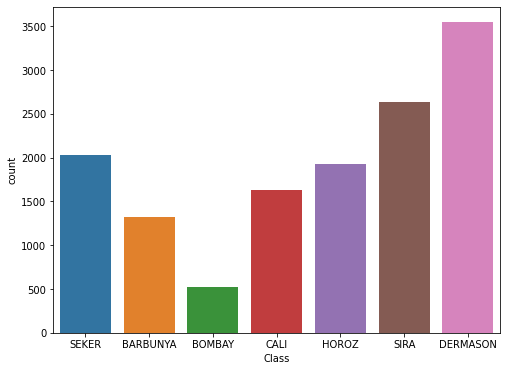
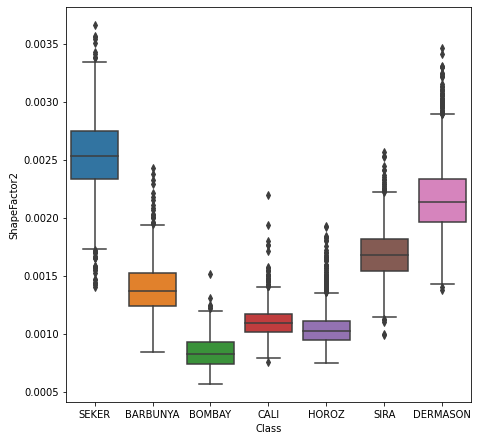
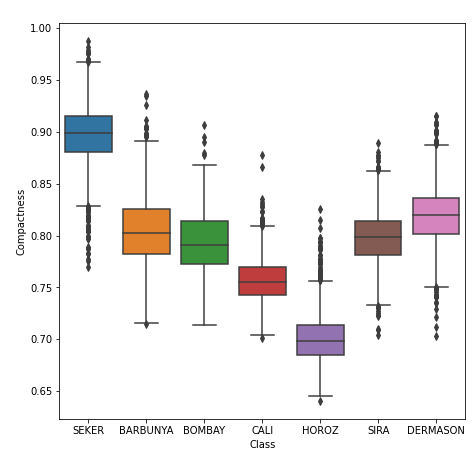
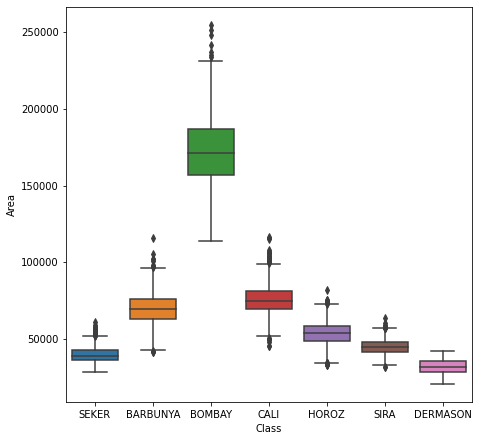
**Dry Bean Dataset - Data Analysis Report**

Data Overview: The dataset had more than 13000 Dry Bean Data with 7 classes.



The “Bombay” class beans in dataset were less but it was enough to make an unbiased model.

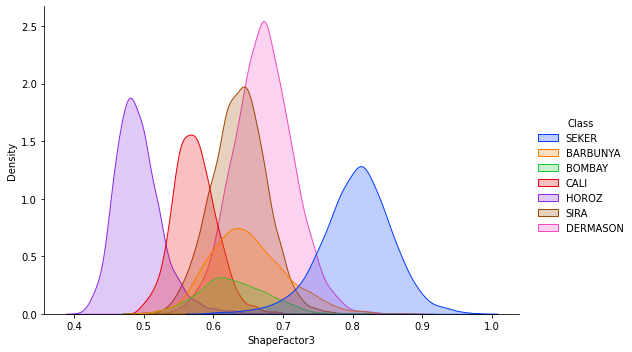
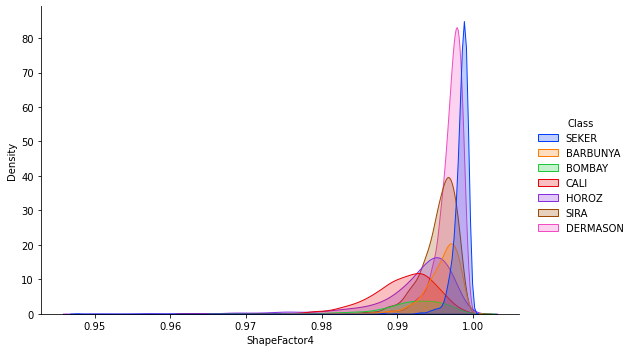
Boxplots were useful to determine the physical characteristic of the dataset.

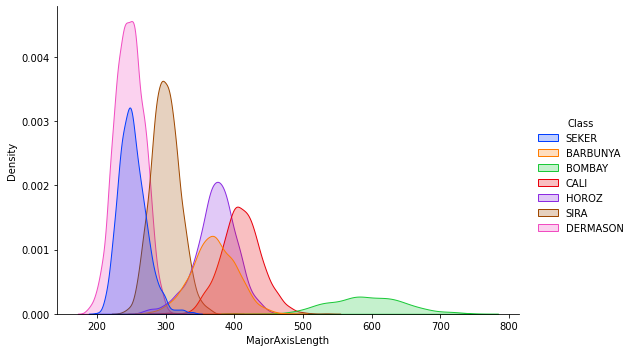
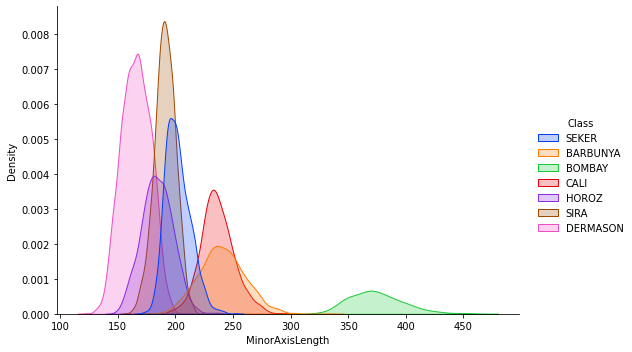
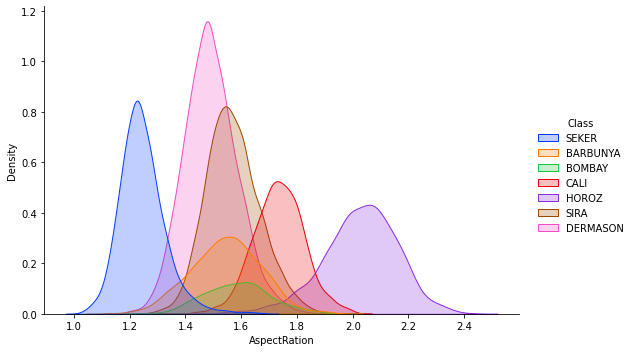
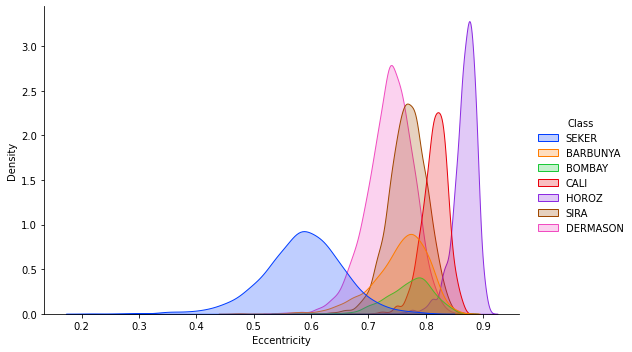
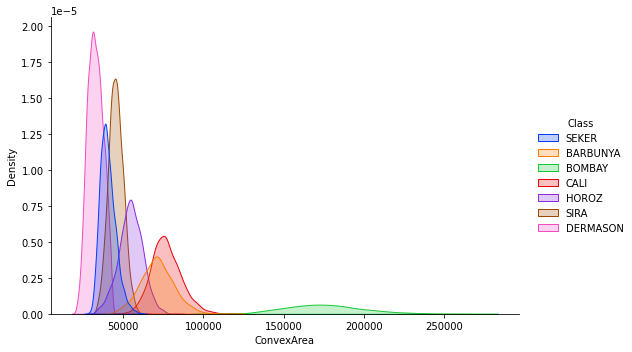
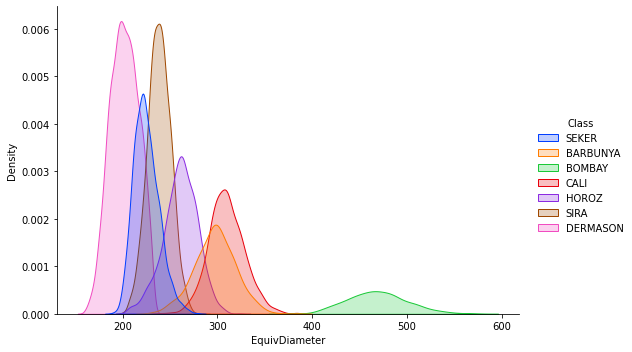
