

# Import the libraries

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

# Import the Dataset

```
In [2]: dataset=pd.read_csv("Titanic-Dataset.csv")
```

```
In [3]: dataset
```

	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	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W.C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

891 rows × 12 columns

```
In [4]: dataset.head()
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.283
2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

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

Out[5]:	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W.C. 6607	23.45
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75

```
In [6]: dataset.shape
```

```
Out[6]: (891, 12)
```

```
In [7]: dataset.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 [8]: `dataset.describe()`

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

In [11]: `corr1=dataset.corr()`  
`corr1`

```
C:\Users\bunny\AppData\Local\Temp\ipykernel_11192\1291402269.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
corr1=dataset.corr()
```

	<b>PassengerId</b>	<b>Survived</b>	<b>Pclass</b>	<b>Age</b>	<b>SibSp</b>	<b>Parch</b>	<b>Fare</b>
<b>PassengerId</b>	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658
<b>Survived</b>	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
<b>Pclass</b>	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
<b>Age</b>	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067
<b>SibSp</b>	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651
<b>Parch</b>	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225
<b>Fare</b>	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000

In [12]: `plt.subplots(figsize=(20,15))`  
`sns.heatmap(corr1, annot=True)`

Out[12]: <Axes: >



```
In [13]: dataset.Fare.value_counts()
```

```
Out[13]:
```

8.0500	43
13.0000	42
7.8958	38
7.7500	34
26.0000	31
..	
35.0000	1
28.5000	1
6.2375	1
14.0000	1
10.5167	1

Name: Fare, Length: 248, dtype: int64

```
In [14]: dataset.Survived.value_counts()
```

```
Out[14]:
```

0	549
1	342

Name: Survived, dtype: int64

```
In [15]: dataset.Pclass.value_counts()
```

```
Out[15]:
```

3	491
1	216
2	184

Name: Pclass, dtype: int64

## Handling null values

```
In [16]: dataset.isnull().any()
```

PassengerId	False
-------------	-------

```
Out[16]: 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 [17]: dataset.isnull().sum()
```

```
Out[17]: 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 [18]: dataset.head()
```

	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	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

## Data Visualization

```
In [19]: pip install seaborn
```

```
Requirement already satisfied: seaborn in d:\python\lib\site-packages (0.  
12.2)Note: you may need to restart the kernel to use updated packages.
```

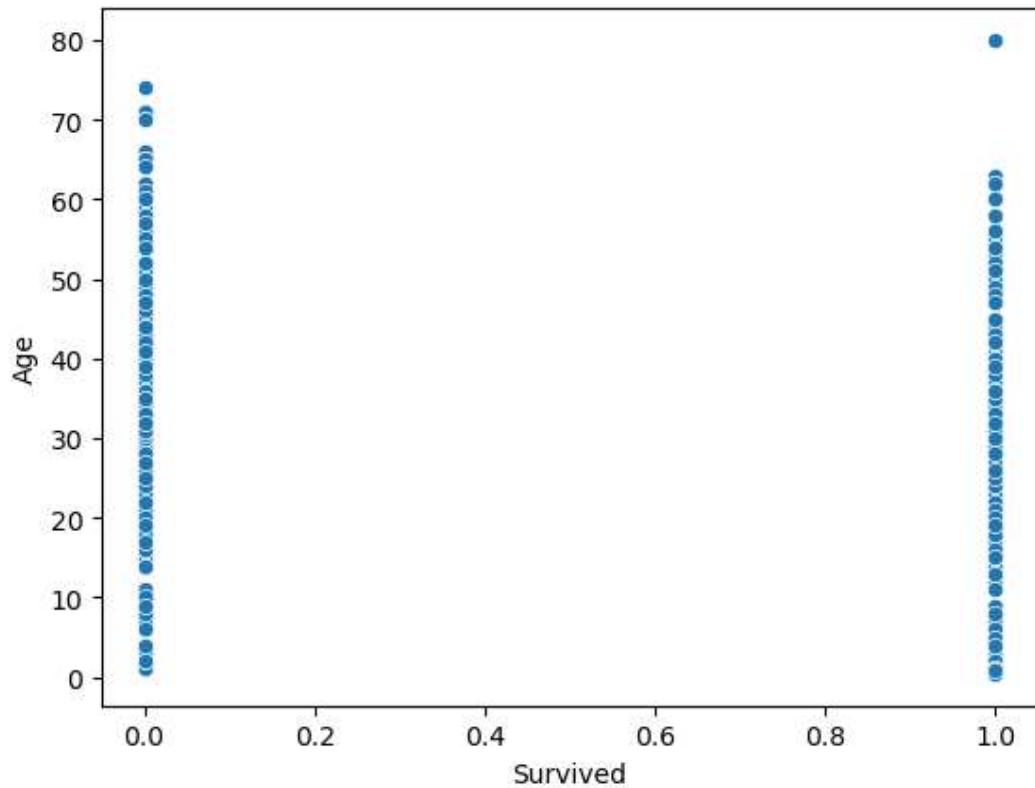
```
Requirement already satisfied: numpy!=1.24.0,>=1.17 in d:\python\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in d:\python\lib\site-packages (from seaborn) (1.5.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in d:\python\lib\site-packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\bunny\appdata\roaming\python\python311\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\python\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\bunny\appdata\roaming\python\python311\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in d:\python\lib\site-packages (from pandas>=0.25->seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in c:\users\bunny\appdata\roaming\python\python311\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

```
In [20]: import seaborn as sns
print(sns.get_dataset_names())

['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes', 'diamonds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glove', 'healthexp', 'iris', 'mpg', 'penguins', 'planets', 'seairce', 'taxis', 'tips', 'titanic']
```

