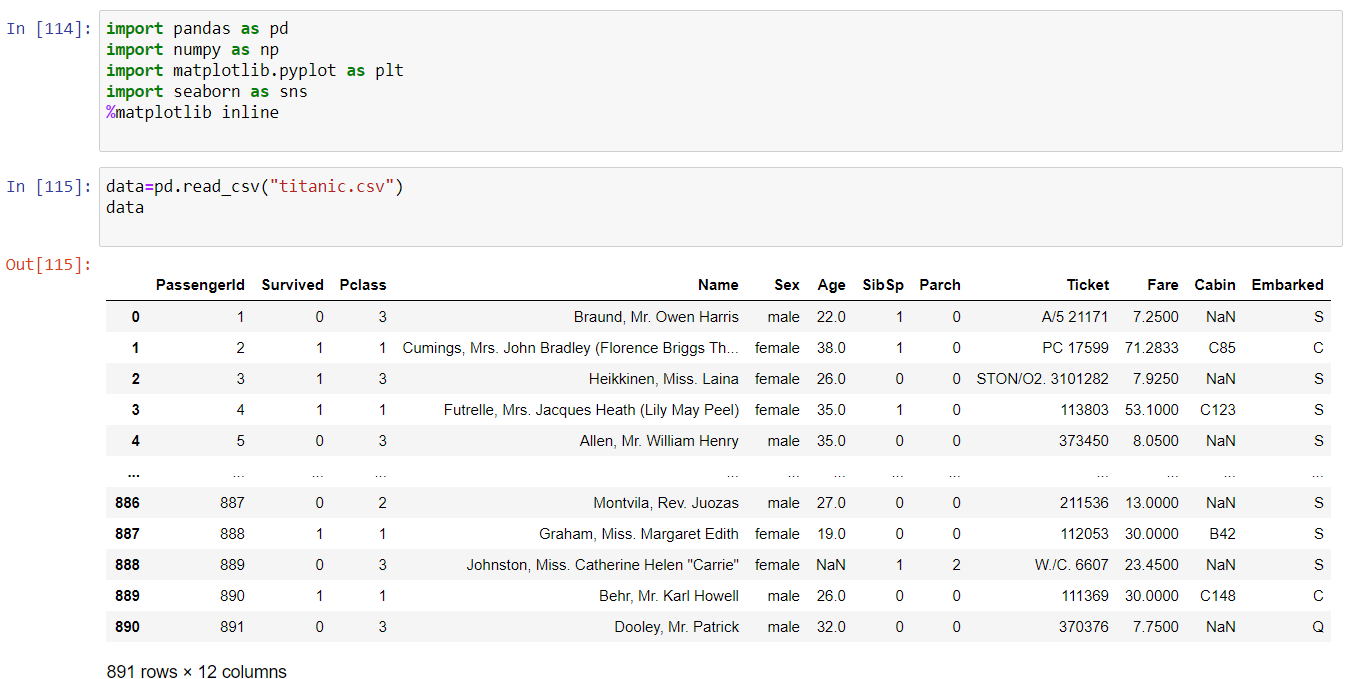
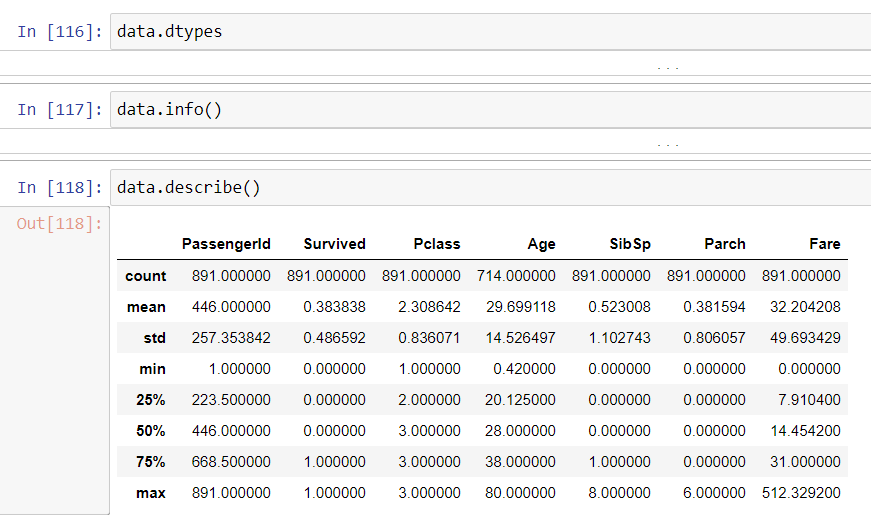
Take any dataset of your own and explore all the data preprocessing steps whichever is required to perform on the data

