=data1['Sex'].map({'male':0,'female':1})
data1
```

Out[12]:

	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	3	0	22.0	7.2500	S
1	1	1	1	38.0	71.2833	C
2	1	3	1	26.0	7.9250	S
3	1	1	1	35.0	53.1000	S
4	0	3	0	35.0	8.0500	S
...
886	0	2	0	27.0	13.0000	S
887	1	1	1	19.0	30.0000	S
888	0	3	1	NaN	23.4500	S
889	1	1	0	26.0	30.0000	C
890	0	3	0	32.0	7.7500	Q

891 rows × 6 columns

```
In [13]: data2=data1.fillna(data1.median())  
data2
```

Out[13]:

	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	3	0	22.0	7.2500	S
1	1	1	1	38.0	71.2833	C
2	1	3	1	26.0	7.9250	S
3	1	1	1	35.0	53.1000	S
4	0	3	0	35.0	8.0500	S
...
886	0	2	0	27.0	13.0000	S
887	1	1	1	19.0	30.0000	S
888	0	3	1	28.0	23.4500	S
889	1	1	0	26.0	30.0000	C
890	0	3	0	32.0	7.7500	Q

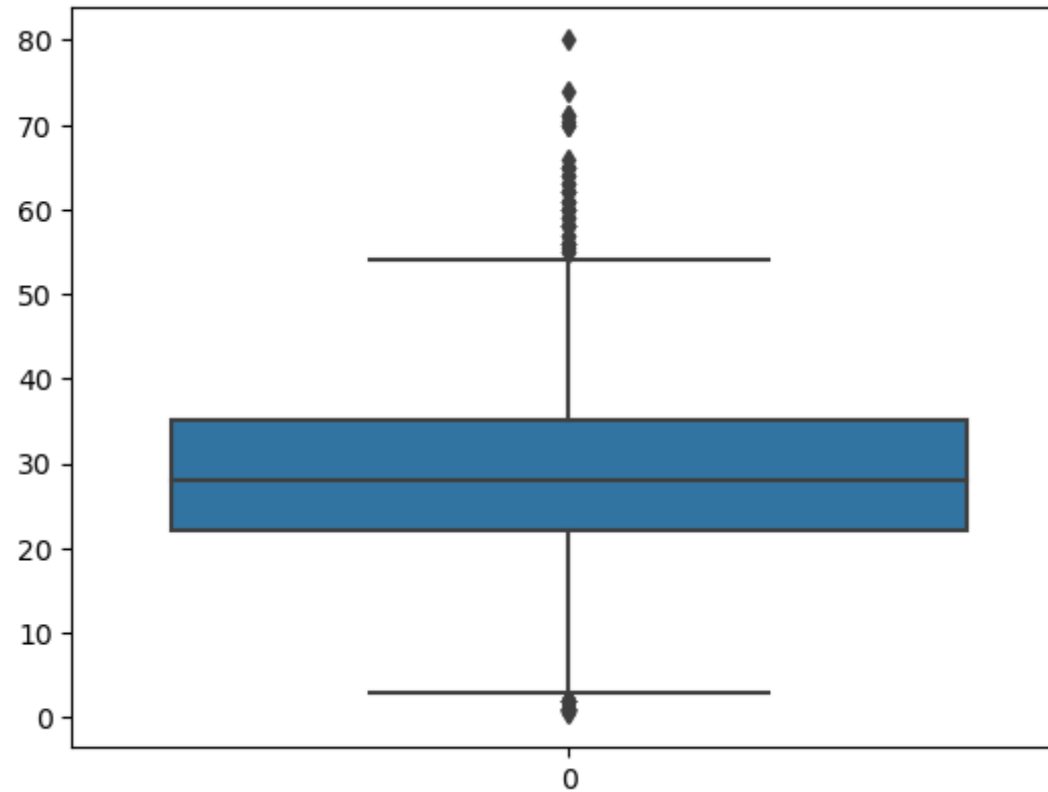
891 rows × 6 columns

