

```
In [2]: import pandas as pd
```

```
In [3]: data=pd.read_csv("/home/placement/Downloads/Titanic Dataset.csv")
```

```
In [4]: data.describe()
```

Out[4]:

	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 [5]: 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 [6]: data.isna().sum()

```
Out[6]: 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 [7]: `data.head(10)`

Out[7]:

	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

In [8]: `list(data)`

Out[8]:

```
[ 'PassengerId',
  'Survived',
  'Pclass',
  'Name',
  'Sex',
  'Age',
  'SibSp',
  'Parch',
  'Ticket',
  'Fare',
  'Cabin',
  'Embarked' ]
```

```
In [9]: data.Pclass.unique()
```

```
Out[9]: array([3, 1, 2])
```

```
In [10]: data.Survived.unique()
```

```
Out[10]: array([0, 1])
```

```
In [11]: data.SibSp.unique()
```

```
Out[11]: array([1, 0, 3, 4, 2, 5, 8])
```

```
In [12]: data['Age'].unique()
```

```
Out[12]: 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 [13]: data1=data.drop(['Name', 'PassengerId', 'Ticket', 'Cabin', 'SibSp', 'Parch'],axis=1)
```

In [14]: data1

Out[14]:

	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 [15]: data1['Sex']=data1['Sex'].map({'male':1,'female':0})  
data1['Pclass'].unique()

Out[15]: array([3, 1, 2])

In [16]: data1=data1.fillna(data1.median())

/tmp/ipykernel\_4868/2430846586.py:1: FutureWarning: The default value of numeric\_only in DataFrame.median is deprecated. In a future version, it will default to False. In addition, specifying 'numeric\_only=None' is deprecated. Select only valid columns or specify the value of numeric\_only to silence this warning.  
data1=data1.fillna(data1.median())

```
In [17]: data1
```

```
Out[17]:
```

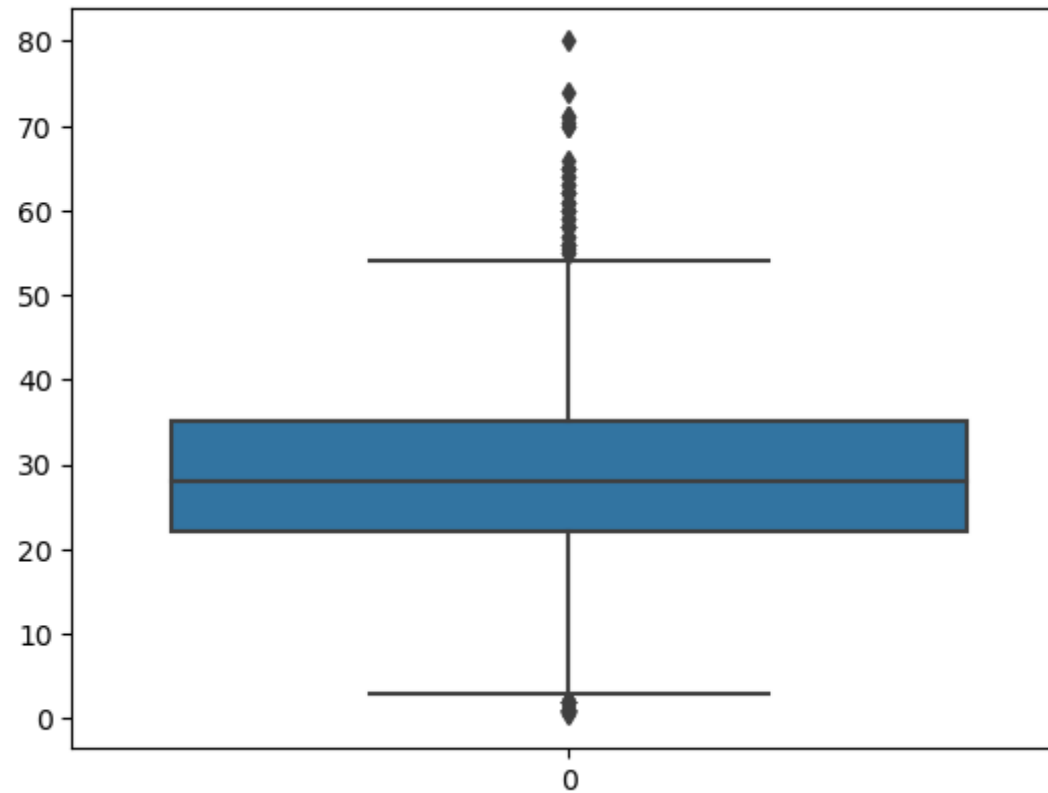
	Survived	Pclass	Sex	Age	Fare	Embarked
0	0	3	1	22.0	7.2500	S
1	1	1	0	38.0	71.2833	C
2	1	3	0	26.0	7.9250	S
3	1	1	0	35.0	53.1000	S
4	0	3	1	35.0	8.0500	S
...	...	...	...	...	...	...
886	0	2	1	27.0	13.0000	S
887	1	1	0	19.0	30.0000	S
888	0	3	0	28.0	23.4500	S
889	1	1	1	26.0	30.0000	C
890	0	3	1	32.0	7.7500	Q

891 rows × 6 columns