* Check the data types and get the summary stats
* If you have any missing values in the numerical columns try imputing those columns
  + Create a dataframe by imputing  the missing values with mean
  + Create a dataframe by imputing the missing values with median
* If any missing values in categorical columns, impute it with mode
* For the categorical columns apply  different kinds of encoding techniques
* If there any duplicate values, drop the duplicate values(Do not drop from the original dataframe)
* Visualize the columns with different kinds of plots and write the patterns you have observed.
* Using boxplot, visualize the outliers if present in the dataset
* Identify the Output column in the dataset and divide the entire dataset into Input and Output; create the training and testing datasets.
* If your data requires scaling to be done, apply the feature scaling

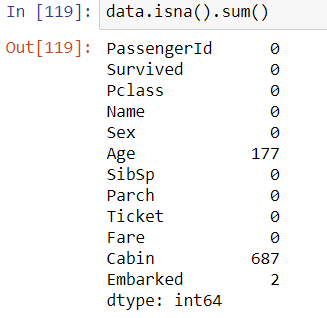
**Import the Libraries and read dataset**



**Check the data types and get the summary stats**



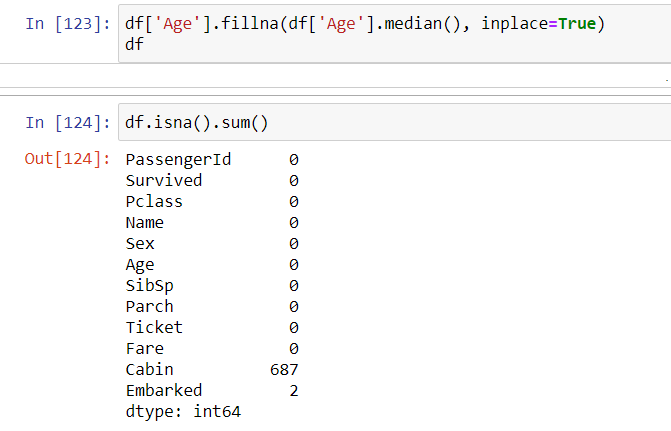
**Find the number of missing values in the dataset**

