

```
In [46]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score
```

```
In [2]: df_titanic = pd.read_csv(r"C:\Users\neetu27\Downloads\archive (1)\tested.csv")
```

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

```
Out[4]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

```
In [5]: df_titanic.shape
```

```
Out[5]: (418, 12)
```

```
In [6]: df_titanic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      418 non-null    int64
1   Survived         418 non-null    int64
2   Pclass           418 non-null    int64
3   Name             418 non-null    object
4   Sex              418 non-null    object
5   Age              332 non-null    float64
6   SibSp            418 non-null    int64
7   Parch            418 non-null    int64
8   Ticket           418 non-null    object
9   Fare             417 non-null    float64
10  Cabin            91 non-null     object
11  Embarked         418 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

```
In [7]: df_titanic.isnull().sum()
```

```
Out[7]: PassengerId      0
Survived      0
Pclass        0
Name          0
Sex           0
Age           86
SibSp         0
Parch         0
Ticket        0
Fare          1
Cabin        327
Embarked      0
dtype: int64
```

Handling of missing values

```
In [9]: df_titanic= df_titanic.drop(columns='Cabin',axis=1)
```

```
In [10]: df_titanic['Age'].fillna(df_titanic['Age'].mean(),inplace=True)
```

```
In [11]: print(df_titanic['Embarked'].mode())
```

```
0    S
Name: Embarked, dtype: object
```

```
In [12]: print(df_titanic['Embarked'].mode()[0])
```

```
S
```

```
In [15]: df_titanic['Embarked'].fillna(df_titanic['Embarked'].mode()[0],inplace=True)
```

```
In [16]: df_titanic.isnull().sum()
```

```
Out[16]: PassengerId    0
Survived      0
Pclass        0
Name          0
Sex           0
Age           0
SibSp         0
Parch         0
Ticket        0
Fare          1
Embarked      0
dtype: int64
```

DATA ANALYSIS

```
In [17]: df_titanic.describe()
```

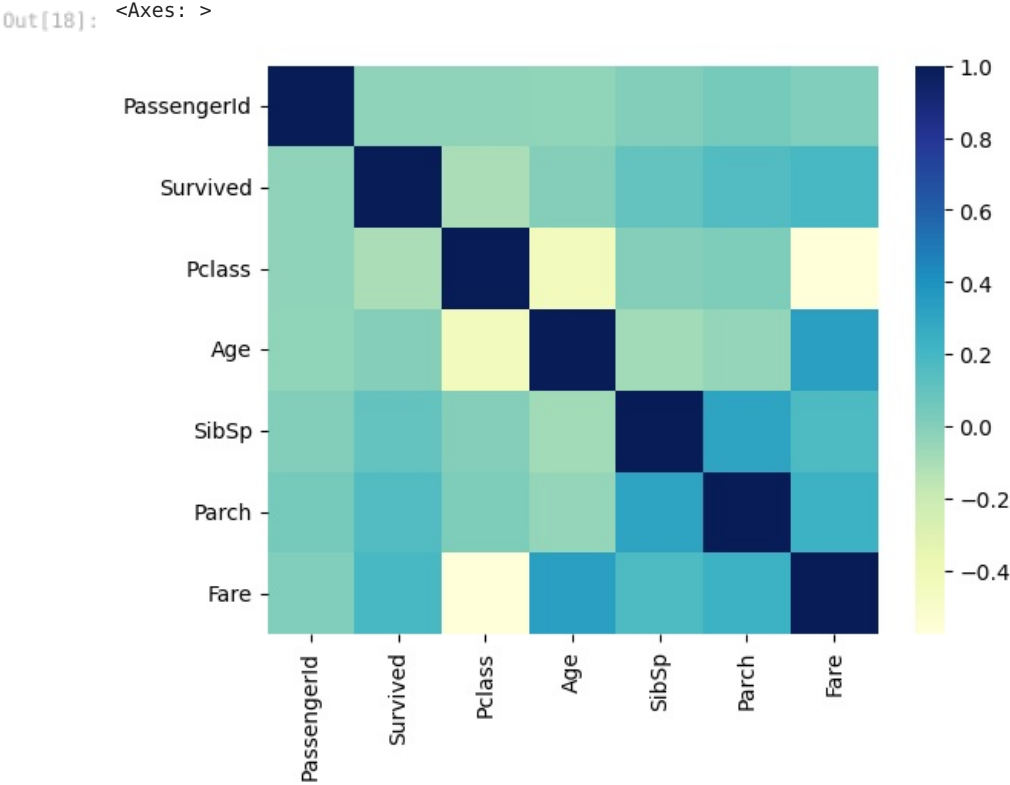
Out[17]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	418.000000	418.000000	418.000000	417.000000
mean	1100.500000	0.363636	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.481622	0.841838	12.634534	0.896760	0.981429	55.907576
min	892.000000	0.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	0.000000	1.000000	23.000000	0.000000	0.000000	7.895800
50%	1100.500000	0.000000	3.000000	30.272590	0.000000	0.000000	14.454200
75%	1204.750000	1.000000	3.000000	35.750000	1.000000	0.000000	31.500000
max	1309.000000	1.000000	3.000000	76.000000	8.000000	9.000000	512.329200

```
In [18]: sns.heatmap(df_titanic.corr(),cmap = "YlGnBu")
```

