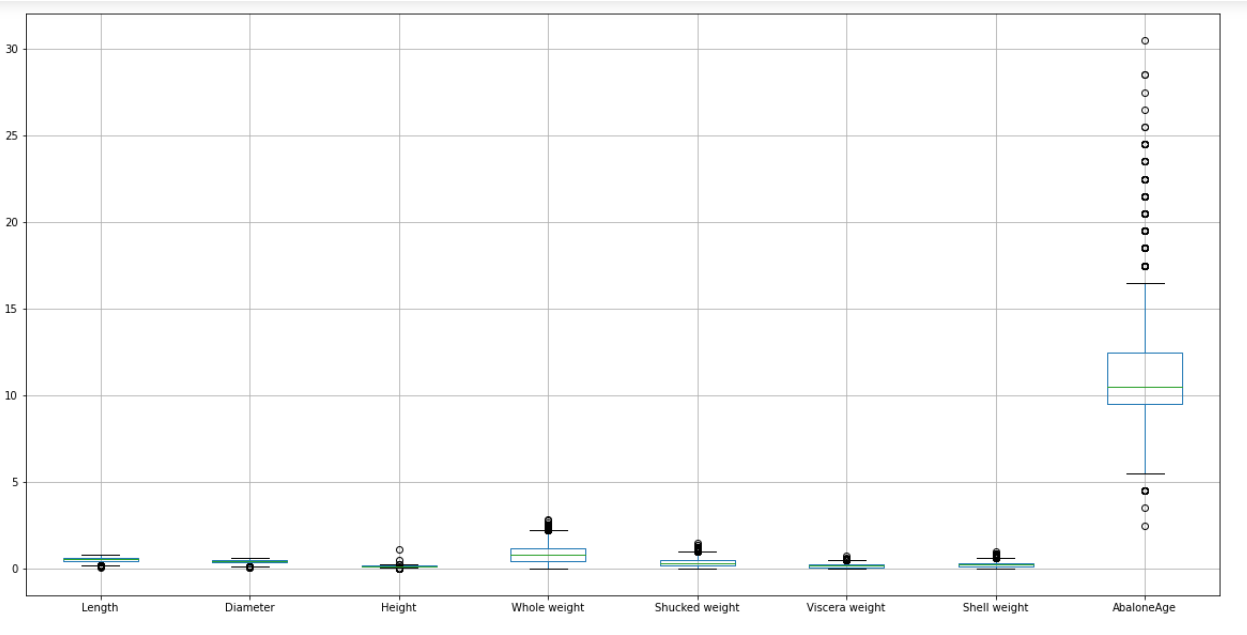
For this project the abalone data set is considered that consisted different details of the abalones as sample where there are initially 9 attributes or columns. The attributes are Sex, Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, Shell weight and the count of Rings. From the analysis of the selected dataset it is found that, it does not contain any null values. In case of height attribute consist minimum value as 0 and not any other attribute has the minimum value as 0.