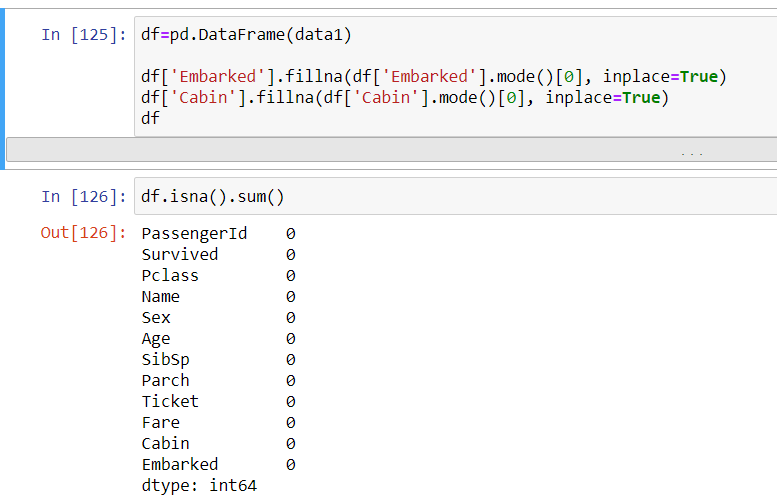


**Create a dataframe by imputing  the missing values with mean**

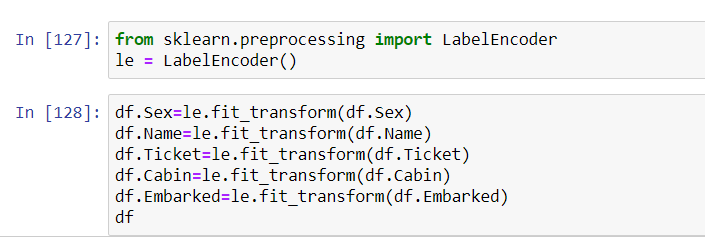
****

**Create a dataframe by imputing the missing values with median**

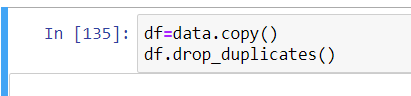
****

**If any missing values in categorical columns, impute