```
In [14]: data2.isna().sum()
```

```
Out[14]: Survived    0  
Pclass    0  
Sex        0  
Age        0  
Fare        0  
Embarked    2  
dtype: int64
```

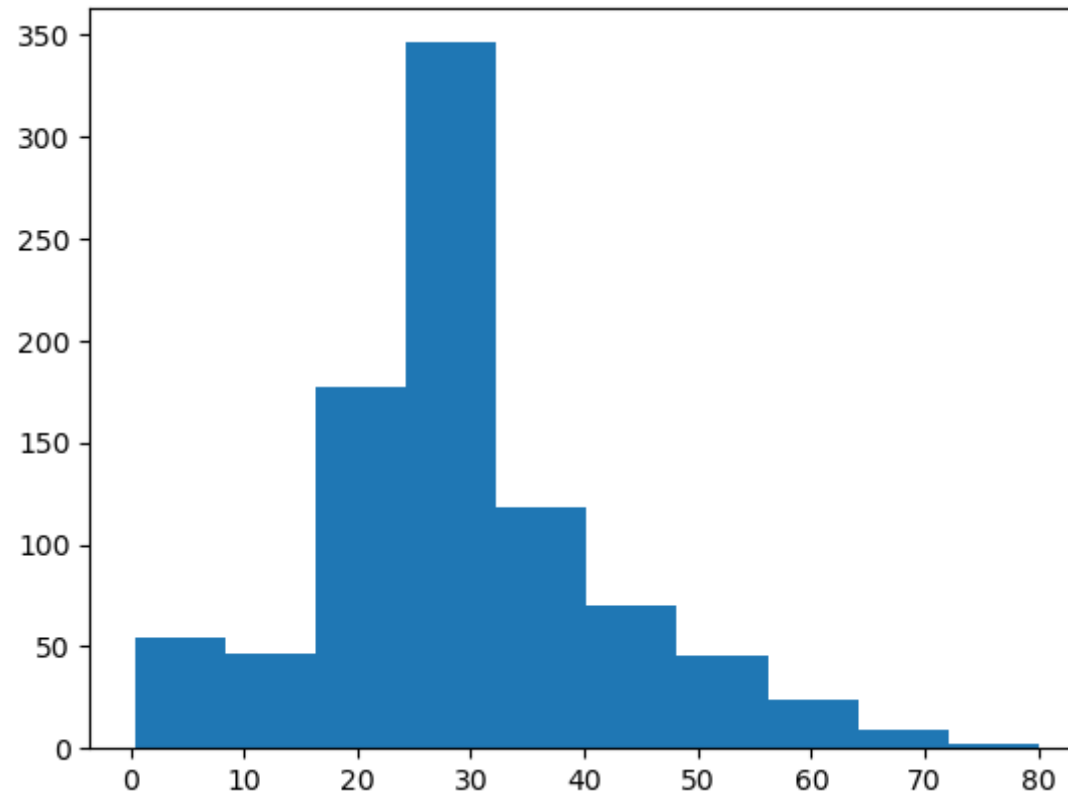
```
In [15]: import seaborn as sns  
import matplotlib.pyplot as plt  
sns.boxplot(data2.Age)
```

Out[15]: <Axes: >



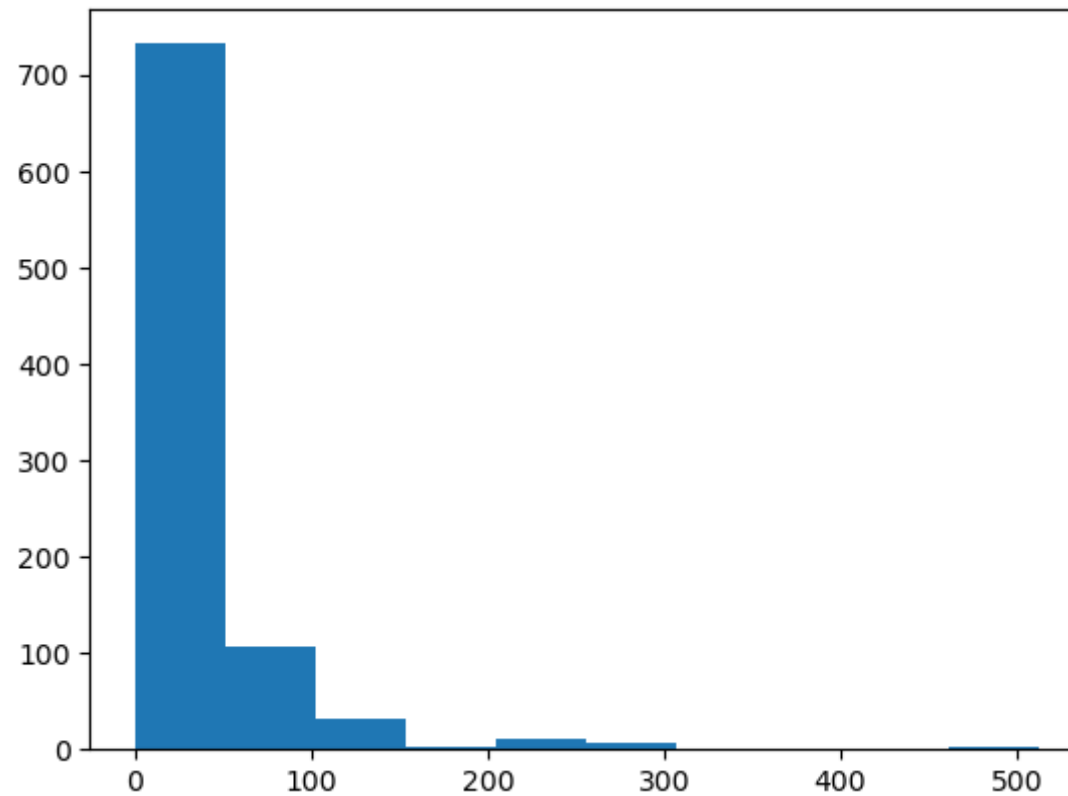