```
In [18]: import warnings
warnings.filterwarnings("ignore")
```

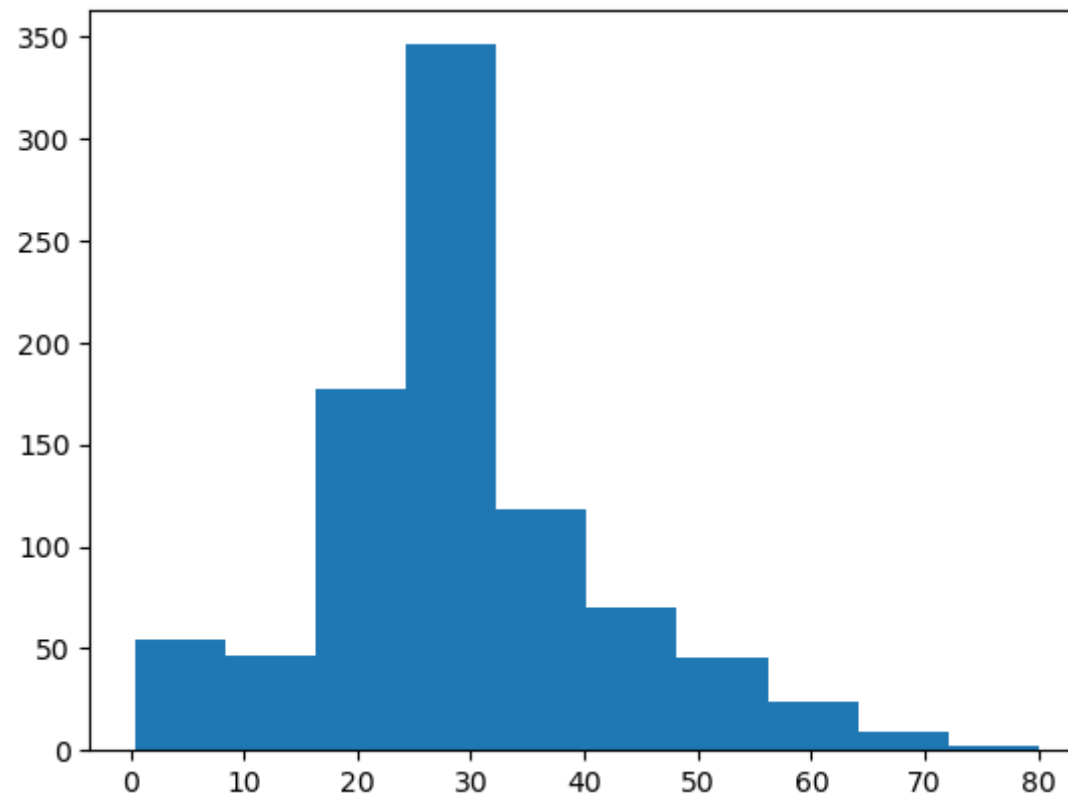
```
In [19]: import seaborn as hh  
import matplotlib.pyplot as plt  
hh.boxplot(data1.Age)
```

Out[19]: <Axes: >