it with mode**

**Using LabelEncoder to encode the columns of given dataset**

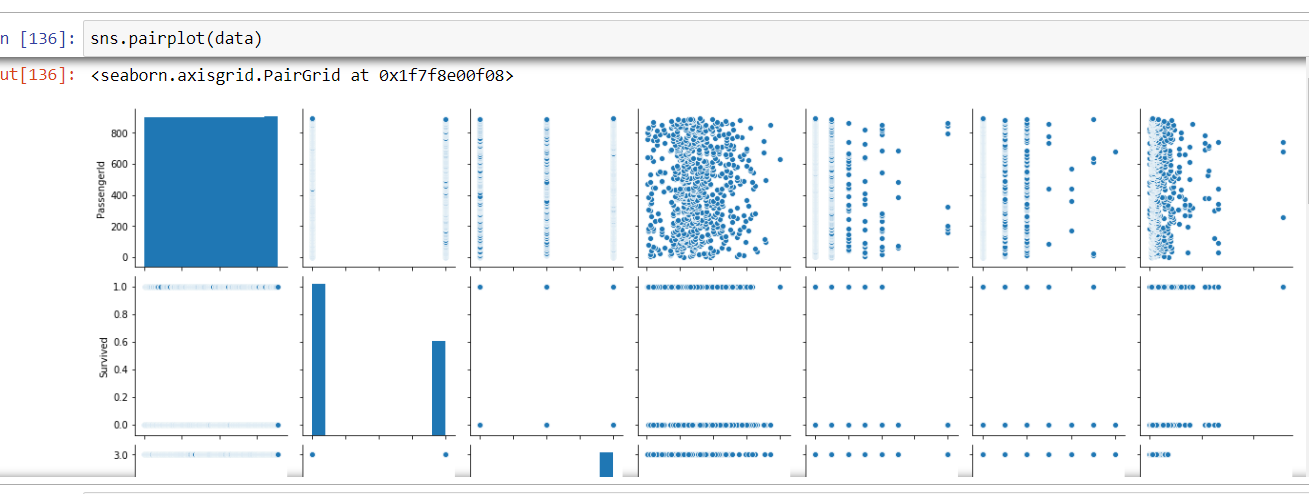


**Drop duplicate values in the dataset(No duplicate values in my dataset)**

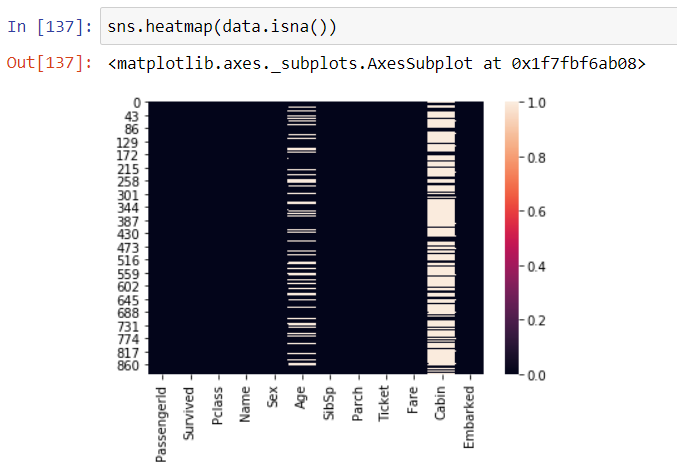


**Visualizing the columns in different kinds of plots:**

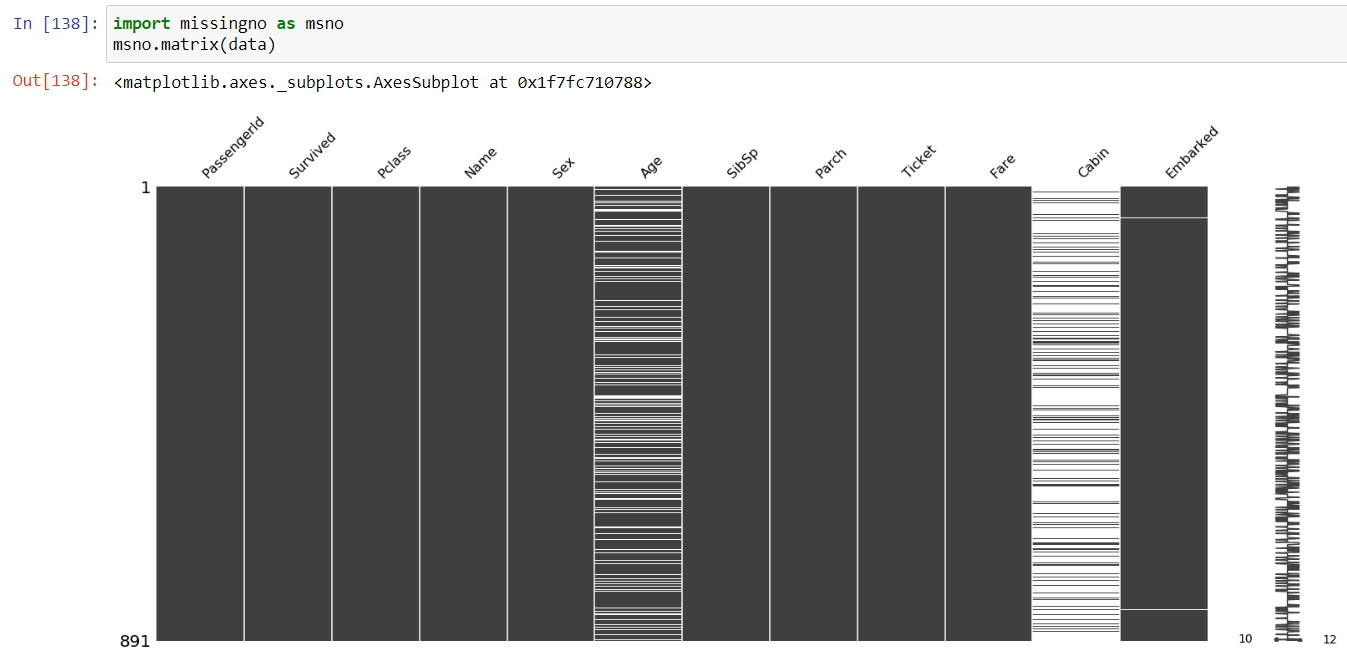