```
In [16]: plt.hist(data2['Age'])
```

```
Out[16]: (array([ 54.,  46., 177., 346., 118.,  70.,  45.,  24.,   9.,   2.]),  
array([ 0.42 ,  8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126,  
        64.084, 72.042, 80.   ]),  
<BarContainer object of 10 artists>)
```



```
In [17]: plt.hist(data2['Fare'])
```

```
Out[17]: (array([732., 106., 31., 2., 11., 6., 0., 0., 0., 3.]),  
array([ 0., 51.23292, 102.46584, 153.69876, 204.93168, 256.1646 ,  
307.39752, 358.63044, 409.86336, 461.09628, 512.3292 ]),  
<BarContainer object of 10 artists>)
```



```
In [18]: data2.isna().sum()
```

```
Out[18]: Survived      0  
Pclass      0  
Sex         0  
Age         0  
Fare        0  
Embarked    2  
dtype: int64
```

```
In [19]: data2.describe()
```

```
Out[19]:
```

	Survived	Pclass	Sex	Age	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	0.352413	29.361582	32.204208
std	0.486592	0.836071	0.477990	13.019697	49.693429
min	0.000000	1.000000	0.000000	0.420000	0.000000
25%	0.000000	2.000000	0.000000	22.000000	7.910400
50%	0.000000	3.000000	0.000000	28.000000	14.454200
75%	1.000000	3.000000	1.000000	35.000000	31.000000
max	1.000000	3.000000	1.000000	80.000000	512.329200

```
In [20]: data2['Age'].unique()
```

```
Out[20]: array([22. , 38. , 26. , 35. , 28. , 54. , 2. , 27. , 14. ,
        4. , 58. , 20. , 39. , 55. , 31. , 34. , 15. , 8. ,
        19. , 40. , 66. , 42. , 21. , 18. , 3. , 7. , 49. ,
        29. , 65. , 28.5 , 5. , 11. , 45. , 17. , 32. , 16. ,
        25. , 0.83, 30. , 33. , 23. , 24. , 46. , 59. , 71. ,
        37. , 47. , 14.5 , 70.5 , 32.5 , 12. , 9. , 36.5 , 51. ,
        55.5 , 40.5 , 44. , 1. , 61. , 56. , 50. , 36. , 45.5 ,
        20.5 , 62. , 41. , 52. , 63. , 23.5 , 0.92, 43. , 60. ,
        10. , 64. , 13. , 48. , 0.75, 53. , 57. , 80. , 70. ,
        24.5 , 6. , 0.67, 30.5 , 0.42, 34.5 , 74. ])
```

```
In [21]: data2.groupby(['Age']).count()
```

```
Out[21]:
```

	Survived	Pclass	Sex	Fare	Embarked
Age					
0.42	1	1	1	1	1
0.67	1	1	1	1	1
0.75	2	2	2	2	2
0.83	2	2	2	2	2
0.92	1	1	1	1	1
...
70.00	2	2	2	2	2
70.50	1	1	1	1	1
71.00	2	2	2	2	2
74.00	1	1	1	1	1
80.00	1	1	1	1	1

88 rows × 5 columns

