

```
In [1]: #Importing required library
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math
%matplotlib inline

In [2]: import warnings
warnings.filterwarnings("ignore")

In [3]: ## 1. Collecting the data

In [4]: titanic_data=pd.read_csv("https://gist.githubusercontent.com/fyyying/4aa5b471868321d7b47f0881898162b7/raw/6987fb3a38bfbb6fccf3a8b1edf90e39714d14f/titanic_dataset.csv")
titanic_data.head()

Out[4]:
```


	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	Cummings, 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

```


In [5]: titanic_data.shape
Out[5]: (891, 12)

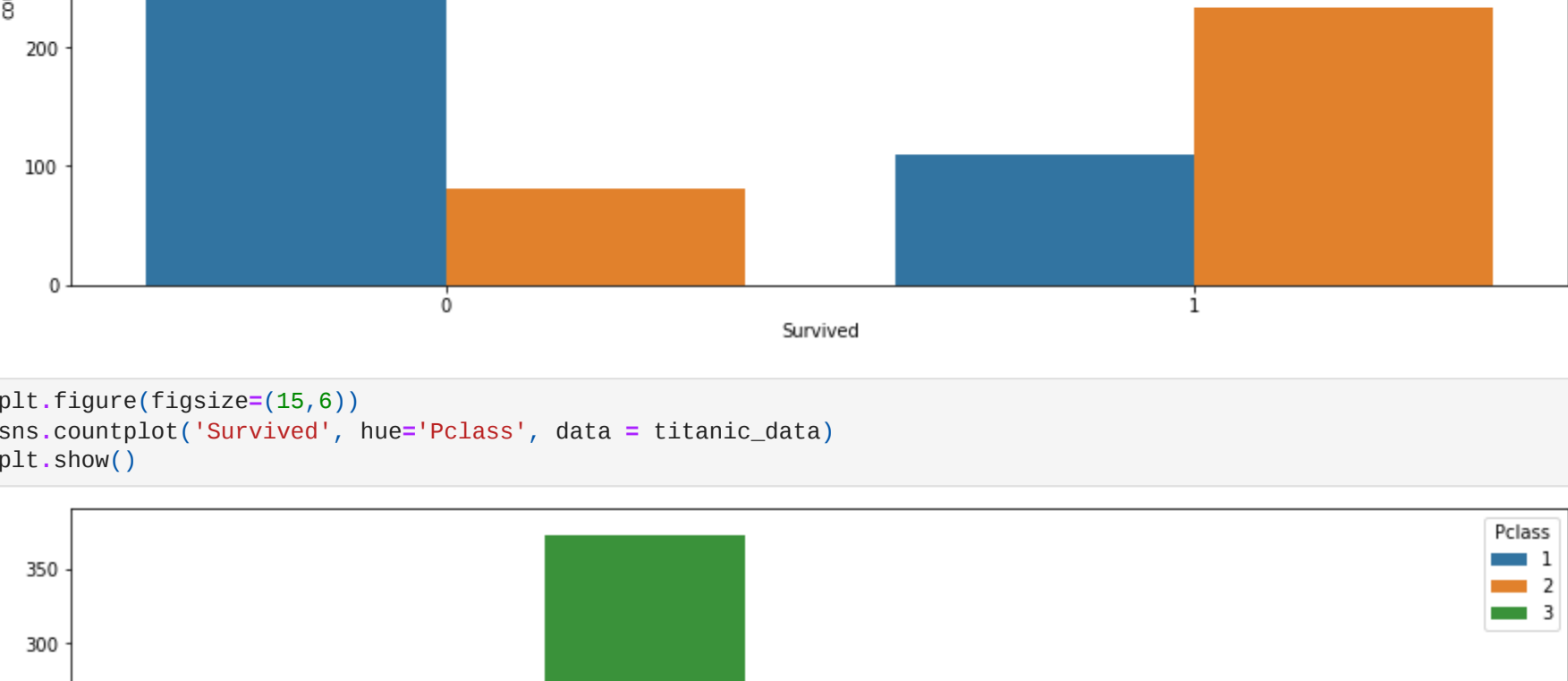
In [6]: ## 2. Analyzing the data

In [7]: plt.figure(figsize=(15,6))
sns.countplot('Survived', data = titanic_data)
plt.show()
```



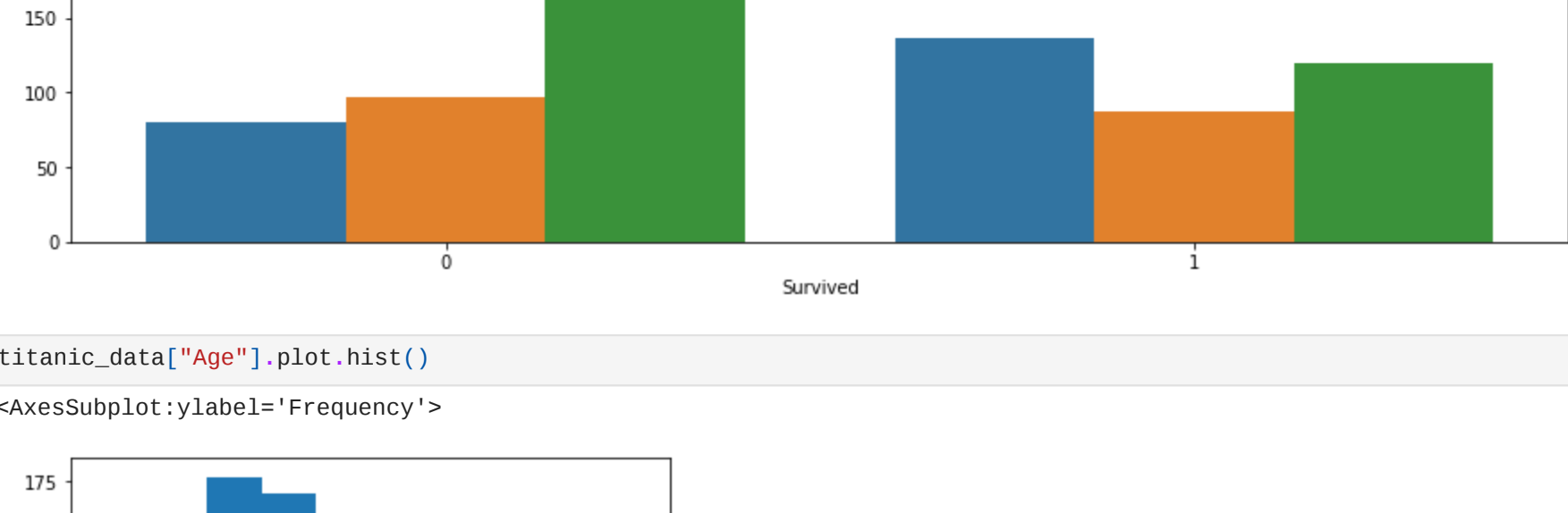
```


In [8]: plt.figure(figsize=(15,6))
sns.countplot('Survived', hue='Sex', data = titanic_data)
plt.show()
```



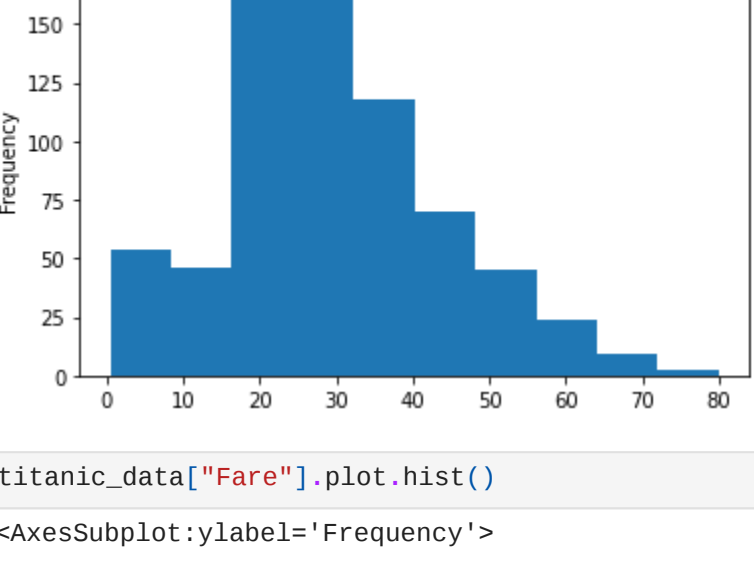
```


In [9]: plt.figure(figsize=(15,6))
sns.countplot('Survived', hue='Pclass', data = titanic_data)
plt.show()
```



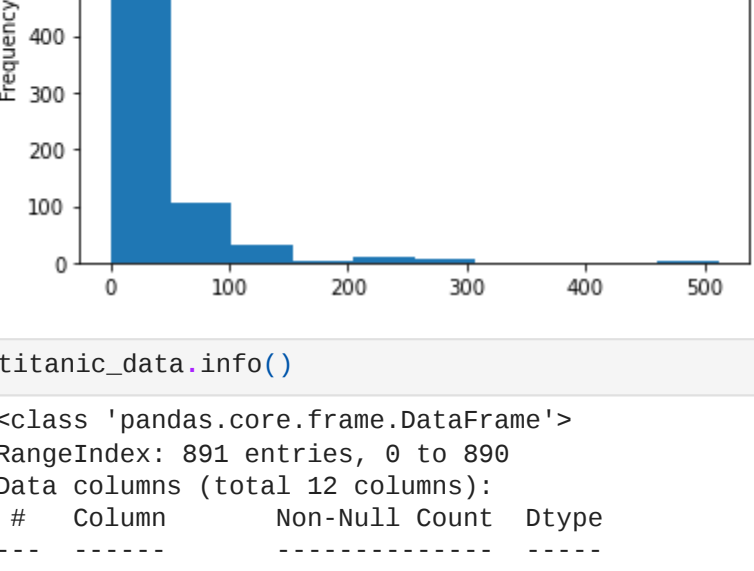
```


In [10]: titanic_data["Age"].plot.hist()
Out[10]: <AxesSubplot:ylabel='Frequency'>
```



```

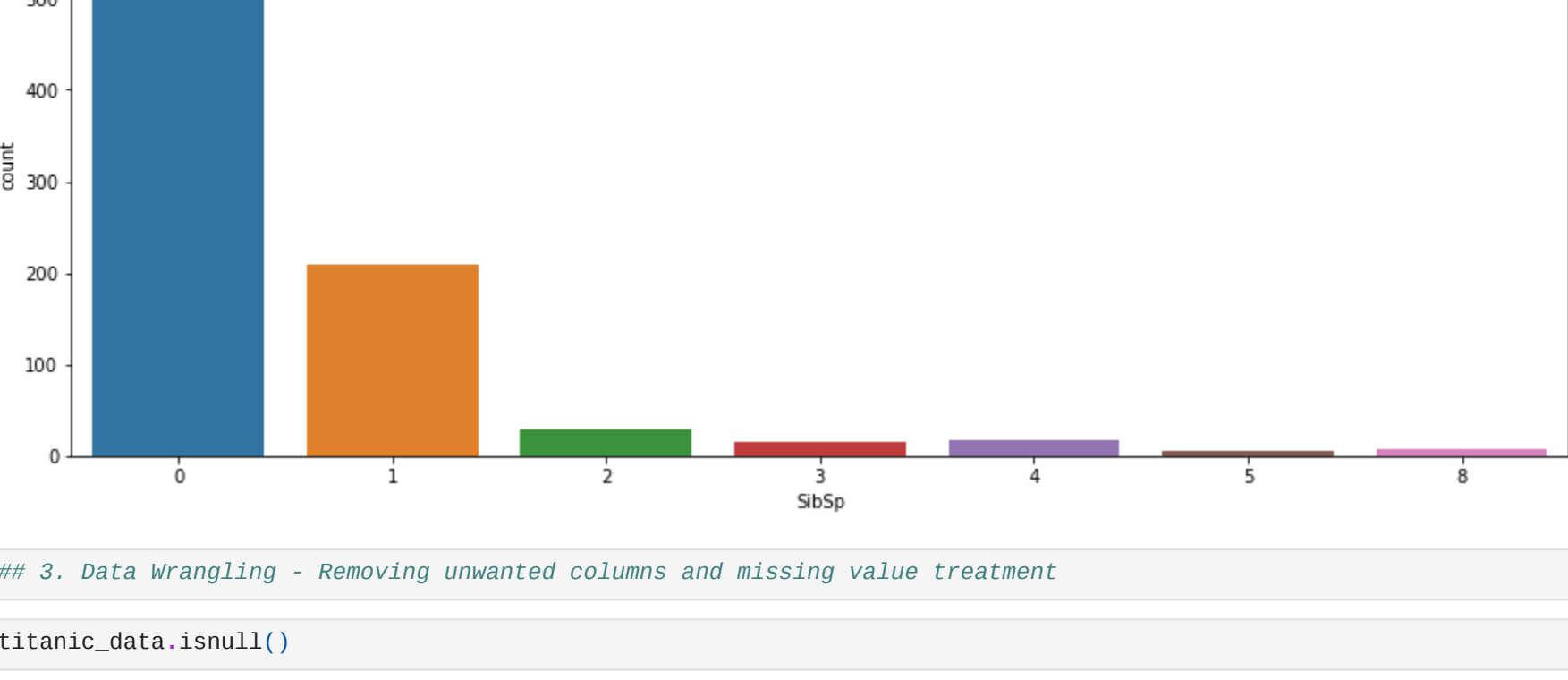

In [11]: titanic_data["Fare"].plot.hist()
Out[11]: <AxesSubplot:ylabel='Frequency'>
```



```


In [12]: titanic_data.info()
Out[12]:
<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 [13]: plt.figure(figsize=(15,6))
sns.countplot('SibSp', data = titanic_data)
plt.show()
```



```


In [14]: ## 3. Data Wrangling - Removing unwanted columns and missing value treatment

In [15]: titanic_data.isnull()

Out[15]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	True	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...
886	False	False	False	False	False	False	False	False	False	True	False	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	True	False	False	False	False	True	False	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	True	False	False


891 rows × 12 columns