**Pair Plot**

****

**Heat Map**

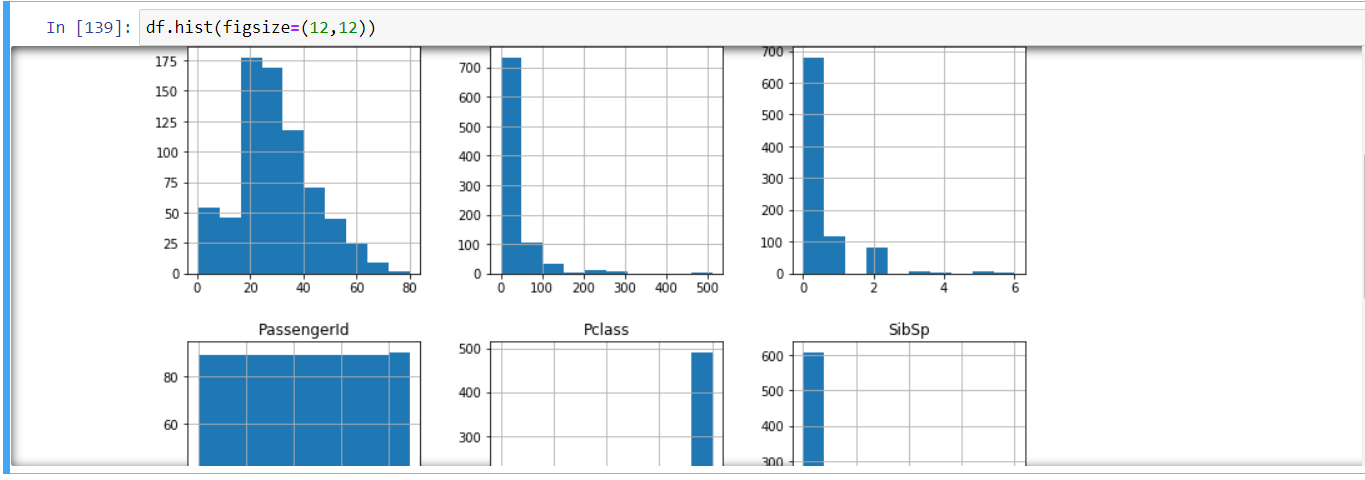


**Missing values plot using missingno**

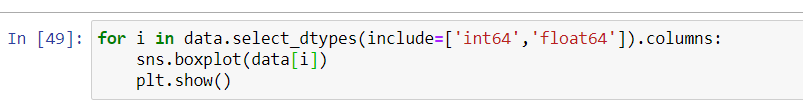
****

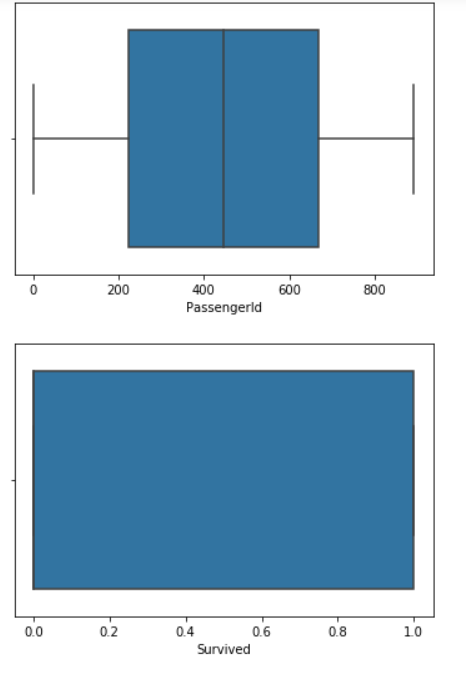
In the above plots Age,Cabin have Null values where Embarked have only 2 null values