```
In [21]: sns.__version__
Out[21]: '0.12.2'
```

```
In [32]: sns.scatterplot(x="Survived", y="Age", data=dataset)
Out[32]: <Axes: xlabel='Survived', ylabel='Age'>
```



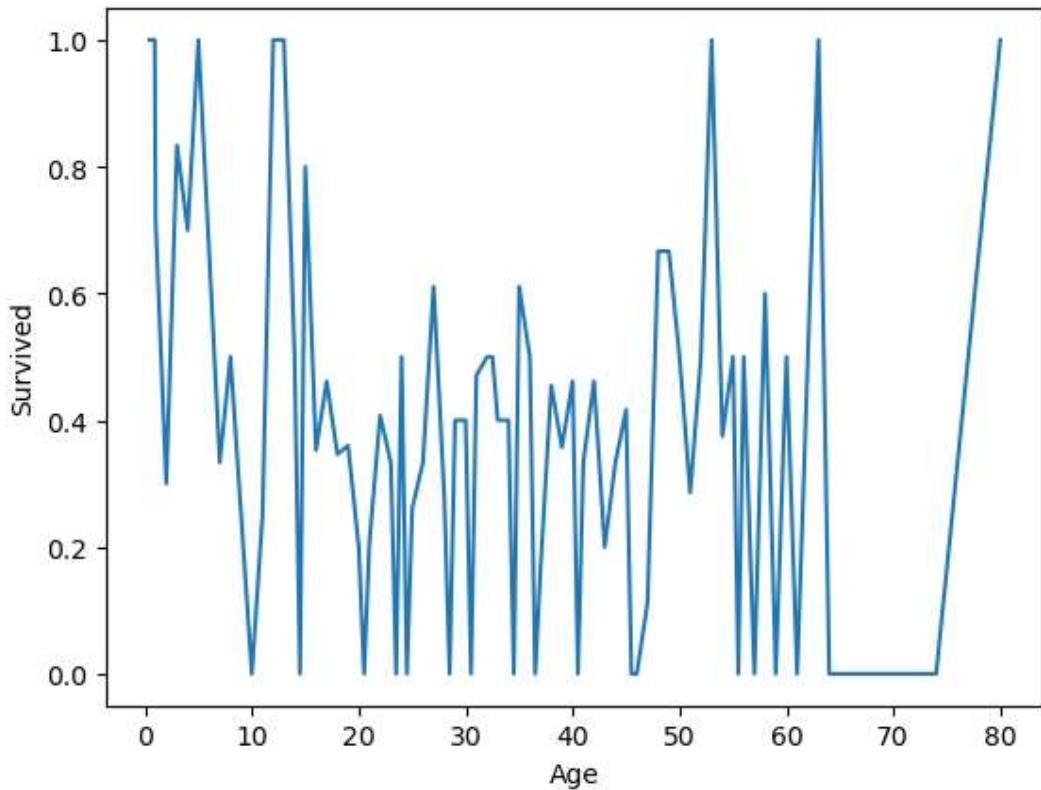
```
In [34]: sns.lineplot(x="Age", y="Survived", data=dataset, ci=None)
```

```
C:\Users\bunny\AppData\Local\Temp\ipykernel_11192\2977963200.py:1: FutureWarning:
```

```
The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.
```

```
    sns.lineplot(x="Age", y="Survived", data=dataset, ci=None)
```

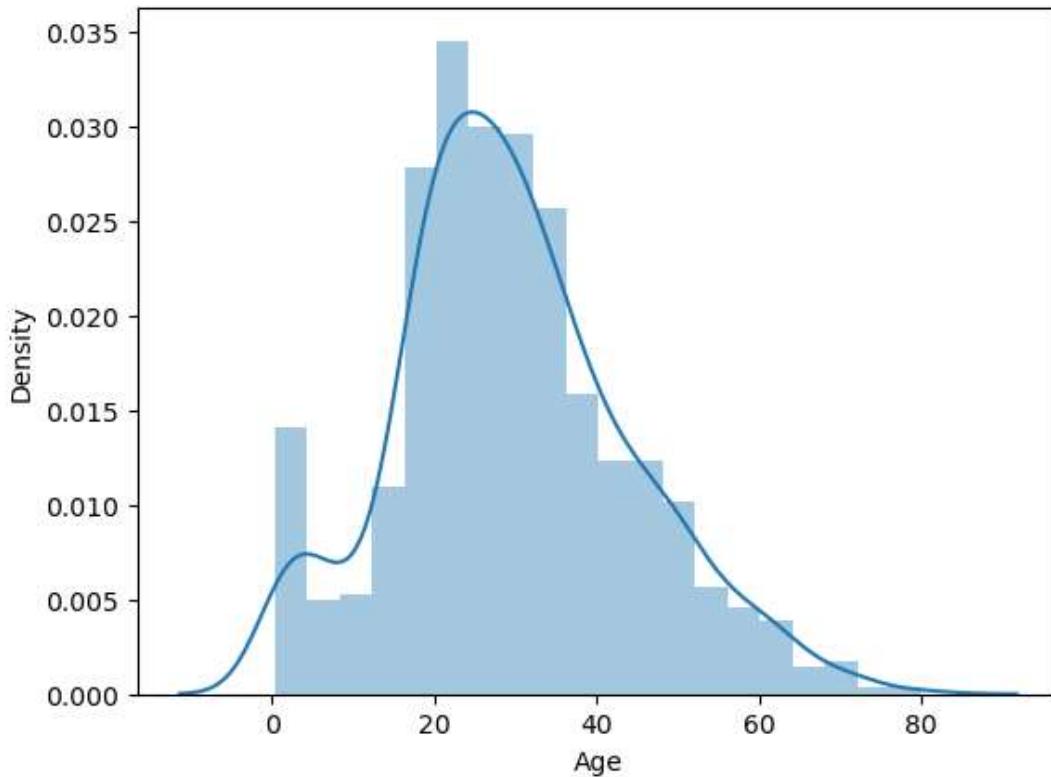
```
Out[34]: <Axes: xlabel='Age', ylabel='Survived'>
```