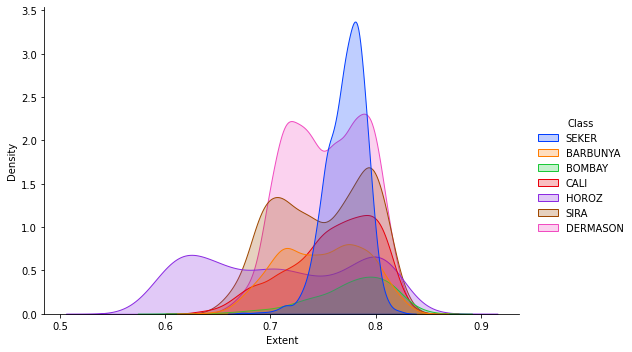
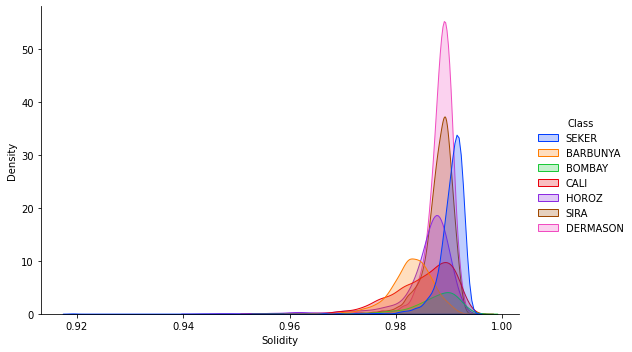
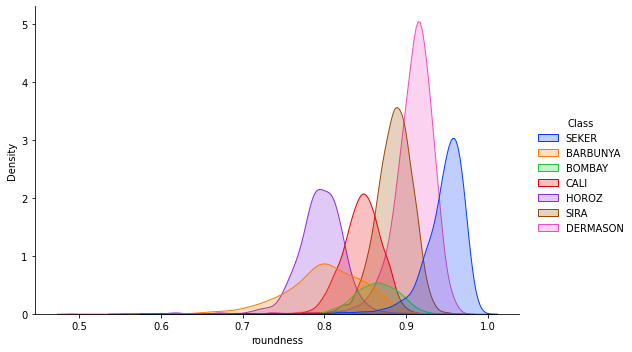
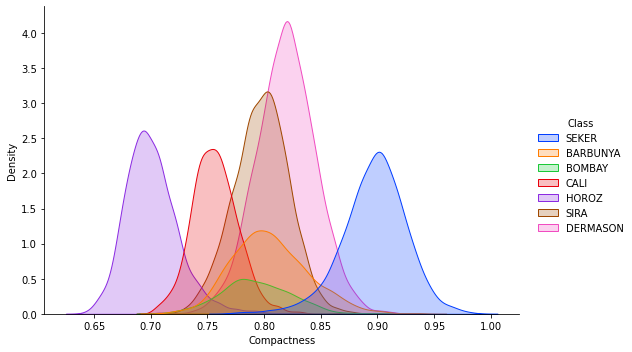
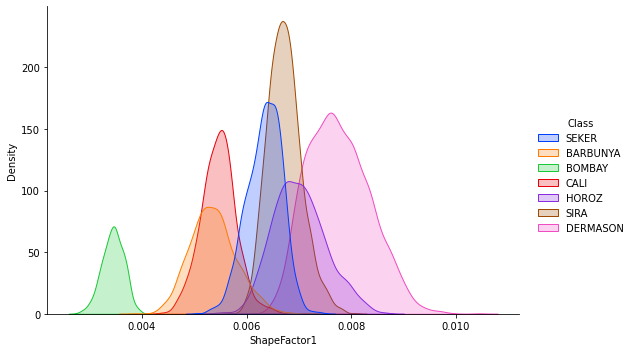
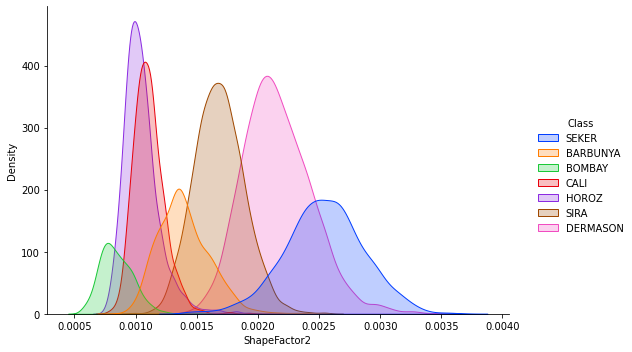


