

# ot-september-15th-correctedversion

September 28, 2023

Importing the necessary libraries

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

Import the dataset

```
[60]: dataset = pd.read_csv("/content/Titanic-Dataset.csv")
dataset
```

```
[60]:      PassengerId  Survived  Pclass  \
0             1         0         3
1             2         1         1
2             3         1         3
3             4         1         1
4             5         0         3
..          ...          ...      ...
886          887         0         2
887          888         1         1
888          889         0         3
889          890         1         1
890          891         0         3
```

```
      Name      Sex  Age  SibSp  \
0  Braund, Mr. Owen Harris    male  22.0    1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0    1
2  Heikkinen, Miss. Laina    female  26.0    0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    female  35.0    1
4  Allen, Mr. William Henry    male  35.0    0
..          ...          ...      ...
886  Montvila, Rev. Juozas    male  27.0    0
887  Graham, Miss. Margaret Edith    female  19.0    0
888  Johnston, Miss. Catherine Helen "Carrie"    female   NaN    1
889  Behr, Mr. Karl Howell    male  26.0    0
890  Dooley, Mr. Patrick    male  32.0    0
```

|     | Parch | Ticket           | Fare    | Cabin | Embarked |
|-----|-------|------------------|---------|-------|----------|
| 0   | 0     | A/5 21171        | 7.2500  | NaN   | S        |
| 1   | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2   | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3   | 0     | 113803           | 53.1000 | C123  | S        |
| 4   | 0     | 373450           | 8.0500  | NaN   | S        |
| ..  | ...   | ...              | ...     | ...   | ...      |
| 886 | 0     | 211536           | 13.0000 | NaN   | S        |
| 887 | 0     | 112053           | 30.0000 | B42   | S        |
| 888 | 2     | W./C. 6607       | 23.4500 | NaN   | S        |
| 889 | 0     | 111369           | 30.0000 | C148  | C        |
| 890 | 0     | 370376           | 7.7500  | NaN   | Q        |

[891 rows x 12 columns]

```
[61]: dataset.head()
```

```
[61]: PassengerId  Survived  Pclass  \
0             1         0         3
1             2         1         1
2             3         1         3
3             4         1         1
4             5         0         3
```

|   | Name  | Sex    | Age  | SibSp | \ |
|---|---|--------|------|-------|---|
| 0 | Braund, Mr. Owen Harris                           | male   | 22.0 | 1     |   |
| 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1     |   |
| 2 | Heikkinen, Miss. Laina                            | female | 26.0 | 0     |   |
| 3 | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35.0 | 1     |   |
| 4 | Allen, Mr. William Henry                          | male   | 35.0 | 0     |   |

|   | Parch | Ticket           | Fare    | Cabin | Embarked |
|---|-------|------------------|---------|-------|----------|
| 0 | 0     | A/5 21171        | 7.2500  | NaN   | S        |
| 1 | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2 | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3 | 0     | 113803           | 53.1000 | C123  | S        |
| 4 | 0     | 373450           | 8.0500  | NaN   | S        |

```
[62]: dataset.tail()
```

```
[62]: PassengerId  Survived  Pclass  Name  \
886          887         0         2  Montvila, Rev. Juozas
887          888         1         1  Graham, Miss. Margaret Edith
888          889         0         3  Johnston, Miss. Catherine Helen "Carrie"
889          890         1         1  Behr, Mr. Karl Howell
890          891         0         3  Dooley, Mr. Patrick
```

|     | Sex    | Age  | SibSp | Parch | Ticket     | Fare  | Cabin | Embarked |
|-----|--------|------|-------|-------|------------|-------|-------|----------|
| 886 | male   | 27.0 | 0     | 0     | 211536     | 13.00 | NaN   | S        |
| 887 | female | 19.0 | 0     | 0     | 112053     | 30.00 | B42   | S        |
| 888 | female | NaN  | 1     | 2     | W./C. 6607 | 23.45 | NaN   | S        |
| 889 | male   | 26.0 | 0     | 0     | 111369     | 30.00 | C148  | C        |
| 890 | male   | 32.0 | 0     | 0     | 370376     | 7.75  | NaN   | Q        |

```
[63]: dataset.shape
```

```
[63]: (891, 12)
```

```
[64]: 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
```

```
[65]: dataset.describe()
```

```
[65]:
```