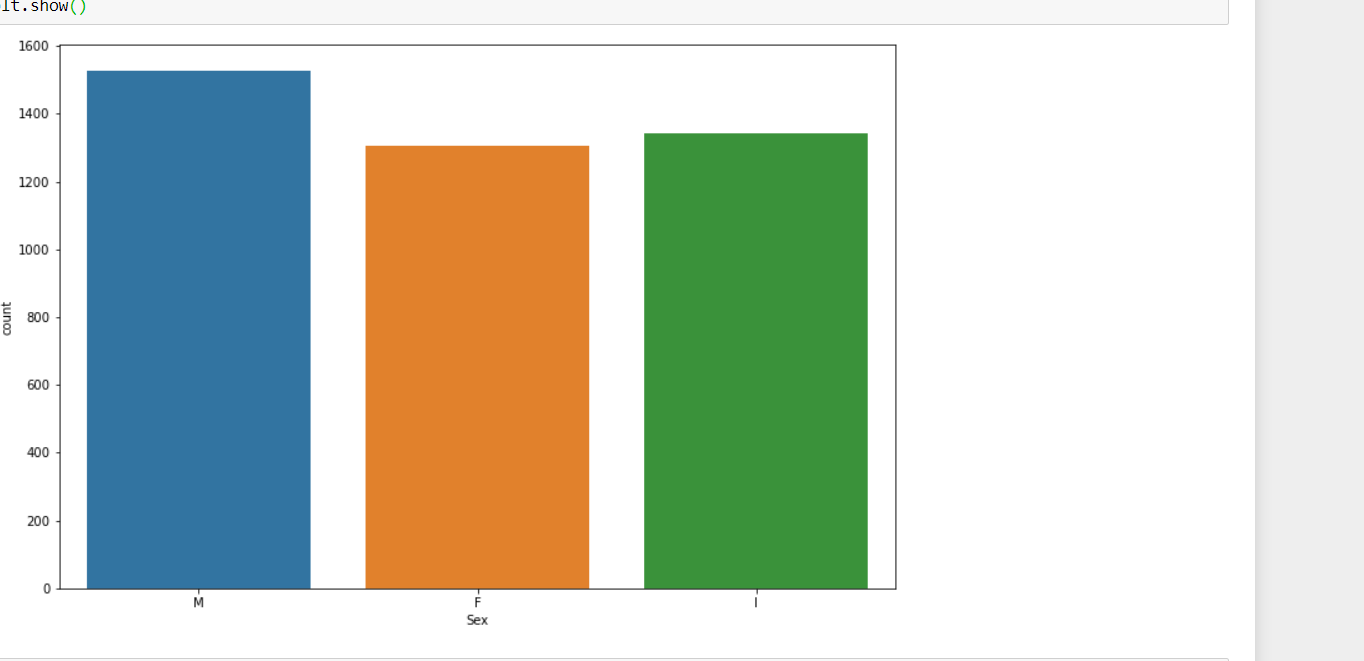
Except the sex all other attributes are numerical values and the sex is categorical attribute for the abalones. As the age can be deduced by adding 1.5 to the number of rings on the abalones shells thus a new column Age is created as the target feature for this data analysis project. While trying to pre-process and cleanse the data, it is found that Age column consisted of lot of outliers, so does few other features that are intended to be used in the further analysis.



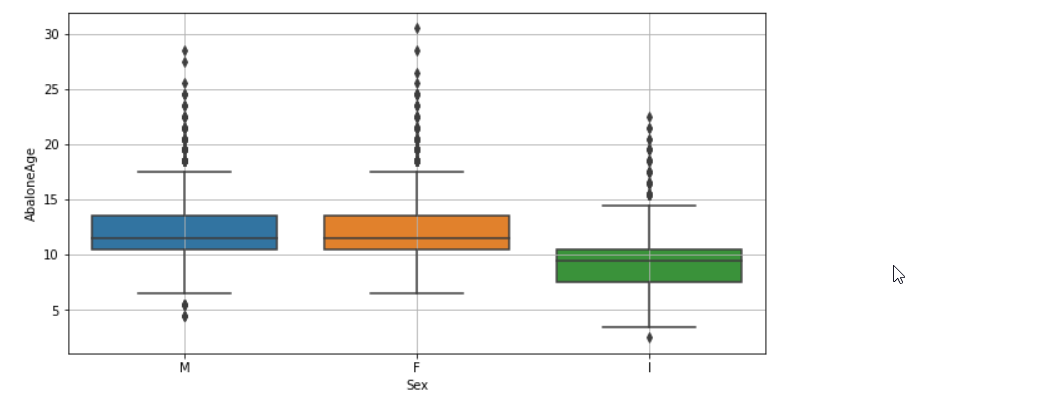
Following is the descriptive statistics for the selected dataset;