```
In [36]: sns.distplot(dataset["Age"])
```

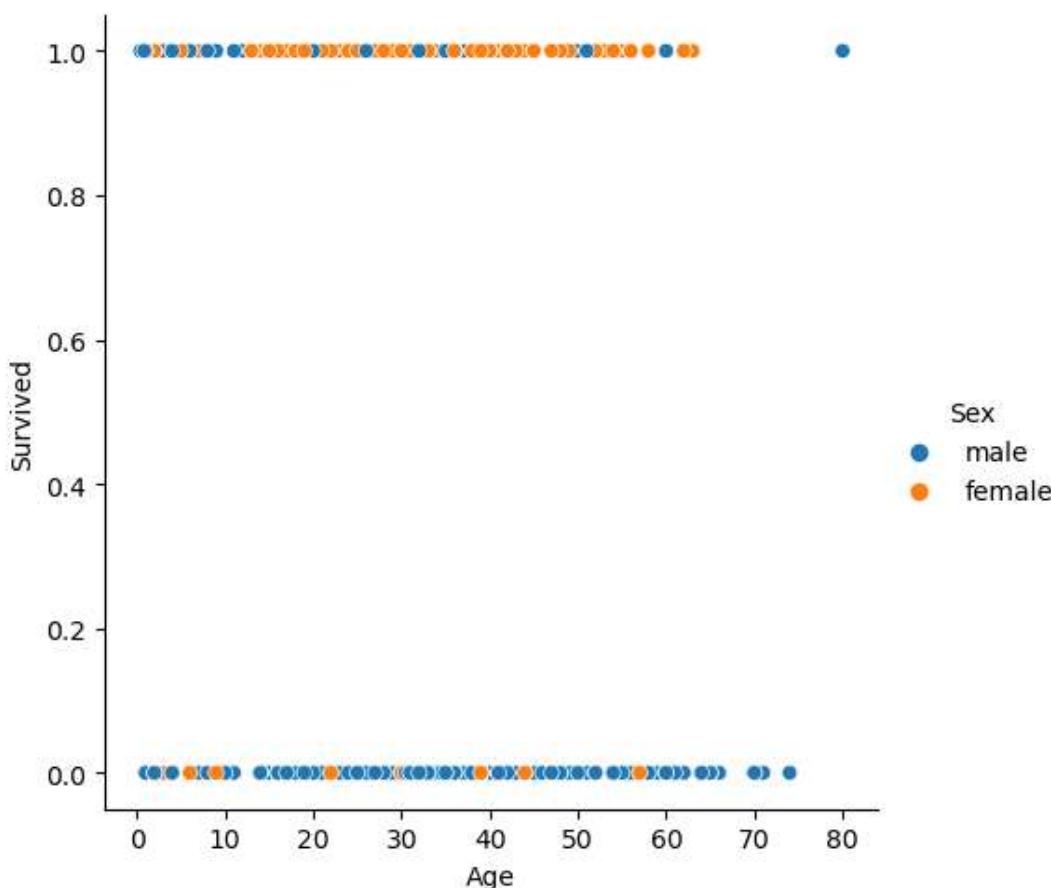
```
C:\Users\bunny\AppData\Local\Temp\ipykernel_11192\642544636.py:1: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.  
  
Please adapt your code to use either `displot` (a figure-level function with  
similar flexibility) or `histplot` (an axes-level function for histograms).  
  
For a guide to updating your code to use the new functions, please see  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
    sns.distplot(dataset["Age"])
Out[36]: <Axes: xlabel='Age', ylabel='Density'>
```



```
In [44]: sns.relplot(x="Age", y="Survived", data=dataset, hue="Sex")
```

```
Out[44]: <seaborn.axisgrid.FacetGrid at 0x22ef5f667d0>
```



```
In [45]: dataset["Fare"].value_counts()
```

```
Out[45]: 8.0500    43
         13.0000   42
         7.8958    38
         7.7500    34
         26.0000   31
           ..
        35.0000    1
        28.5000    1
        6.2375    1
       14.0000    1
      10.5167    1
Name: Fare, Length: 248, dtype: int64
```

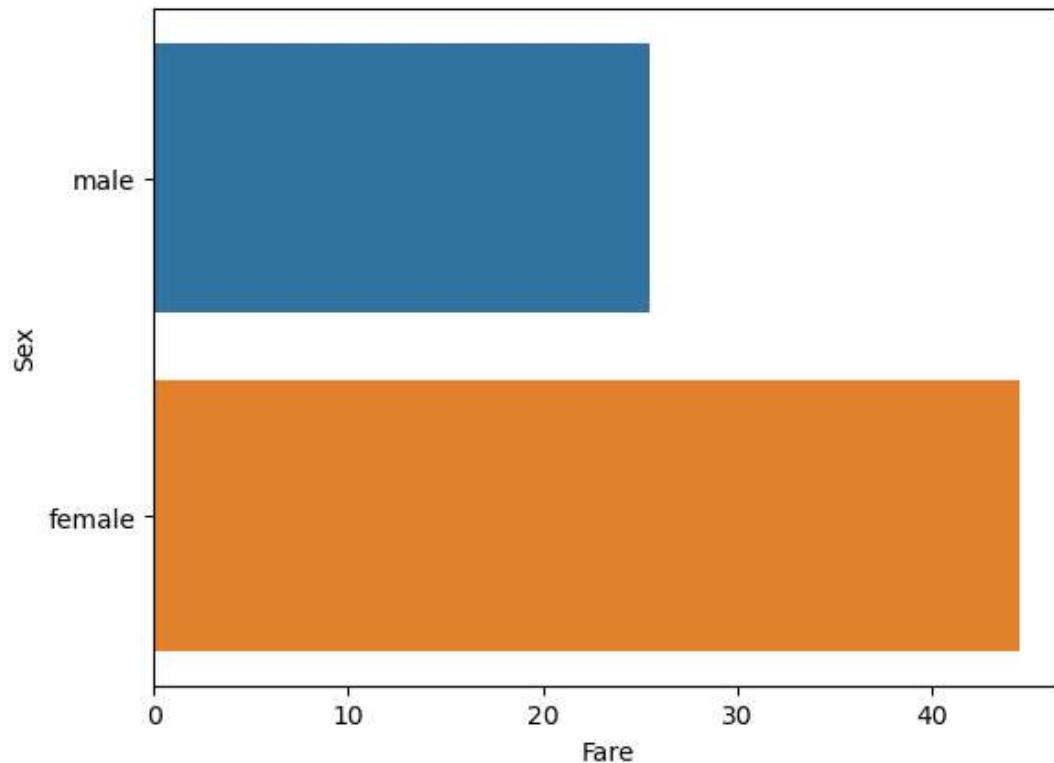
```
In [50]: sns.barplot(data=dataset,x="Fare",y="Sex",ci=None)
```

```
C:\Users\bunny\AppData\Local\Temp\ipykernel_11192\3052055135.py:1: Future
Warning:
```

```
The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.
```

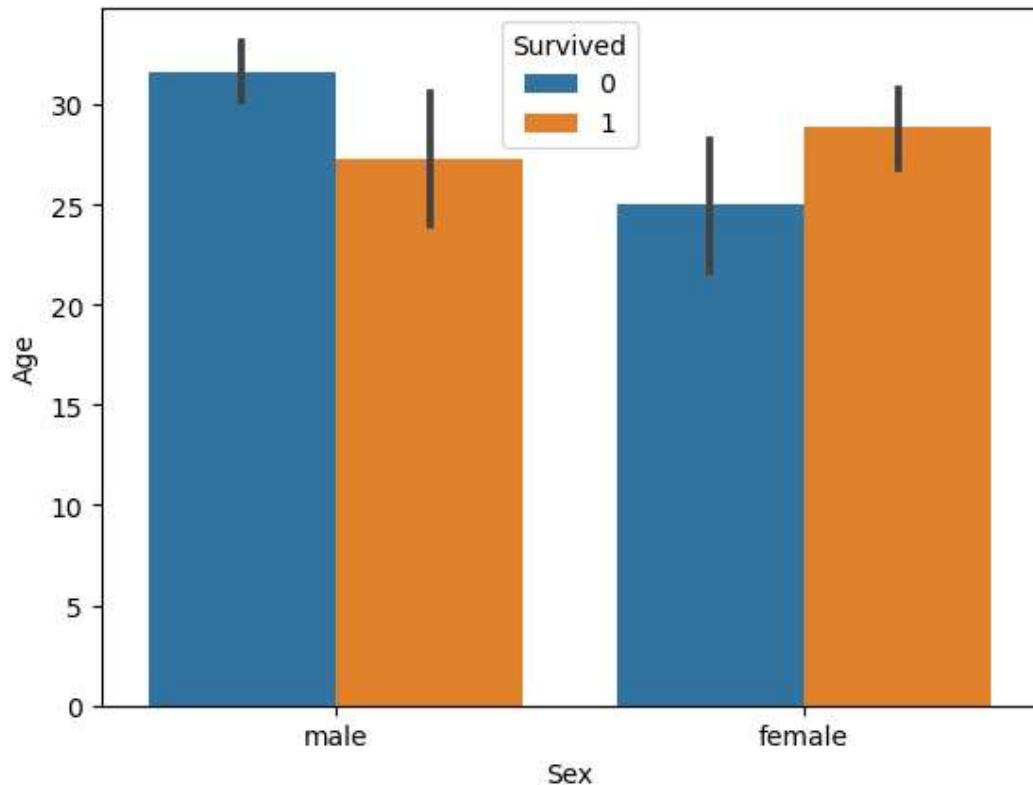
```
sns.barplot(data=dataset,x="Fare",y="Sex",ci=None)
```

```
Out[50]: <Axes: xlabel='Fare', ylabel='Sex'>
```



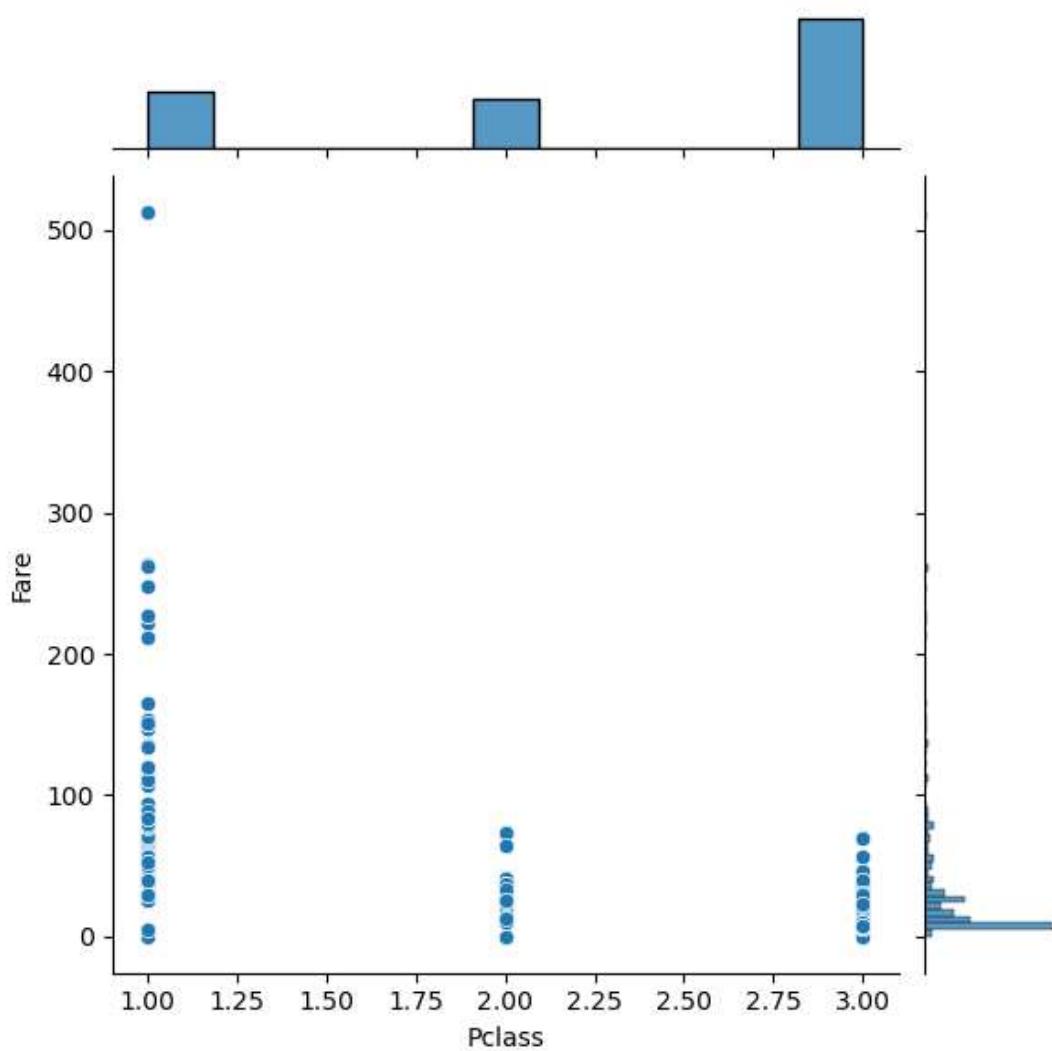
```
In [57]: sns.barplot(data=dataset,x="Sex",y="Age",hue="Survived")
```

```
Out[57]: <Axes: xlabel='Sex', ylabel='Age'>
```



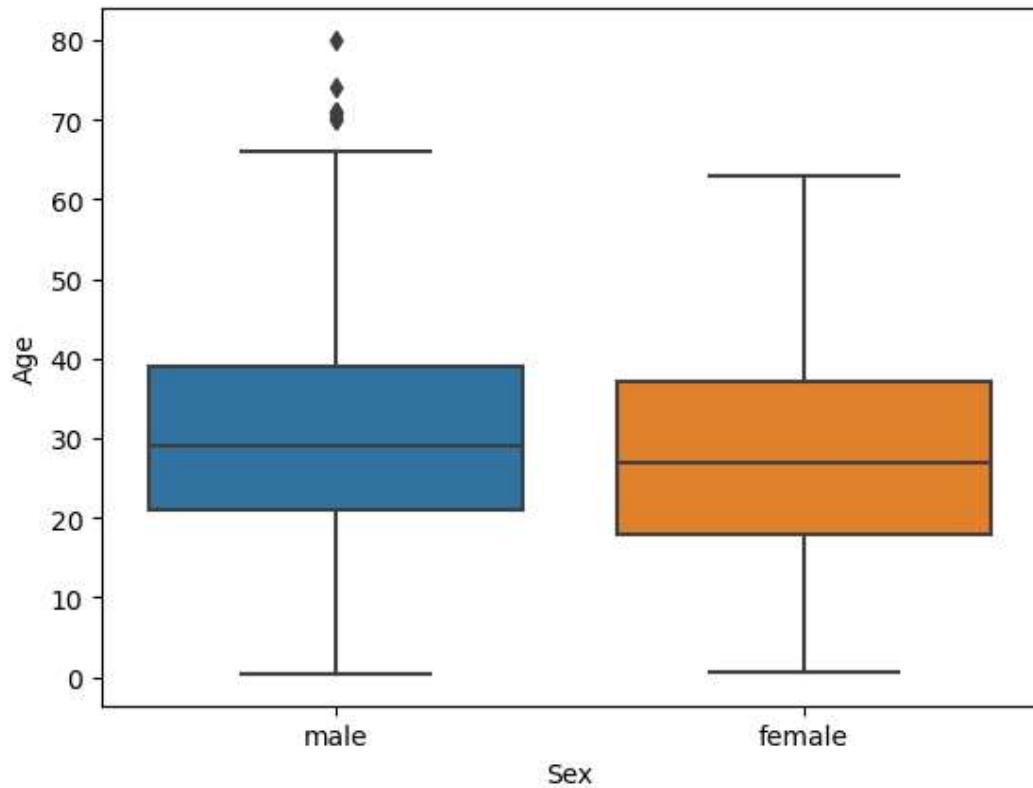
```
In [58]: sns.jointplot(x="Pclass", y="Fare", data=dataset)
```

```
Out[58]: <seaborn.axisgrid.JointGrid at 0x22e81faee10>
```



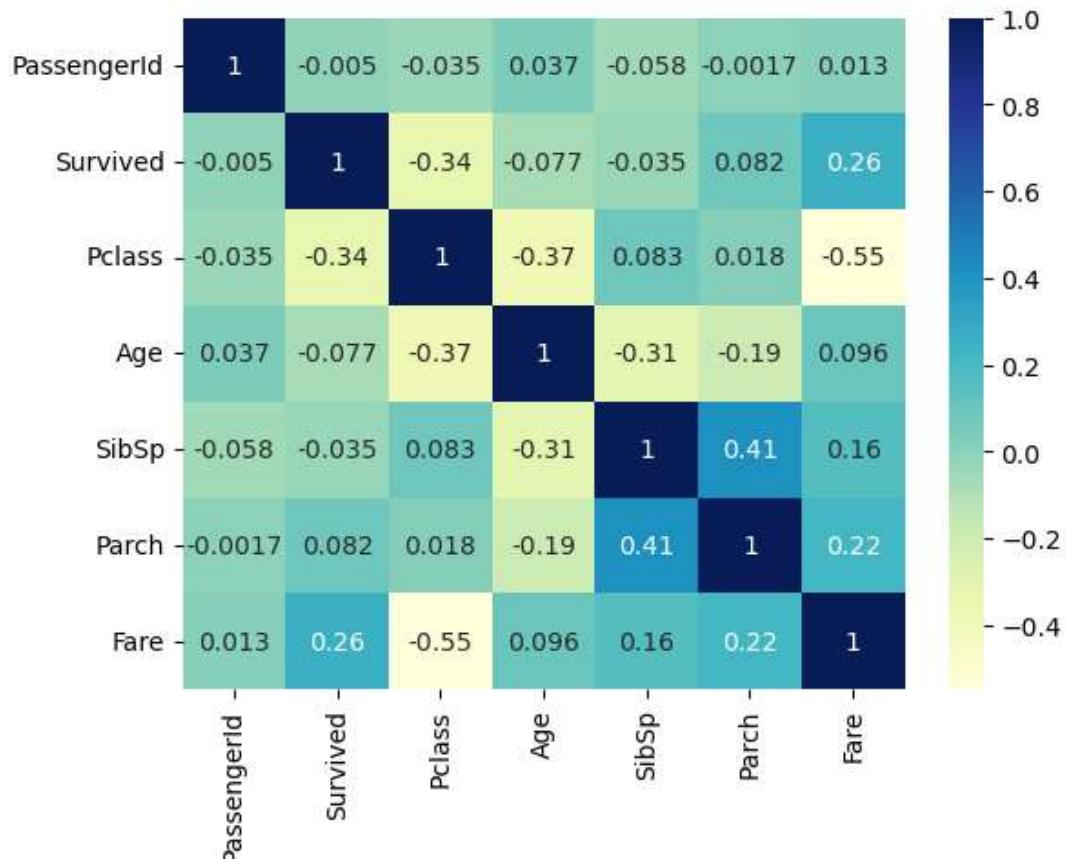
```
In [63]: sns.boxplot(x="Sex", y="Age", data=dataset)
```

```
Out[63]: <Axes: xlabel='Sex', ylabel='Age'>
```



```
In [64]: sns.heatmap(corr1, annot=True, cmap="YlGnBu")
```

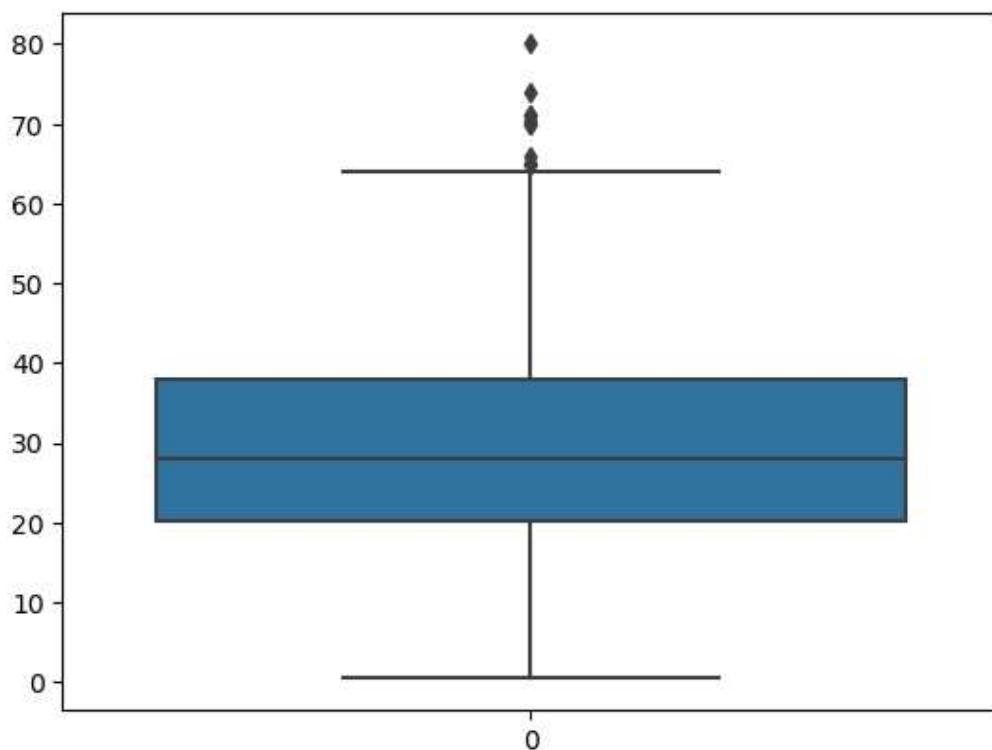
```
Out[64]: <Axes: >
```



# Outlier Detection

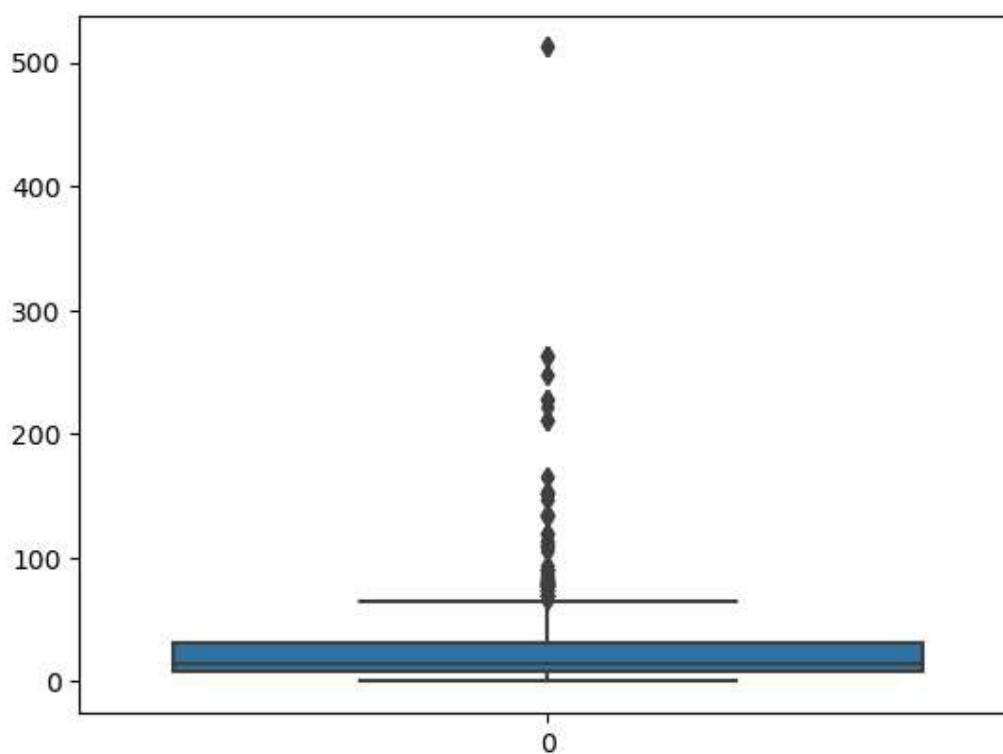
```
In [65]: sns.boxplot(dataset.Age)
```

```
Out[65]: <Axes: >
```



```
In [75]: sns.boxplot(dataset.Fare)
```

```
Out[75]: <Axes: >
```



# Splitting dependent and independent variables

```
In [213]: x=dataset.iloc[:, 3:13]
y=dataset.iloc[:, 10:12]
```

```
In [214]: x.head()
```

```
Out[214]:
```

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	Cumings, Mrs. John Bradley (Florence Briggs Th... er)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
In [185]: y.head()
```

```
Out[185]:
```

	Cabin	Embarked
0	NaN	S
1	C85	C
2	NaN	S
3	C123	S
4	NaN	S

## Perform Encoding

```
In [76]: from sklearn.preprocessing import LabelEncoder
```

```
In [77]: le=LabelEncoder()
```

```
In [190...]: x[ "Name" ]=le.fit_transform(x[ "Name" ])

x[ "Name" ]
```

```
Out[190]:
```

0	108
1	190
2	353
3	272
4	15
	...
886	548

```
887      303
888      413
889       81
890      220
Name: Name, Length: 891, dtype: int32
```

```
In [170...]: x["Sex"]
```

```
Out[170]: 0      male
1    female
2    female
3    female
4      male
...
886      male
887  female
888  female
889      male
890      male
Name: Sex, Length: 891, dtype: object
```

```
In [172...]: x["Sex"].value_counts()
```

```
Out[172]: male      577
female     314
Name: Sex, dtype: int64
```

```
In [158...]: x.Name.value_counts()
```

```
Out[158]: 108      1
98       1
267      1
284      1
566      1
...
431      1
518      1
411      1
428      1
220      1
Name: Name, Length: 891, dtype: int64
```

```
In [173...]: x["Sex"].nunique()
```

```
Out[173]: 2
```

```
In [177...]: Name=pd.get_dummies(x["Name"], drop_first=True)
```

```
In [178...]: Name
```

```
Out[178]:   1  2  3  4  5  6  7  8  9  10 ... 881 882 883 884 885 886 887 888 889
  0  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
  1  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
  2  0  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
  3  0  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
  4  0  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
...
886  0  0  0  0  0  0  0  0  0  0 ...  0  0  0  0  0  0  0  0  0
```

887	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
888	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
889	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
890	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0

891 rows × 890 columns

In [203]: `x=pd.concat([x,Name],axis=1)`

In [204]: `x`

Out[204]:

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

891 rows × 899 columns

# splitting into training and testing set

```
In [206...]: from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test_split(k,y,test_size=0.3,random_s
```

```
In [216...]: x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

Out[216]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	1	...
857	Daly, Mr. Peter Denis	male	51.0	0	0	113055	26.5500	E17	S	0	...
52	Harper, Mrs. Henry Sleeper (Myra Haxton)	female	49.0	1	0	PC 17572	76.7292	D33	C	0	...
386	Goodwin, Master. Sidney Leonard	male	1.0	5	2	CA 2144	46.9000	NaN	S	0	...
124	White, Mr. Percival Wayland	male	54.0	0	1	35281	77.2875	D26	S	0	...
578	Caram, Mrs. Joseph (Maria Elias)	female	NaN	1	0	2689	14.4583	NaN	C	0	...
...	...	...	...	...	...	...	...	...	...	...	...
835	Compton, Miss. Sara Rebecca	female	39.0	1	1	PC 17756	83.1583	E49	C	0	...
192	Andersen-Jensen, Miss. Carla Christine Nielsine	female	19.0	1	0	350046	7.8542	NaN	S	0	...
629	O'Connell, Mr. Patrick D	male	NaN	0	0	334912	7.7333	NaN	Q	0	...
559	de Messemaeker, Mrs. Guillaume Joseph (Emma)	female	36.0	1	0	345572	17.4000	NaN	S	0	...
684	Brown, Mr. Thomas William Solomon	male	60.0	1	1	29750	39.0000	NaN	S	0	...

623 rows × 899 columns

```
In [223...]: from sklearn.preprocessing import StandardScaler  
sc=StandardScaler()
```

```
In [224...]: x_train
```

Out[224]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	1	...
857	Daly, Mr. Peter Denis	male	51.0	0	0	113055	26.5500	E17	S	0	...

52	Harper, Mrs. Henry Sleeper (Myra Haxton)	female	49.0	1	0	PC 17572	76.7292	D33	C	0	...
386	Goodwin, Master. Sidney Leonard	male	1.0	5	2	CA 2144	46.9000	NaN	S	0	...
124	White, Mr. Percival Wayland	male	54.0	0	1	35281	77.2875	D26	S	0	...
578	Caram, Mrs. Joseph (Maria Elias)	female	Nan	1	0	2689	14.4583	NaN	C	0	...
...	...	...	...	...	...	...	...	...	...	...	...
835	Compton, Miss. Sara Rebecca	female	39.0	1	1	PC 17756	83.1583	E49	C	0	...
192	Andersen- Jensen, Miss. Carla Christine Nielsine	female	19.0	1	0	350046	7.8542	NaN	S	0	...
629	O'Connell, Mr. Patrick D	male	Nan	0	0	334912	7.7333	NaN	Q	0	...
559	de Messemaeker, Mrs. Guillaume Joseph (Emma)	female	36.0	1	0	345572	17.4000	NaN	S	0	...
684	Brown, Mr. Thomas William Solomon	male	60.0	1	1	29750	39.0000	NaN	S	0	...

623 rows × 899 columns

In [225...]	x_test										
Out[225]:											
495	Yousseff, Mr. Gerious	male	Nan	0	0	2627	14.4583	NaN	C	0	...
648	Willey, Mr. Edward	male	Nan	0	0	S.O./P.P. 751	7.5500	NaN	S	0	...
278	Rice, Master. Eric	male	7.0	4	1	382652	29.1250	NaN	Q	0	...
31	Spencer, Mrs. William Augustus (Marie Eugenie)	female	Nan	1	0	PC 17569	146.5208	B78	C	0	...
255	Touma, Mrs. Darwis	female	29.0	0	2	2650	15.2458	NaN	C	0	...

(Hanne  
Youssef  
Razi)

...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
263	Harrison, Mr. William	male	40.0	0	0	112059	0.0000	B94	S	0	...			
718	McEvoy, Mr. Michael	male	NaN	0	0	36568	15.5000	NaN	Q	0	...			
620	Yasbeck, Mr. Antoni	male	27.0	1	0	2659	14.4542	NaN	C	0	...			
786	Sjoblom, Miss. Anna Sofia	female	18.0	0	0	3101265	7.4958	NaN	S	0	...			
64	Stewart, Mr. Albert A	male	NaN	0	0	PC 17605	27.7208	NaN	C	0	...			

268 rows × 899 columns

In [226]: y\_train

	Cabin	Embarked
857	E17	S
52	D33	C
386	NaN	S
124	D26	S
578	NaN	C
...	...	...
835	E49	C
192	NaN	S
629	NaN	Q
559	NaN	S
684	NaN	S

623 rows × 2 columns

In [227]: y\_test

	Cabin	Embarked
495	NaN	C
648	NaN	S
278	NaN	Q
31	B78	C
255	NaN	C

...	...	...
263	B94	S
718	NaN	Q
620	NaN	C
786	NaN	S
64	NaN	C

268 rows × 2 columns

```
In [234]: x.columns = x.columns.astype(str)
x.columns
Out[234]: Index(['Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin',
       'Embarked'],
              dtype='object')
```

```
In [ ]: x_train=sc.fit_transform(x_train)
x_test=sc.fit_transform(x_test)
```

```
In [228]: a=[3,4,5,6,7,8]
b=[0,1,3,4,2,7]

for i in range(10):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.3,random_state=i)
    print("with random state",a_train)

with random state [7, 6, 8, 3]
```

```
In [121]: a=[3,4,5,6,7,8]
b=[0,1,3,4,2,7]

for i in range(10):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.3,random_state=None)
    print("without random state",a_train)

without random state [8, 4, 5, 7]
without random state [5, 8, 4, 7]
without random state [7, 4, 5, 8]
without random state [8, 7, 6, 3]
without random state [5, 7, 6, 3]
without random state [5, 4, 8, 6]
without random state [4, 3, 5, 6]
without random state [8, 5, 4, 3]
without random state [3, 5, 7, 8]
without random state [8, 6, 4, 5]
```