|       | PassengerId | Survived   | Pclass     | Age        | SibSp      | \ |
|-------|-------------|------------|------------|------------|------------|---|
| count | 891.000000  | 891.000000 | 891.000000 | 714.000000 | 891.000000 |   |
| mean  | 446.000000  | 0.383838   | 2.308642   | 29.699118  | 0.523008   |   |
| std   | 257.353842  | 0.486592   | 0.836071   | 14.526497  | 1.102743   |   |
| min   | 1.000000    | 0.000000   | 1.000000   | 0.420000   | 0.000000   |   |
| 25%   | 223.500000  | 0.000000   | 2.000000   | 20.125000  | 0.000000   |   |
| 50%   | 446.000000  | 0.000000   | 3.000000   | 28.000000  | 0.000000   |   |
| 75%   | 668.500000  | 1.000000   | 3.000000   | 38.000000  | 1.000000   |   |
| max   | 891.000000  | 1.000000   | 3.000000   | 80.000000  | 8.000000   |   |

|       | Parch      | Fare       |
|-------|------------|------------|
| count | 891.000000 | 891.000000 |

|      |          |            |
|------|----------|------------|
| mean | 0.381594 | 32.204208  |
| std  | 0.806057 | 49.693429  |
| min  | 0.000000 | 0.000000   |
| 25%  | 0.000000 | 7.910400   |
| 50%  | 0.000000 | 14.454200  |
| 75%  | 0.000000 | 31.000000  |
| max  | 6.000000 | 512.329200 |

## Data Visualization

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

```
<ipython-input-66-388b185de253>:1: UserWarning:
```

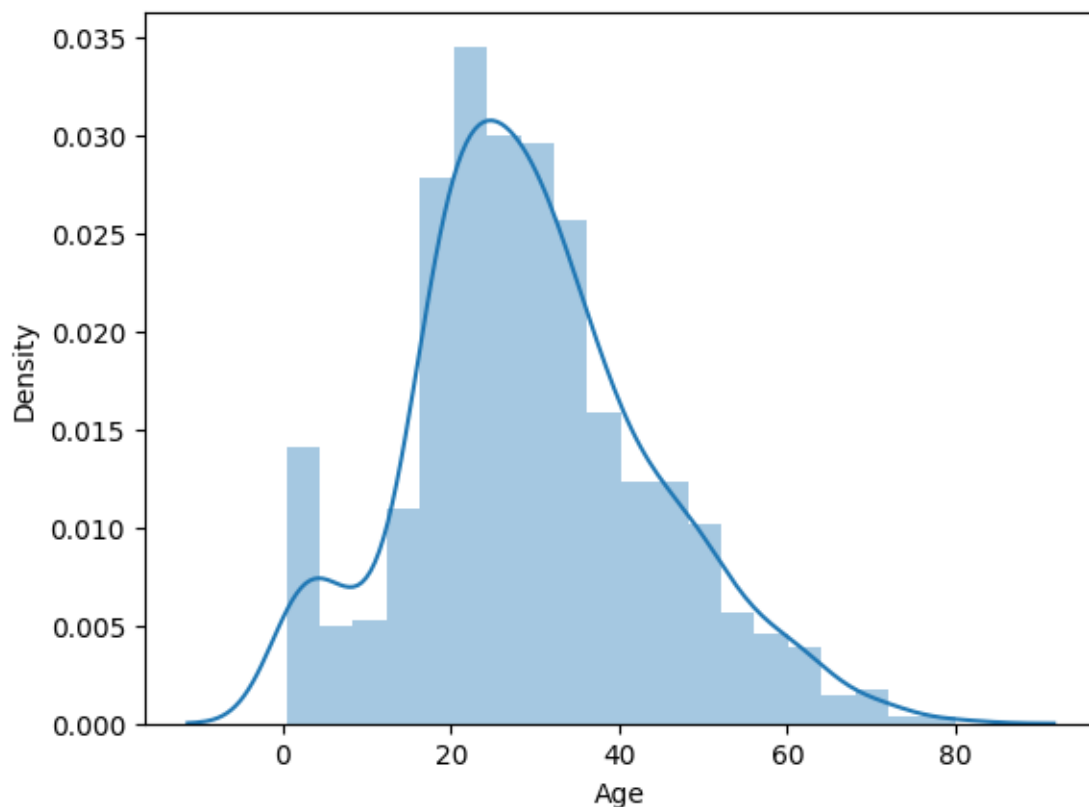
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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"])
```

```
[66]: <Axes: xlabel='Age', ylabel='Density'>
```



```
[67]: corr = dataset.corr()
      corr
```

<ipython-input-67-721137479888>: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.