```
In [22]: data2['Pclass']=data1['Pclass'].map({1:'F',2:'S',3:'T'})
```

In [23]: data2

Out[23]:

	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	T	0	22.0	7.2500	S
1	1	F	1	38.0	71.2833	C
2	1	T	1	26.0	7.9250	S
3	1	F	1	35.0	53.1000	S
4	0	T	0	35.0	8.0500	S
...
886	0	S	0	27.0	13.0000	S
887	1	F	1	19.0	30.0000	S
888	0	T	1	28.0	23.4500	S
889	1	F	0	26.0	30.0000	C
890	0	T	0	32.0	7.7500	Q

891 rows × 6 columns

```
In [24]: data2=pd.get_dummies(data2)
data2
```

Out[24]:

	Survived	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_T	Embarked_C	Embarked_Q	Embarked_S
0	0	0	22.0	7.2500	0	0	1	0	0	1
1	1	1	38.0	71.2833	1	0	0	1	0	0
2	1	1	26.0	7.9250	0	0	1	0	0	1
3	1	1	35.0	53.1000	1	0	0	0	0	1
4	0	0	35.0	8.0500	0	0	1	0	0	1
...
886	0	0	27.0	13.0000	0	1	0	0	0	1
887	1	1	19.0	30.0000	1	0	0	0	0	1
888	0	1	28.0	23.4500	0	0	1	0	0	1
889	1	0	26.0	30.0000	1	0	0	1	0	0
890	0	0	32.0	7.7500	0	0	1	0	1	0

891 rows × 10 columns

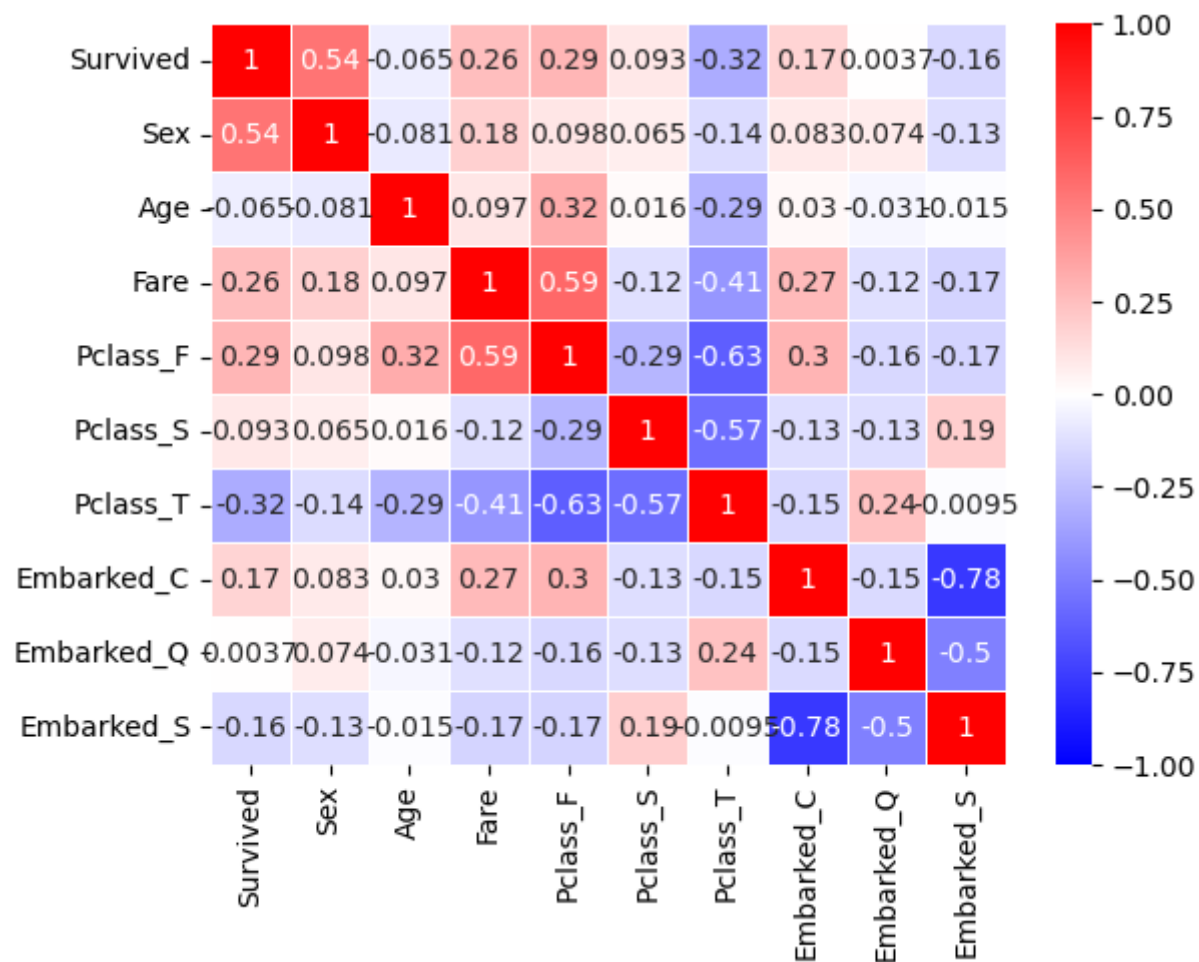
```
In [25]: cor=data2.corr()  
cor
```

Out[25]:

	Survived	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_T	Embarked_C	Embarked_Q	Embarked_S
Survived	1.000000	0.543351	-0.064910	0.257307	0.285904	0.093349	-0.322308	0.168240	0.003650	-0.155660
Sex	0.543351	1.000000	-0.081163	0.182333	0.098013	0.064746	-0.137143	0.082853	0.074115	-0.125722
Age	-0.064910	-0.081163	1.000000	0.096688	0.323896	0.015831	-0.291955	0.030248	-0.031415	-0.014665
Fare	0.257307	0.182333	0.096688	1.000000	0.591711	-0.118557	-0.413333	0.269335	-0.117216	-0.166603
Pclass_F	0.285904	0.098013	0.323896	0.591711	1.000000	-0.288585	-0.626738	0.296423	-0.155342	-0.170379
Pclass_S	0.093349	0.064746	0.015831	-0.118557	-0.288585	1.000000	-0.565210	-0.125416	-0.127301	0.192061
Pclass_T	-0.322308	-0.137143	-0.291955	-0.413333	-0.626738	-0.565210	1.000000	-0.153329	0.237449	-0.009511
Embarked_C	0.168240	0.082853	0.030248	0.269335	0.296423	-0.125416	-0.153329	1.000000	-0.148258	-0.778359
Embarked_Q	0.003650	0.074115	-0.031415	-0.117216	-0.155342	-0.127301	0.237449	-0.148258	1.000000	-0.496624
Embarked_S	-0.155660	-0.125722	-0.014665	-0.166603	-0.170379	0.192061	-0.009511	-0.778359	-0.496624	1.000000