```
In [20]: #plt.hist(data['Age'])  
plt.hist(data1['Age'])
```

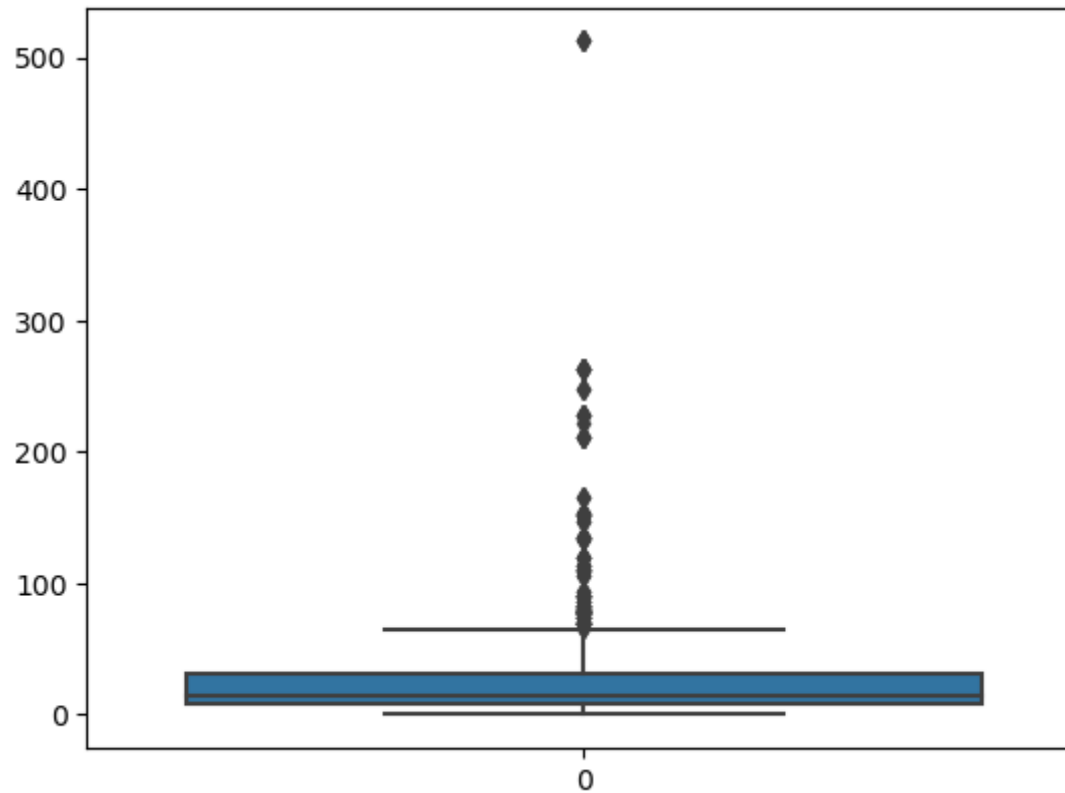
```
Out[20]: (array([ 54.,  46., 177., 346., 118.,  70.,  45.,  24.,   9.,   2.]),  
array([ 0.42 ,  8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126,  
        64.084, 72.042, 80.   ]),  
<BarContainer object of 10 artists>)
```





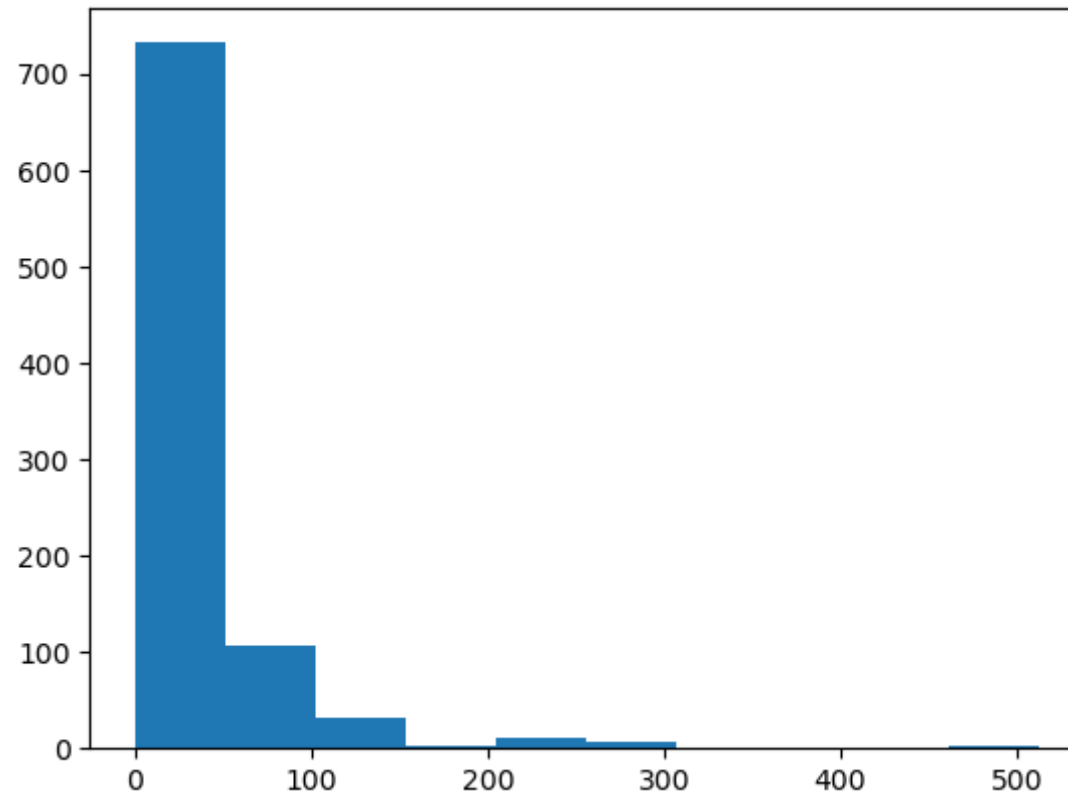
```
In [21]: import seaborn as hh  
import matplotlib.pyplot as plt  
hh.boxplot(data1.Fare)
```

Out[21]: <Axes: >