C:\Users\neetu27\AppData\Local\Temp\ipykernel_13784\1368390462.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.

```
sns.heatmap(df_titanic.corr(),cmap = "YlGnBu")
```

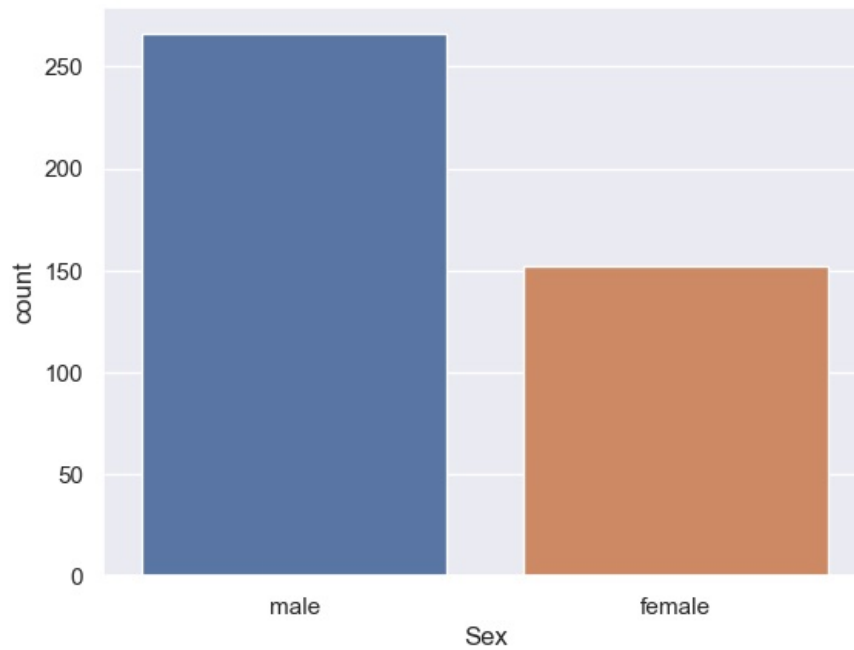


Data Visulaization

```
In [19]: sns.set()
```

```
In [21]: sns.countplot(x = 'Sex', data = df_titanic)
```

```
Out[21]: <Axes: xlabel='Sex', ylabel='count'>
```



```
In [22]: df_titanic['Sex'].value_counts()
```

```
Out[22]: male      266
female    152
Name: Sex, dtype: int64
```

```
In [23]: df_titanic['Embarked'].value_counts()
```

```
Out[23]: S      270
C      102
Q       46
Name: Embarked, dtype: int64
```

```
In [24]: df_titanic.replace({'Sex':{'male':0,'female':1}, 'Embarked':{'S':0,'C':1,'Q':2}},inplace=True)
```

```
In [25]: df_titanic.head()
```

```
Out[25]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	892	0	3	Kelly, Mr. James	0	34.5	0	0	330911	7.8292	2
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	1	47.0	1	0	363272	7.0000	0
2	894	0	2	Myles, Mr. Thomas Francis	0	62.0	0	0	240276	9.6875	2
3	895	0	3	Wirz, Mr. Albert	0	27.0	0	0	315154	8.6625	0
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	1	22.0	1	1	3101298	12.2875	0

```
In [31]: df_titanic = pd.get_dummies(df_titanic,drop_first=True)
df_titanic = df_titanic.drop(columns=['Embarked'])
df_titanic
```

Out[31]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Name_Abelseth, Miss. Karen Marie	Name_Abelseth, Mr. Olaus Jorgensen	...	Ticket_SOTON/OQ 392083	Ticket_2.
0	892	0	3	0	34.50000	0	0	7.8292	0	0	...	0	
1	893	1	3	1	47.00000	1	0	7.0000	0	0	...	0	
2	894	0	2	0	62.00000	0	0	9.6875	0	0	...	0	
3	895	0	3	0	27.00000	0	0	8.6625	0	0	...	0	
4	896	1	3	1	22.00000	1	1	12.2875	0	0	...	0	
...
413	1305	0	3	0	30.27259	0	0	8.0500	0	0	...	0	
414	1306	1	1	1	39.00000	0	0	108.9000	0	0	...	0	
415	1307	0	3	0	38.50000	0	0	7.2500	0	0	...	0	
416	1308	0	3	0	30.27259	0	0	8.0500	0	0	...	0	
417	1309	0	3	0	30.27259	1	1	22.3583	0	0	...	0	

418 rows × 787 columns

In [67]: x= df_titanic[['Pclass','Sex','Age','SibSp','Parch','Fare']]
Y= df_titanic['Survived']

In [68]: x_train,x_test,Y_train, Y_test = train_test_split(x,Y,test_size= 0.25,random_state=2)

In [69]: lr = LogisticRegression(max_iter=150)

In [71]: Y_test.shape

Out[71]: (105,)

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js