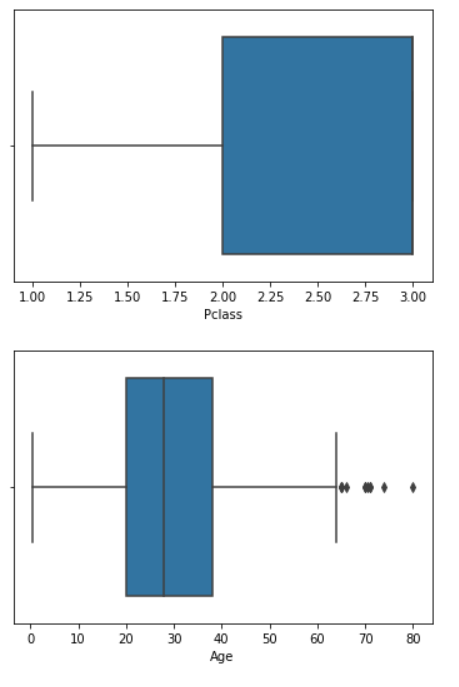
**Histogram**

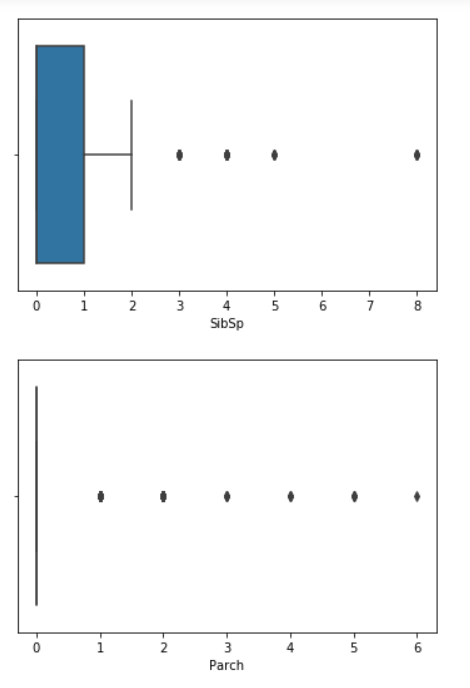
****

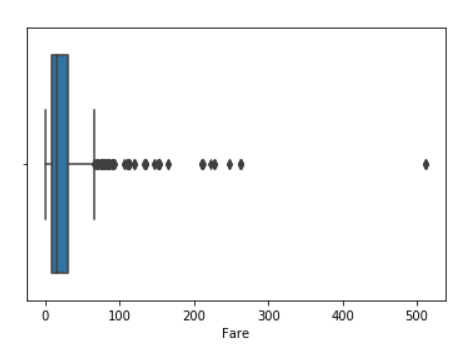
**Boxplot for Each  Column**











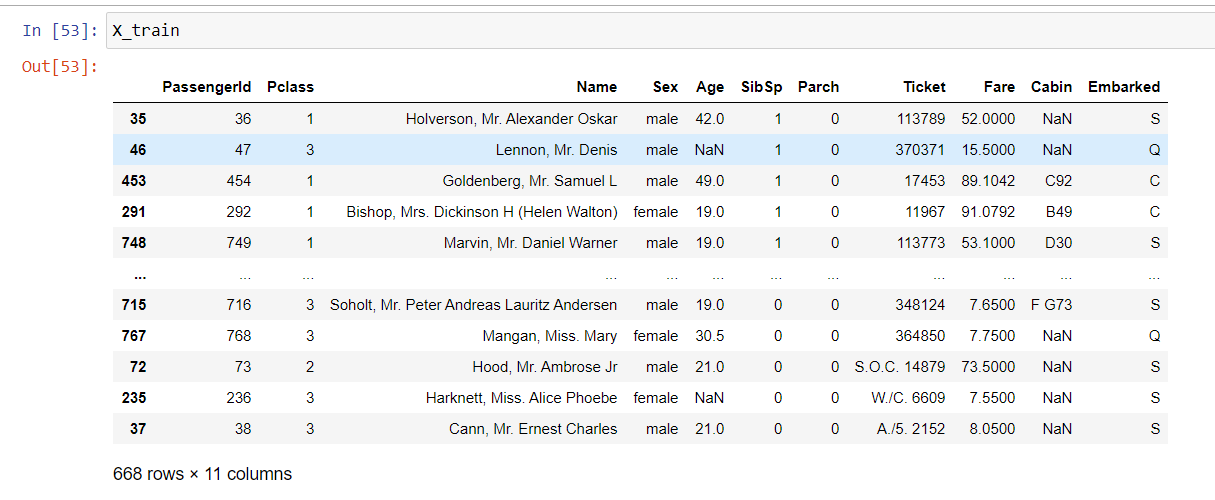
Pclass,Survived and PassengerID do not have any Outlier

**Identify the Output column in the dataset and divide the entire dataset into Input and Output; create the training and testing datasets.**

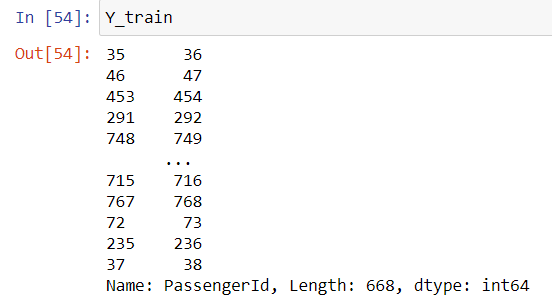
**Crete the input and output of Training and testing dataset**

****

**Input Training Data**



**Output Training Data**

****