```
In [26]: import seaborn as sns
sns.heatmap(cor, vmax=1, vmin=-1, annot=True, linewidths=.5, cmap='bwr')
```

Out[26]: <Axes: >




```
In [27]: data2.groupby('Survived').count()
```

```
Out[27]:
```

	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_T	Embarked_C	Embarked_Q	Embarked_S
Survived									
0	549	549	549	549	549	549	549	549	549
1	342	342	342	342	342	342	342	342	342

```
In [28]: #adding to seperate dataframe the value,we want to predict
y=data2['Survived']
#removing the value we want to predict from original dataframe
x=data2.drop('Survived',axis=1)
```

```
In [29]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.33,random_state=42)
```

```
In [30]: from sklearn.linear_model import LogisticRegression
classifier=LogisticRegression()
classifier.fit(x_train,y_train)
```

```
Out[30]:
```

```
▼ LogisticRegression
LogisticRegression()
```

```
In [31]: y_pred=classifier.predict(x_test)
y_pred
```

```
Out[31]: array([0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
                1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
                1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
                0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1,
                0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
                1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0,
                0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1,
                0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0,
                0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0,
                1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0,
                0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1,
                0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0,
                0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
                1, 0, 0, 0, 0, 0, 1, 1, 0])
```

```
In [32]: from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

```
Out[32]: array([[154,  21],
                [ 37,  83]])
```

```
In [33]: from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

```
Out[33]: 0.8033898305084746
```

```
In [ ]:
```