```
In [22]: plt.hist(data1['Fare'])
```

```
Out[22]: (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 [23]: data1.isna().sum()
```

```
Out[23]: Survived      0
          Pclass       0
          Sex          0
          Age          0
          Fare         0
          Embarked     2
          dtype: int64
```

```
In [24]: data1.describe()
```

```
Out[24]:
```

	Survived	Pclass	Sex	Age	Fare
<b>count</b>	891.000000	891.000000	891.000000	891.000000	891.000000
<b>mean</b>	0.383838	2.308642	0.647587	29.361582	32.204208
<b>std</b>	0.486592	0.836071	0.477990	13.019697	49.693429
<b>min</b>	0.000000	1.000000	0.000000	0.420000	0.000000
<b>25%</b>	0.000000	2.000000	0.000000	22.000000	7.910400
<b>50%</b>	0.000000	3.000000	1.000000	28.000000	14.454200
<b>75%</b>	1.000000	3.000000	1.000000	35.000000	31.000000
<b>max</b>	1.000000	3.000000	1.000000	80.000000	512.329200

```
In [25]: data1['Age'].unique()
```

```
Out[25]: 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 [26]: data1.groupby(['Age']).count()
```

```
Out[26]:
```

	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 [27]: data1['Pclass']=data1['Pclass'].map({1:'F',2:'S',3:'Third'})
```

```
In [28]: data1
```

```
Out[28]:
```

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

891 rows × 6 columns

```
In [29]: data1=pd.get_dummies(data1)
```

```
In [30]: data1.shape
```

```
Out[30]: (891, 10)
```

```
In [31]: data1
```

```
Out[31]:
```

