Assignment 3- Kanishka Verma(21BCE1412, VIT Chennai)

Importing the Libraries

```
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.preprocessing import StandardScaler, LabelEncoder
```

Importing the dataset

```
path="/content/drive/MyDrive/Data/Titanic-Dataset.csv"
data = pd.read_csv(path)
data.describe
```

```
<bound method NDFrame.describe of      PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
                                 3
...
                                 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]>
```

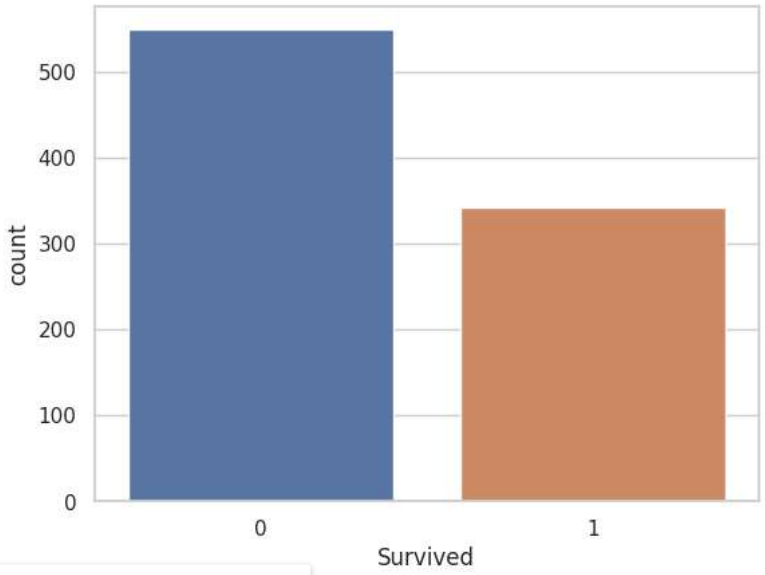
Checking for Null Values

```
null_values = data.isnull().sum()
print(null_values)
```

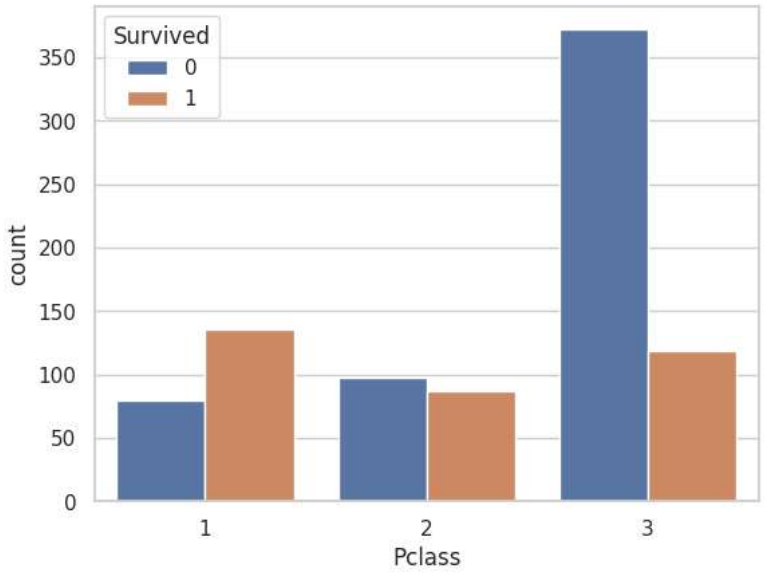
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

Data Visualization

```
sns.set(style="whitegrid")
sns.countplot(x='Survived', data=data)
plt.show()
```



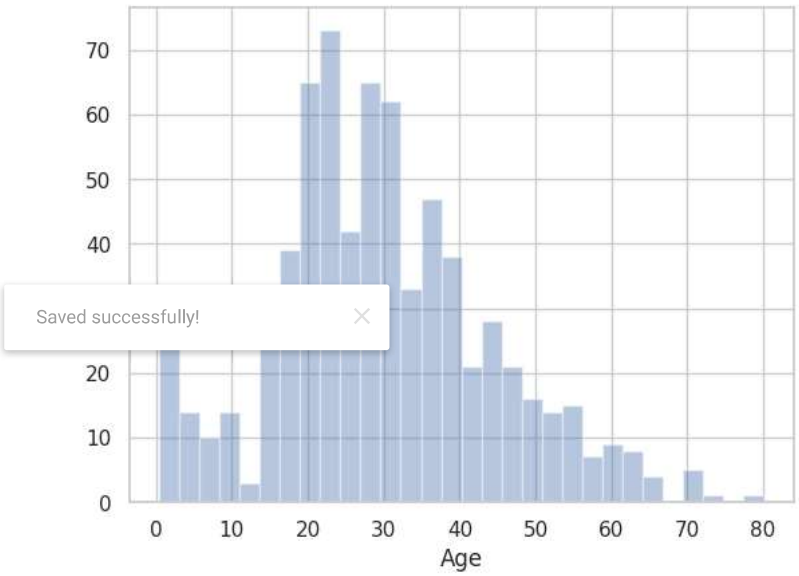

```
sns.countplot(x='Pclass', hue='Survived', data=data)
plt.show()
```




```
sns.countplot(x='Sex', hue='Survived', data=data)
plt.show()
```

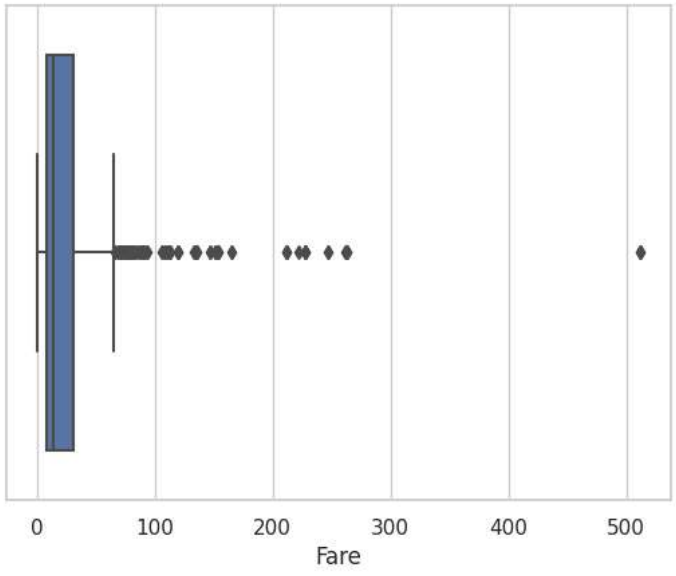
```
sns.distplot(data['Age'].dropna(), kde=False, bins=30)
plt.show()
```

```
<ipython-input-7-87e869ee960a>: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(data['Age'].dropna(), kde=False, bins=30)
```



Outlier Detection

```
sns.boxplot(x='Fare', data=data)
plt.show()
```



Splitting Dependent and Independent variables

```
X = data.drop('Survived', axis=1)  # Independent variables
y = data['Survived']  # Dependent variable
```

Perform Encoding

```
encoder = LabelEncoder()
X['Sex'] = encoder.fit_transform(X['Sex'])
```

```
X = pd.get_dummies(X, columns=['Embarked'], drop_first=True)
```

Feature Scaling

```
scaler = StandardScaler()
X[['Age', 'Fare']] = scaler.fit_transform(X[['Age', 'Fare']])
```

Splitting Data into Train and Test

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of y_train:", y_train.shape)
print("Shape of y_test:", y_test.shape)
```

```
Shape of X_train: (712, 12)
Shape of X_test: (179, 12)
Shape of y_train: (712,)
Shape of y_test: (179,)
```

```
print("\nX_train head:\n", X_train.head())
                                    d())
                                   ead())
                                    d())
```



```
X_train head:
      PassengerId  Pclass                                Name  Sex       Age  \
331          332       1               Partner, Mr. Austen    1  1.088491
733          734       2          Berriman, Mr. William John    1 -0.461489
382          383       3                  Tikkanen, Mr. Juho    1  0.158503
704          705       3              Hansen, Mr. Henrik Juul    1 -0.254825
813          814       3  Andersson, Miss. Ebba Iris Alfrida    0 -1.632584

     SibSp  Parch             Ticket      Fare Cabin  Embarked_Q  Embarked_S
331      0      0             113043 -0.074583  C124           0           1
733      0      0              28425 -0.386671   NaN           0           1
382      0      0  STON/O 2. 3101293 -0.488854   NaN           0           1
704      1      0             350025 -0.490280   NaN           0           1
813      4      2             347082 -0.018709   NaN           0           1

X_test head:
      PassengerId  Pclass                                          Name  \
709          710       3  Moubarek, Master. Halim Gonios ("William George")
439          440       2             Kvillner, Mr. Johan Henrik Johannesson
840          841       3                         Alhomaki, Mr. Ilmari Rudolf
720          721       2                Harper, Miss. Annie Jessie "Nina"
39            40       3                      Nicola-Yarred, Miss. Jamila

     Sex       Age  SibSp  Parch            Ticket      Fare Cabin  \
709    1       NaN      1      1              2661 -0.341452   NaN
439    1  0.089615      0      0        C.A. 18723 -0.437007   NaN
840    1 -0.668153      0      0  SOTON/O2 3101287 -0.488854   NaN
720    0 -1.632584      0      1            248727  0.016023   NaN
39     0 -1.081480      1      0              2651 -0.422074   NaN

     Embarked_Q  Embarked_S
709           0           0
439           0           1
840           0           1
720           0           1
39            0           0

y_train head:
 331    0
733    0
382    0
704    0
813    0
Name: Survived, dtype: int64

y_test head:
 709    1
439    0
840    0
720    1
39     1
Name: Survived, dtype: int64
```