```
corr = dataset.corr()
```

```
[67]:
```

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

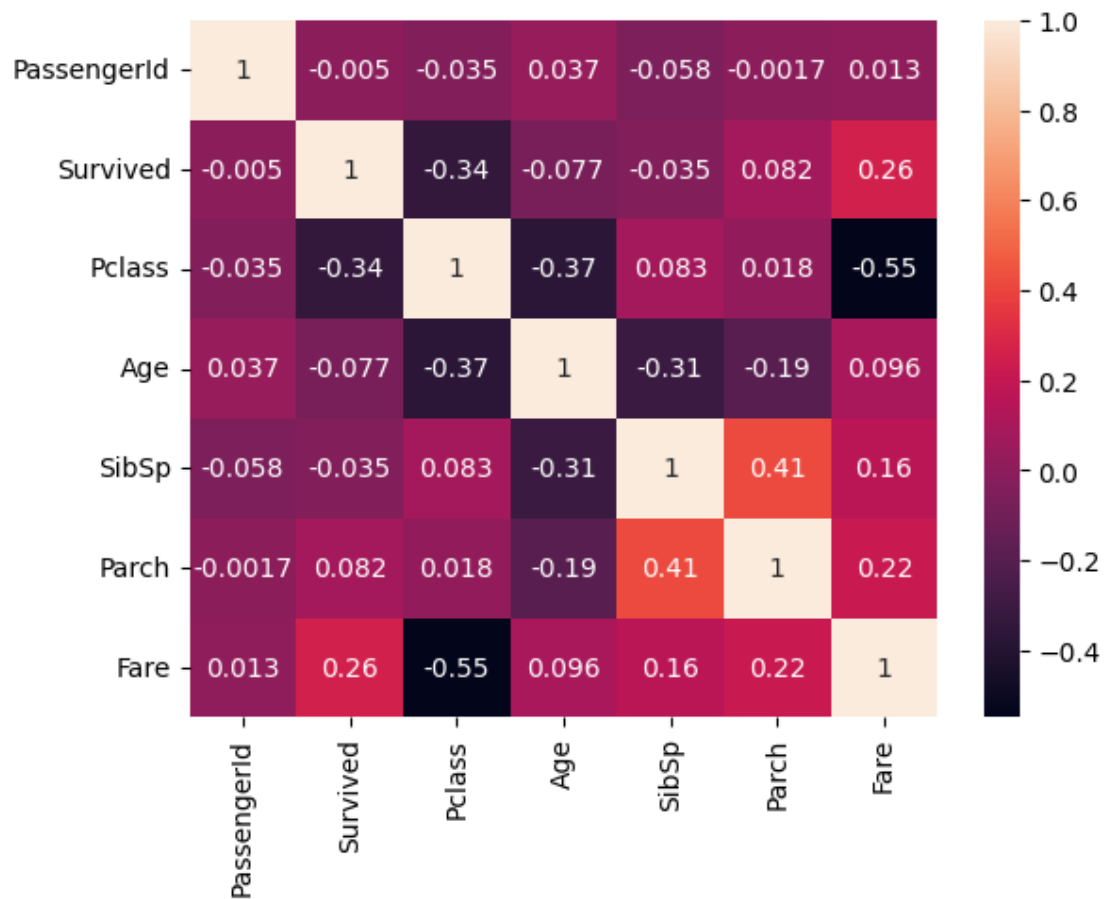
  

|             | Fare     |
|-------------|----------|
| PassengerId | 0.012658 |
| Survived    | 0.257307 |

```
Pclass      -0.549500
Age          0.096067
SibSp        0.159651
Parch        0.216225
Fare         1.000000
```

```
[68]: sns.heatmap(corr,annot=True)
```

```
[68]: <Axes: >
```



```
[69]: dataset.Embarked.value_counts()
```

```
[69]: S      644
      C      168
      Q       77
      Name: Embarked, dtype: int64
```

```
[70]: dataset.Sex.value_counts()
```

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

```
[71]: dataset.SibSp.value_counts()
```

```
[71]: 0      608
      1      209
      2       28
      4       18
      3       16
      8        7
      5         5
      Name: SibSp, dtype: int64
```

Handling Null Values

```
[72]: dataset.isnull().any()
```

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

```
[73]: dataset.isnull().sum()
```

```
[73]: 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

```
[74]: dataset["Age"].fillna(dataset["Age"].mean(),inplace=True)
```

```
[75]: dataset.head()
```

```
[75]:
```

|   | PassengerId | Survived | Pclass | \ | Name  | Sex    | Age  | SibSp | \ |
|---|-------------|----------|--------|---|---|--------|------|-------|---|
| 0 | 1           | 0        | 3      |   | Braund, Mr. Owen Harris                           | male   | 22.0 | 1     |   |
| 1 | 2           | 1        | 1      |   | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1     |   |
| 2 | 3           | 1        | 3      |   | Heikkinen, Miss. Laina                            | female | 26.0 | 0     |   |
| 3 | 4           | 1        | 1      |   | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35.0 | 1     |   |
| 4 | 5           | 0        | 3      |   | Allen, Mr. William Henry                          | male   | 35.0 | 0     |   |

|   | Parch | Ticket           | Fare    | Cabin | Embarked |
|---|-------|------------------|---------|-------|----------|
| 0 | 0     | A/5 21171        | 7.2500  | NaN   | S        |
| 1 | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2 | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3 | 0     | 113803           | 53.1000 | C123  | S        |
| 4 | 0     | 373450           | 8.0500  | NaN   | S        |

```
[76]: dataset["Cabin"].fillna(dataset["Cabin"].mode()[0],inplace=True)
```

```
[77]: dataset["Embarked"].fillna(dataset["Embarked"].mode()[0],inplace=True)
```

```
[78]: dataset.isnull().any()
```

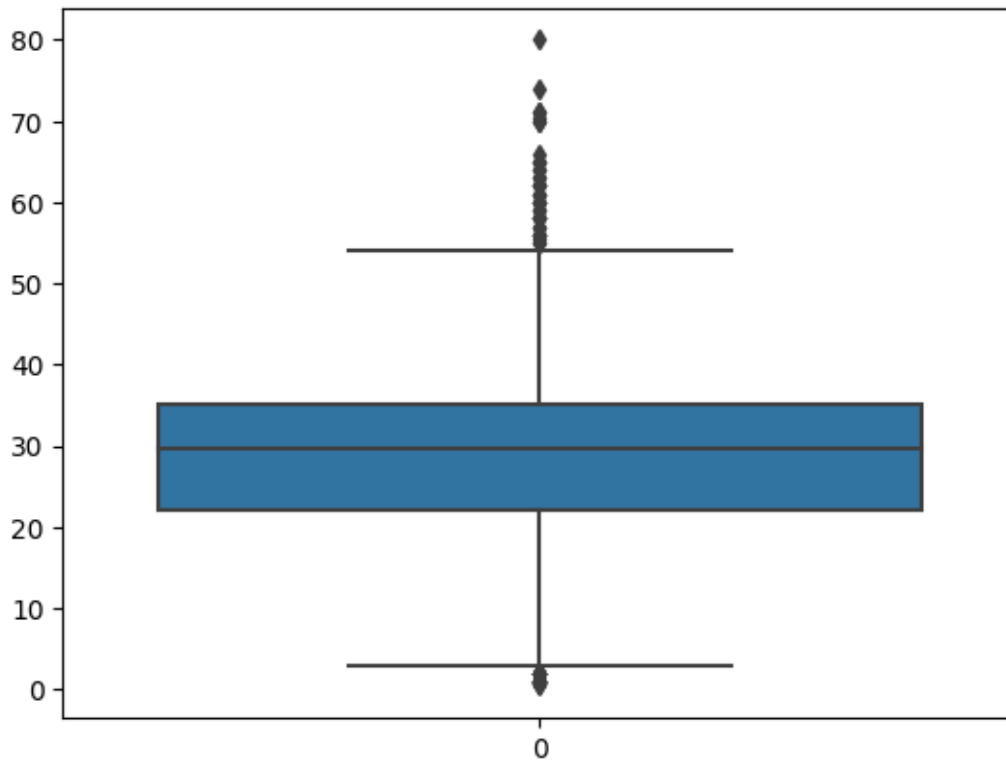
```
[78]: PassengerId    False
Survived         False
Pclass           False
Name             False
Sex              False
Age             False
SibSp            False
Parch           False
Ticket          False
Fare            False
Cabin           False
Embarked        False
dtype: bool
```



### Outlier Detection

```
[79]: sns.boxplot(dataset.Age)
```

```
[79]: <Axes: >
```



### Outlier Handling using IQR

```
[80]: q1 = dataset.Age.quantile(0.25)
      q3 = dataset.Age.quantile(0.75)
```

```
[81]: IQR = q3-q1
```

```
[82]: upper_limit = q3 + 1.5*IQR
```

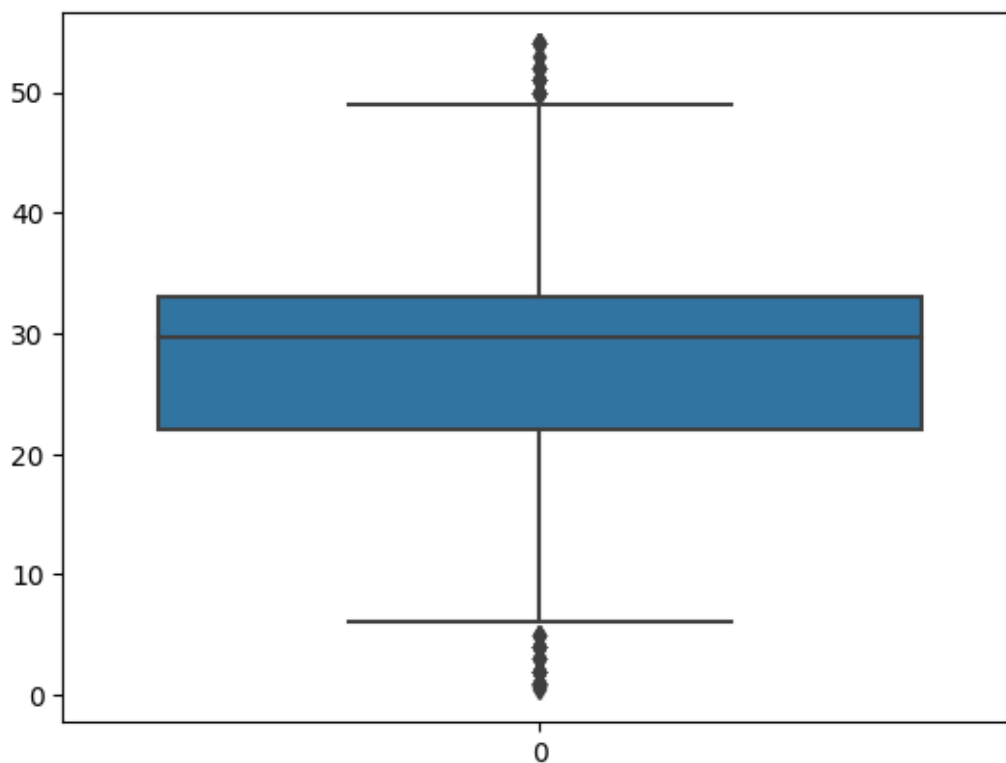
```
[83]: upper_limit
```

```
[83]: 54.5
```

```
[84]: dataset = dataset[dataset.Age<upper_limit]
```

```
[85]: sns.boxplot(dataset.Age)
```

[85]: <Axes: >



[86]: `dataset.shape`

[86]: (849, 12)

Separating Dependent and Independent Variables

[87]: `dataset`

[87]:

|     | PassengerId | Survived | Pclass | \ |
|-----|-------------|----------|--------|---|
| 0   | 1           | 0        | 3      |   |
| 1   | 2           | 1        | 1      |   |
| 2   | 3           | 1        | 3      |   |
| 3   | 4           | 1        | 1      |   |
| 4   | 5           | 0        | 3      |   |
| ..  | ...         | ...      | ...    |   |
| 886 | 887         | 0        | 2      |   |
| 887 | 888         | 1        | 1      |   |
| 888 | 889         | 0        | 3      |   |
| 889 | 890         | 1        | 1      |   |
| 890 | 891         | 0        | 3      |   |

|     |   | Name                         | Sex    | Age \     |
|-----|---|------------------------------|--------|-----------|
| 0   |   | Braund, Mr. Owen Harris      | male   | 22.000000 |
| 1   | Cumings, Mrs. John Bradley (Florence Briggs Th... |                              | female | 38.000000 |
| 2   |   | Heikkinen, Miss. Laina       | female | 26.000000 |
| 3   | Futrelle, Mrs. Jacques Heath (Lily May Peel)      |                              | female | 35.000000 |
| 4   |   | Allen, Mr. William Henry     | male   | 35.000000 |
| ..  |   | ...                          | ...    | ...       |
| 886 |   | Montvila, Rev. Juozas        | male   | 27.000000 |
| 887 |   | Graham, Miss. Margaret Edith | female | 19.000000 |
| 888 | Johnston, Miss. Catherine Helen "Carrie"          |                              | female | 29.699118 |
| 889 |   | Behr, Mr. Karl Howell        | male   | 26.000000 |
| 890 |   | Dooley, Mr. Patrick          | male   | 32.000000 |

|     | SibSp | Parch | Ticket           | Fare    | Cabin   | Embarked |
|-----|-------|-------|------------------|---------|---------|----------|
| 0   | 1     | 0     | A/5 21171        | 7.2500  | B96 B98 | S        |
| 1   | 1     | 0     | PC 17599         | 71.2833 | C85     | C        |
| 2   | 0     | 0     | STON/O2. 3101282 | 7.9250  | B96 B98 | S        |
| 3   | 1     | 0     | 113803           | 53.1000 | C123    | S        |
| 4   | 0     | 0     | 373450           | 8.0500  | B96 B98 | S        |
| ..  | ...   | ...   | ...              | ...     | ...     | ...      |
| 886 | 0     | 0     | 211536           | 13.0000 | B96 B98 | S        |
| 887 | 0     | 0     | 112053           | 30.0000 | B42     | S        |
| 888 | 1     | 2     | W./C. 6607       | 23.4500 | B96 B98 | S        |
| 889 | 0     | 0     | 111369           | 30.0000 | C148    | C        |
| 890 | 0     | 0     | 370376           | 7.7500  | B96 B98 | Q        |

[849 rows x 12 columns]

```
[88]: x = dataset.iloc[:,2:12]
      y = dataset.iloc[:,1:2]
```

```
[89]: x.head()
```

```
[89]: Pclass      Name      Sex      Age \
0      3      Braund, Mr. Owen Harris      male      22.0
1      1  Cumings, Mrs. John Bradley (Florence Briggs Th...      female      38.0
2      3      Heikkinen, Miss. Laina      female      26.0
3      1      Futrelle, Mrs. Jacques Heath (Lily May Peel)      female      35.0
4      3      Allen, Mr. William Henry      male      35.0

      SibSp  Parch      Ticket      Fare      Cabin Embarked
0      1      0      A/5 21171      7.2500  B96 B98      S
1      1      0      PC 17599      71.2833      C85      C
2      0      0  STON/O2. 3101282      7.9250  B96 B98      S
3      1      0      113803      53.1000      C123      S
4      0      0      373450      8.0500  B96 B98      S
```

```
[90]: y.head()
```

```
[90]:      Survived
0         0
1         1
2         1
3         1
4         0
```

```
[91]: dataset.shape
```

```
[91]: (849, 12)
```

```
[92]: x.shape
```

```
[92]: (849, 10)
```

```
[93]: y.shape
```

```
[93]: (849, 1)
```

Encoding

Label Encoding on Sex column

```
[94]: from sklearn.preprocessing import LabelEncoder
```

```
[95]: le = LabelEncoder()
```

```
[96]: x["Sex"]=le.fit_transform(x["Sex"])
      x["Sex"]
```

```
[96]: 0         1
      1         0
      2         0
      3         0
      4         1
      ..
      886        1
      887         0
      888         0
      889         1
      890         1
      Name: Sex, Length: 849, dtype: int64
```

```
[97]: x.head()
```

```
[97]:
```

|   | Pclass | Name  | Sex | Age  | \ |
|---|--------|---|-----|------|---|
| 0 | 3      | Braund, Mr. Owen Harris                           | 1   | 22.0 |   |
| 1 | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0   | 38.0 |   |
| 2 | 3      | Heikkinen, Miss. Laina                            | 0   | 26.0 |   |
| 3 | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | 0   | 35.0 |   |
| 4 | 3      | Allen, Mr. William Henry                          | 1   | 35.0 |   |

|   | SibSp | Parch | Ticket           | Fare    | Cabin   | Embarked |
|---|-------|-------|------------------|---------|---------|----------|
| 0 | 1     | 0     | A/5 21171        | 7.2500  | B96 B98 | S        |
| 1 | 1     | 0     | PC 17599         | 71.2833 | C85     | C        |
| 2 | 0     | 0     | STON/O2. 3101282 | 7.9250  | B96 B98 | S        |
| 3 | 1     | 0     | 113803           | 53.1000 | C123    | S        |
| 4 | 0     | 0     | 373450           | 8.0500  | B96 B98 | S        |

```
[98]: x["Sex"].value_counts()
```

```
[98]: 1    545
      0    304
      Name: Sex, dtype: int64
```

```
[99]: x["Sex"].nunique()
```

```
[99]: 2
```

One hot encoding on Pclass column

```
[100]: Pclass=pd.get_dummies(x["Pclass"],drop_first=True)
```

```
[101]: Pclass
```

```
[101]:
```

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

[849 rows x 2 columns]

```
[102]: x=pd.concat([x,Pclass],axis=1)
```

```
[103]: x.head()
```

```
[103]:
```

|   | Pclass | Name  | Sex | Age  | \ |
|---|--------|---|-----|------|---|
| 0 | 3      | Braund, Mr. Owen Harris                           | 1   | 22.0 |   |
| 1 | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0   | 38.0 |   |
| 2 | 3      | Heikkinen, Miss. Laina                            | 0   | 26.0 |   |
| 3 | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | 0   | 35.0 |   |
| 4 | 3      | Allen, Mr. William Henry                          | 1   | 35.0 |   |

|   | SibSp | Parch | Ticket           | Fare    | Cabin   | Embarked | 2 | 3 |
|---|-------|-------|------------------|---------|---------|----------|---|---|
| 0 | 1     | 0     | A/5 21171        | 7.2500  | B96 B98 | S        | 0 | 1 |
| 1 | 1     | 0     | PC 17599         | 71.2833 | C85     | C        | 0 | 0 |
| 2 | 0     | 0     | STON/O2. 3101282 | 7.9250  | B96 B98 | S        | 0 | 1 |
| 3 | 1     | 0     | 113803           | 53.1000 | C123    | S        | 0 | 0 |
| 4 | 0     | 0     | 373450           | 8.0500  | B96 B98 | S        | 0 | 1 |

```
[104]: x.shape
```

```
[104]: (849, 12)
```

```
[105]: x.drop(["Pclass"],axis=1,inplace=True)
```

```
[106]: x.head(10)
```

```
[106]:
```

|   | Name  | Sex | Age       | SibSp | \ |
|---|---|-----|-----------|-------|---|
| 0 | Braund, Mr. Owen Harris                           | 1   | 22.000000 | 1     |   |
| 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0   | 38.000000 | 1     |   |
| 2 | Heikkinen, Miss. Laina                            | 0   | 26.000000 | 0     |   |
| 3 | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | 0   | 35.000000 | 1     |   |
| 4 | Allen, Mr. William Henry                          | 1   | 35.000000 | 0     |   |
| 5 | Moran, Mr. James                                  | 1   | 29.699118 | 0     |   |
| 6 | McCarthy, Mr. Timothy J                           | 1   | 54.000000 | 0     |   |
| 7 | Palsson, Master. Gosta Leonard                    | 1   | 2.000000  | 3     |   |
| 8 | Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg) | 0   | 27.000000 | 0     |   |
| 9 | Nasser, Mrs. Nicholas (Adele Achem)               | 0   | 14.000000 | 1     |   |

|   | Parch | Ticket           | Fare    | Cabin   | Embarked | 2 | 3 |
|---|-------|------------------|---------|---------|----------|---|---|
| 0 | 0     | A/5 21171        | 7.2500  | B96 B98 | S        | 0 | 1 |
| 1 | 0     | PC 17599         | 71.2833 | C85     | C        | 0 | 0 |
| 2 | 0     | STON/O2. 3101282 | 7.9250  | B96 B98 | S        | 0 | 1 |
| 3 | 0     | 113803           | 53.1000 | C123    | S        | 0 | 0 |
| 4 | 0     | 373450           | 8.0500  | B96 B98 | S        | 0 | 1 |
| 5 | 0     | 330877           | 8.4583  | B96 B98 | Q        | 0 | 1 |
| 6 | 0     | 17463            | 51.8625 | E46     | S        | 0 | 0 |
| 7 | 1     | 349909           | 21.0750 | B96 B98 | S        | 0 | 1 |
| 8 | 2     | 347742           | 11.1333 | B96 B98 | S        | 0 | 1 |
| 9 | 0     | 237736           | 30.0708 | B96 B98 | C        | 1 | 0 |

Splitting into Training and Testing set

```
[107]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

```
[108]: x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
[108]: ((594, 11), (255, 11), (594, 1), (255, 1))
```

```
[109]: a=[1,2,3,4,5,6]
b=[1,0,1,5,6,3]

for i in range(5):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.
↪3,random_state=100)
    print("with random state",a_train)
```

```
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
```

```
[110]: a=[1,2,3,4,5,6]
b=[1,0,1,5,6,3]

for i in range(5):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.3)
    print("without random state",a_train)
```

```
without random state [5, 1, 3, 6]
without random state [5, 4, 6, 3]
without random state [4, 3, 6, 1]
without random state [4, 3, 5, 1]
without random state [5, 3, 2, 1]
```

Feature Scaling

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