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

891 rows × 10 columns

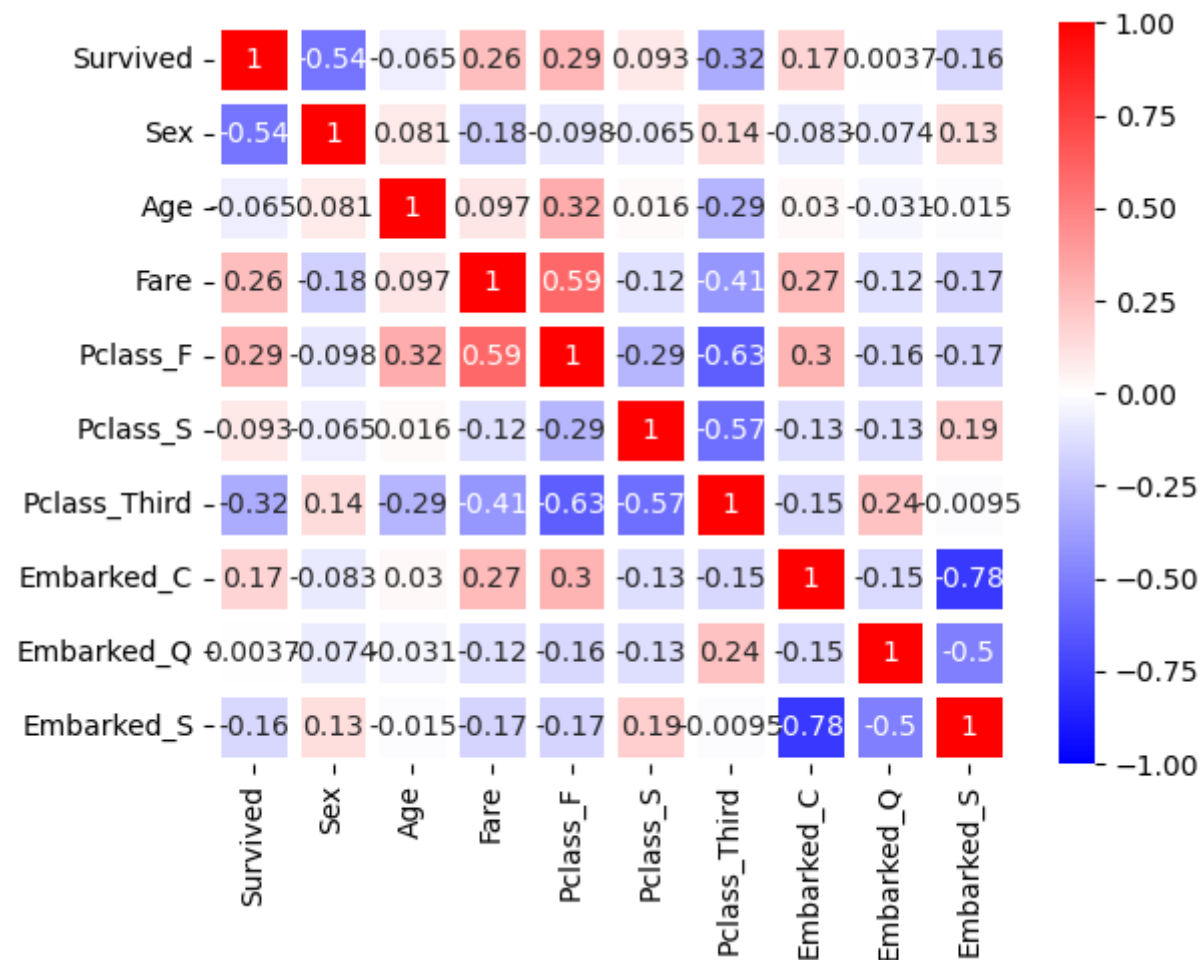
```
In [32]: cor=data1.corr()  
cor
```

Out[32]:

	Survived	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_Third	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_Third	-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 [33]: import seaborn as hh
hh.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidth=5,cmap='bwr')
```

Out[33]: <Axes: >





```
In [34]: data1.groupby('Survived').count()
```

Out[34]:

	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_Third	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 [35]: y=data1['Survived']  
x=data1.drop('Survived',axis=1)
```

```
In [36]: 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 [37]: x_test.head(5)
```

Out[37]:

	Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_Third	Embarked_C	Embarked_Q	Embarked_S
709	1	28.0	15.2458	0	0	1	1	0	0
439	1	31.0	10.5000	0	1	0	0	0	1
840	1	20.0	7.9250	0	0	1	0	0	1
720	0	6.0	33.0000	0	1	0	0	0	1
39	0	14.0	11.2417	0	0	1	1	0	0

```
In [38]: y_test.head(5)
```

```
Out[38]: 709    1
         439    0
         840    0
         720    1
          39    1
         Name: Survived, dtype: int64
```

```
In [39]: x_train.head(5)
```

```
Out[39]:
```

		Sex	Age	Fare	Pclass_F	Pclass_S	Pclass_Third	Embarked_C	Embarked_Q	Embarked_S
	<b>6</b>	1	54.0	51.8625	1	0	0	0	0	1
	<b>718</b>	1	28.0	15.5000	0	0	1	0	1	0
	<b>685</b>	1	25.0	41.5792	0	1	0	1	0	0
	<b>73</b>	1	26.0	14.4542	0	0	1	1	0	0
	<b>882</b>	0	22.0	10.5167	0	0	1	0	0	1

```
In [40]: y_train.head(5)
```

```
Out[40]: 6      0
         718    0
         685    0
          73    0
         882    0
         Name: Survived, dtype: int64
```

```
In [47]: from sklearn.linear_model import LogisticRegression
classifier=LogisticRegression()
classifier.fit(x_train,y_train)
```

```
Out[47]: ▾ LogisticRegression
LogisticRegression()
```

```
In [48]: y_pred=classifier.predict(x_test)
y_pred
```

```
Out[48]: 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, 0, 1, 1, 0, 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, 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 [49]: from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

```
Out[49]: array([[154,  21],
[ 37,  83]])
```

```
In [50]: from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

```
Out[50]: 0.8033898305084746
```

In [51]:

y

Out[51]:

0	0
1	1
2	1
3	1
4	0
	..
886	0
887	1
888	0
889	1
890	0

Name: Survived, Length: 891, dtype: int64

In [ ]: