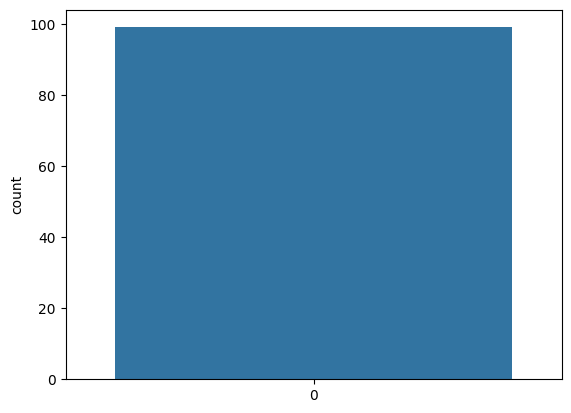
**5. Exploratory Data Analysis (EDA) Process :**

|  |
| --- |
| #Load the required libraries import pandas as pd import numpy as np import seaborn as sns  #Load the data df = pd.read\_csv('/content/titanic - dataset  - Sheet1.csv')   #View the data df.head() df.info() df.describe() |

|  |
| --- |
| <class '**pandas**.**core**.**frame**.**DataFrame**'> **RangeIndex**: 99 entries, 0 to 98 Data columns (total 12 columns):  #   Column       Non-Null Count  Dtype   ---  ------       --------------  -----    0   PassengerId  99 non-null     int64    1   Survived     99 non-null     int64    2   Pclass       99 non-null     int64    3   Name         99 non-null     object   4   Sex          99 non-null     object   5   Age          77 non-null     float64  6   SibSp        99 non-null     int64    7   Parch        99 non-null     int64    8   Ticket       99 non-null     object   9   Fare         99 non-null     float64  10  Cabin        20 non-null     object   11  Embarked     98 non-null     object  dtypes: float64(2), int64(5), object(5) memory usage: 9.4+ KB PassengerId Survived Pclass Age SibSp Parch Fare count 99.000000 99.000000 99.000000 77.000000 99.000000 99.000000 99.000000 mean 50.000000 0.414141 2.404040 27.380909 0.727273 0.444444 29.553157 std 28.722813 0.495080 0.819646 15.360556 1.185096 0.971242 41.179872 min 1.000000 0.000000 1.000000 0.830000 0.000000 0.000000 7.225000 25% 25.500000 0.000000 2.000000 18.000000 0.000000 0.000000 8.050000 50% 50.000000 0.000000 3.000000 26.000000 0.000000 0.000000 15.500000 75% 74.500000 1.000000 3.000000 35.000000 1.000000 0.000000 32.881250 max 99.000000 1.000000 3.000000 71.000000 5.000000 5.000000 263.000000 |

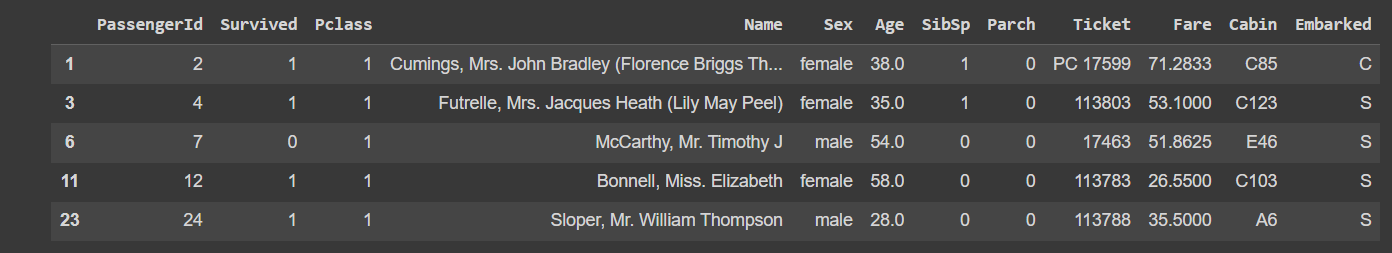
|  |
| --- |
| #plotting sns.countplot(df['Pclass']) |



|  |
| --- |
| #Replace null values df.replace(np.nan,'0',inplace = True)  #Check the changes now df.isnull().sum() |

|  |
| --- |
| PassengerId    0 Survived       0 Pclass         0 Name           0 Sex            0 Age            0 SibSp          0 Parch          0 Ticket         0 Fare           0 Cabin          0 Embarked       0 dtype: int64 |

|  |
| --- |
| df[df['Pclass']==1].head() |



|  |
| --- |
| df.dtypes |

|  |
| --- |
| PassengerId      int64 Survived         int64 Pclass           int64 Name            object Sex             object Age             object SibSp            int64 Parch            int64 Ticket          object Fare           float64 Cabin           object Embarked        object dtype: object |