```


In [16]: titanic_data.isnull().sum()
Out[16]: 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

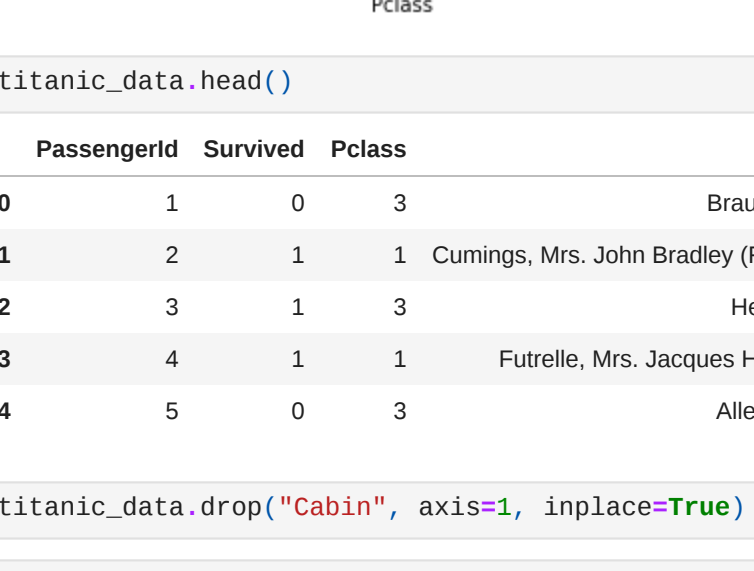
In [17]: import seaborn as sns
sns.heatmap(titanic_data.isnull(), yticklabels = False)

Out[17]: <AxesSubplot:>
```



```


In [18]: sns.boxplot(x="Pclass", y="Age", data=titanic_data)
Out[18]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



```


In [19]: titanic_data.head()
Out[19]:
```

	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	Cummings, 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

```


In [20]: titanic_data.drop("Cabin", axis=1, inplace=True)

In [21]: titanic_data

Out[21]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Q

891 rows × 11 columns

```


In [22]: titanic_data.dropna(inplace=True)

In [23]: titanic_data

Out[23]:
```

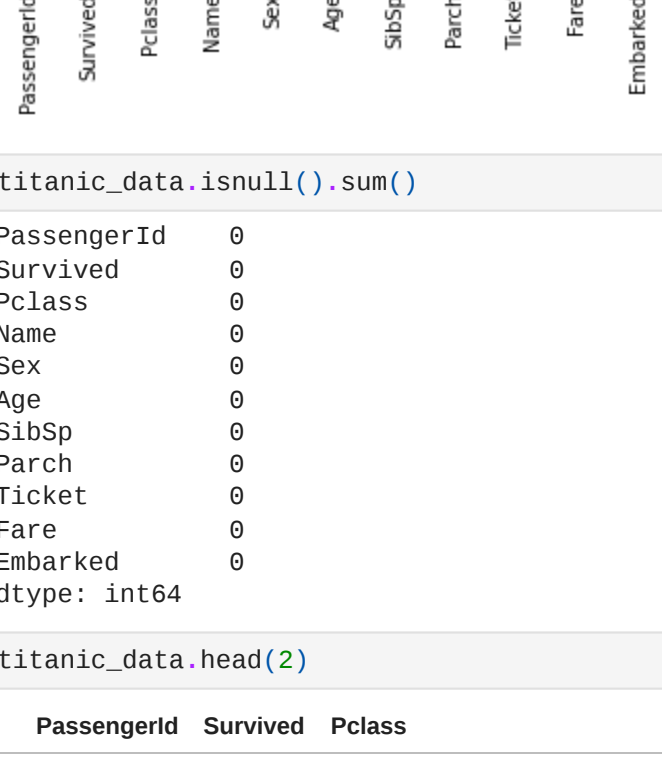
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S
...
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	Q
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Q

712 rows × 11 columns