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Length | Diameter | Height | Whole weight | Shucked weight | Viscera weight | Shell weight | AbaloneAge |
| count | 4177 | 4177 | 4177 | 4177 | 4177 | 4177 | 4177 | 4177 |
| mean | 0.523992 | 0.407881 | 0.139516 | 0.828742 | 0.359367 | 0.180594 | 0.238831 | 11.43368 |
| std | 0.120093 | 0.09924 | 0.041827 | 0.490389 | 0.221963 | 0.109614 | 0.139203 | 3.224169 |
| min | 0.075 | 0.055 | 0 | 0.002 | 0.001 | 0.0005 | 0.0015 | 2.5 |
| 25% | 0.45 | 0.35 | 0.115 | 0.4415 | 0.186 | 0.0935 | 0.13 | 9.5 |
| 50% | 0.545 | 0.425 | 0.14 | 0.7995 | 0.336 | 0.171 | 0.234 | 10.5 |
| 75% | 0.615 | 0.48 | 0.165 | 1.153 | 0.502 | 0.253 | 0.329 | 12.5 |
| max | 0.815 | 0.65 | 1.13 | 2.8255 | 1.488 | 0.76 | 1.005 | 30.5 |

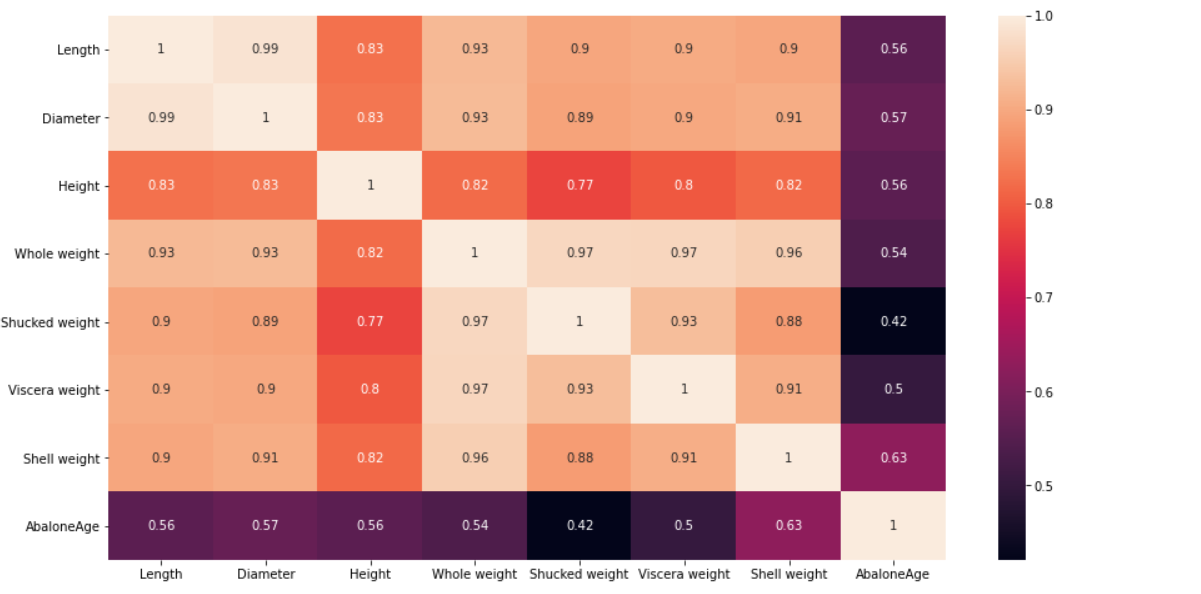
In the further analysis it is found that the selected seven features are not normally distributed in the available dataset. The main objective of this project is to Anticipate or predict the age of abalones from different physical estimations physical data. Usually the he age of any abalone is calculated by counting the carving of the shell through the cone, or recolouring it and finally tallying the total number of rings on the shells using the magnifying instruments. From the dataset it is found that there are highest number of male and infant abalones data are stored in the dataset;



Highest number of outliers are recorded for female abalones as depicted below;



In the next stage the correlation among the attributes are tested using the following heat map;



After which depending on the features a logistic regression model is built whose train and test results are provided below;

Classification Report:

precision recall f1-score support

0 0.77 0.83 0.80 635

1 0.81 0.75 0.78 619

accuracy 0.79 1254

macro avg 0.79 0.79 0.79 1254

weighted avg 0.79 0.79 0.79 1254

Confusion Matrix:

[[525 110]

[156 463]]

Training Score: 0.7858364693807732

Testing Score: 0.7878787878787878