```
[112]: x.head()
```

```
[112]:
```

|   | Name  | Sex | Age  | SibSp | Parch | \ |
|---|---|-----|------|-------|-------|---|
| 0 | Braund, Mr. Owen Harris                           | 1   | 22.0 | 1     | 0     |   |
| 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0   | 38.0 | 1     | 0     |   |
| 2 | Heikkinen, Miss. Laina                            | 0   | 26.0 | 0     | 0     |   |
| 3 | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | 0   | 35.0 | 1     | 0     |   |

|   |  |                          |   |      |   |   |
|---|--|--------------------------|---|------|---|---|
| 4 |  | Allen, Mr. William Henry | 1 | 35.0 | 0 | 0 |
|---|--|--------------------------|---|------|---|---|

|   | Ticket           | Fare    | Cabin   | Embarked | 2 | 3 |
|---|------------------|---------|---------|----------|---|---|
| 0 | A/5 21171        | 7.2500  | B96 B98 | S        | 0 | 1 |
| 1 | PC 17599         | 71.2833 | C85     | C        | 0 | 0 |
| 2 | STON/O2. 3101282 | 7.9250  | B96 B98 | S        | 0 | 1 |
| 3 | 113803           | 53.1000 | C123    | S        | 0 | 0 |
| 4 | 373450           | 8.0500  | B96 B98 | S        | 0 | 1 |

```
[113]: scaler = StandardScaler()

# Define the columns you want to scale (exclude 'Sex', '2', and '3')
numeric_columns = ['Age', 'SibSp', 'Parch', 'Fare']

# Fit and transform the scaler on the selected columns for training data
x_train[numeric_columns] = scaler.fit_transform(x_train[numeric_columns])

# Transform the test data using the same scaler
x_test[numeric_columns] = scaler.transform(x_test[numeric_columns])
```

```
[114]: x_train
```

```
[114]:
```

|     | Name   | Sex | Age       | SibSp    | \ |
|-----|--|-----|-----------|----------|---|
| 272 | Mellinger, Mrs. (Elizabeth Anne Maidment)    | 0   | 1.134388  | -0.48352 |   |
| 19  | Masselmani, Mrs. Fatima                      | 0   | 0.124901  | -0.48352 |   |
| 560 | Morrow, Mr. Thomas Rowan                     | 1   | 0.124901  | -0.48352 |   |
| 704 | Hansen, Mr. Henrik Juul                      | 1   | -0.205535 | 0.36371  |   |
| 599 | Duff Gordon, Sir. Cosmo Edmund ("Mr Morgan") | 1   | 1.849014  | 0.36371  |   |
| ..  | ...  | ... | ...       | ...      |   |
| 876 | Gustafsson, Mr. Alfred Ossian                | 1   | -0.741504 | -0.48352 |   |
| 203 | Youseff, Mr. Gerious                         | 1   | 1.536365  | -0.48352 |   |
| 663 | Coleff, Mr. Peju                             | 1   | 0.687747  | -0.48352 |   |
| 588 | Gilinski, Mr. Eliezer                        | 1   | -0.562848 | -0.48352 |   |
| 721 | Jensen, Mr. Svend Lauritz                    | 1   | -1.009489 | 0.36371  |   |

|     | Parch     | Ticket   | Fare      | Cabin   | Embarked | 2   | 3   |
|-----|-----------|----------|-----------|---------|----------|-----|-----|
| 272 | 0.679535  | 250644   | -0.245822 | B96 B98 | S        | 1   | 0   |
| 19  | -0.487065 | 2649     | -0.490937 | B96 B98 | C        | 0   | 1   |
| 560 | -0.487065 | 372622   | -0.480453 | B96 B98 | Q        | 0   | 1   |
| 704 | -0.487065 | 350025   | -0.478373 | B96 B98 | S        | 0   | 1   |
| 599 | -0.487065 | PC 17485 | 0.501588  | A20     | C        | 0   | 0   |
| ..  | ...       | ...      | ...       | ...     | ...      | ... | ... |
| 876 | -0.487065 | 7534     | -0.438603 | B96 B98 | S        | 0   | 1   |
| 203 | -0.487065 | 2628     | -0.490937 | B96 B98 | C        | 0   | 1   |
| 663 | -0.487065 | 349210   | -0.485529 | B96 B98 | S        | 0   | 1   |
| 588 | -0.487065 | 14973    | -0.474463 | B96 B98 | S        | 0   | 1   |
| 721 | -0.487065 | 350048   | -0.494347 | B96 B98 | S        | 0   | 1   |



[594 rows x 11 columns]