```


In [24]: sns.heatmap(titanic_data.isnull(), yticklabels = False, cbar=False)

Out[24]: <AxesSubplot:>
```



```


In [25]: titanic_data.isnull().sum()
Out[25]: PassengerId      0
Survived            0
Pclass             0
Name               0
Sex               0
Age              0
SibSp            0
Parch            0
Ticket            0
Fare              0
Embarked          0
dtype: int64

In [26]: titanic_data.head(2)
Out[26]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C

```


In [27]: pd.get_dummies(titanic_data["Sex"])
Out[27]:
```

	female	male
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1
...
885	1	0
886	0	1
887	1	0
889	0	1
890	0	1

712 rows × 2 columns

```


In [28]: sex=pd.get_dummies(titanic_data["Sex"], drop_first=True)
sex.head(5)

Out[28]:
```

	male
0	1
1	0
2	0
3	0
4	1

```


In [29]: embark=pd.get_dummies(titanic_data["Embarked"], drop_first=True)
embark.head(5)

Out[29]:
```

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

```


In [30]: titanic_data["Embarked"].value_counts()
Out[30]: S    554
C    138
Q     28
Name: Embarked, dtype: int64

In [31]: pcl=pd.get_dummies(titanic_data["Pclass"], drop_first=True)
pcl.head(5)

Out[31]:
```

	2	3
0	0	1
1	0	0
2	0	1
3	0	0
4	0	1

```


In [32]: titanic_data = pd.concat([titanic_data,sex, embark,pcl],axis=1)

In [33]: titanic_data.head(4)
Out[33]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	male	Q	S	2	3
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S	1	0	1	0	1
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C	0	0	0	0	0
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S	0	0	1	0	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S	0	0	1	0	1

```


In [34]: titanic_data.drop(["PassengerId","Name","Sex","Ticket","Embarked","Pclass"], axis=1, inplace=True)

In [35]: titanic_data

Out[35]:
```

	Survived	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	0	22.0	1	0	7.2500	1	0	1	0	1
1	1	38.0	1	0	71.2833	0	0	0	0	0
2	1	26.0	0	0	7.9250	0	0	1	0	1
3	1	35.0	1	0	53.1000	0	0	1	0	0
4	0	35.0	0	0	8.0500	1	0	1	0	1
...
885	0	39.0	0	5	29.1250	0	1	0	0	1
886	0	27.0	0	0	13.0000	1	0	1	1	0
887	1	19.0	0	0	30.0000	0	0	1	0	0
889	0	26.0	0	0	30.0000	1	0	0	0	0
890	0	32.0	0	0	7.7500	1	1	0	0	1

712 rows × 10 columns

```


In [36]: ## 4. Training and testing the data

In [37]: X=titanic_data[["Survived"], axis=1]
y=titanic_data["Survived"]

In [38]: X

Out[38]:
```

	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	22.0	1	0	7.2500	1	0	1	0	1
1	38.0	1	0	71.2833	0	0	0	0	0
2	26.0	0	0	7.9250	0	0	1	0	1
3	35.0	1	0	53.1000	0	0	1	0	0
4	35.0	0	0	8.0500	1	0	1	0	1
...
885	39.0	0	5	29.1250	0	1	0	0	1
886	27.0	0	0	13.0000	1	0	1	1	0
887	19.0	0	0	30.0000	0	0	1	0	0
889	26.0	0	0	30.0000	1	0	0	0	0
890	32.0	0	0	7.7500	1	1	0	0	1

712 rows × 9 columns

```


In [39]: 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=1)
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
logmodel=LogisticRegression()

In [40]: logmodel.fit(X_train,y_train)

Out[40]: LogisticRegression()

In [41]: predictions=logmodel.predict(X_test)

In [42]: from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,predictions)

Out[42]: array([[102, 24],
       [ 25, 433]], dtype=int64)

In [43]: from sklearn.metrics import accuracy_score
accuracy_score(y_test, predictions)

Out[43]: 0.7710280373831776

In [ ]:
```