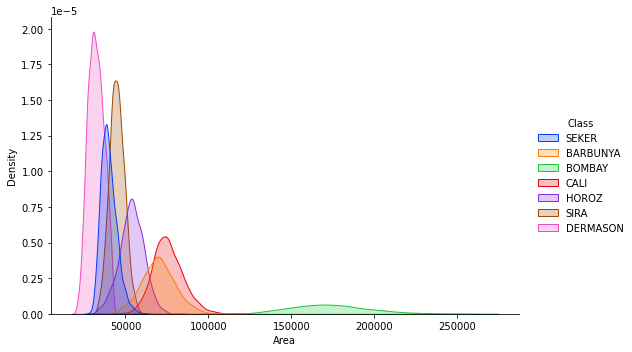
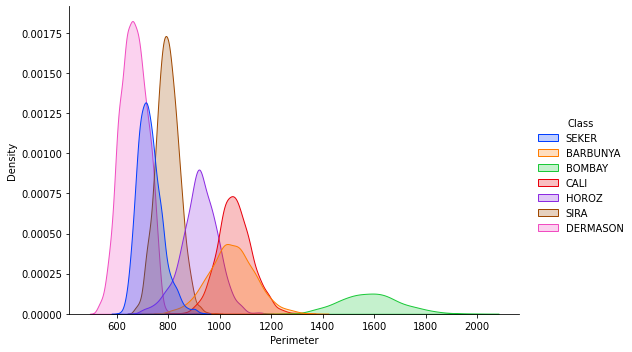
* Bombay Class beans are generally large and Dreamason are smallest in given classes.
* Seker beans are very compact as compared to other classes
* ShapeFactor2 of Seker2 is quite high with a indicator that this feature is quite relatable to the compactness.

Distribution Curves:

The distribution curves were normal for most of the features but there were some classes with \positively skewed and negatively skewed distributions.







The density distribution curves were normal for except some features of a few classes.

* The density distribution curve of the ShapeFactor4 , Roundness, Solidity of HOROZ Class were negatively skewed.
* Density distribution curve of roundness of SEKER was positively skewed.

The variations in distribution and presence of outliers in the dataset were removed by the use of IQR index set to 10 percentile.

Predictive Analysis of was done on 25 percent of the dataset with the help of XGBoostClassifier.

Accuracy obtained from the training dataset was 95.44% and 94.03% from test set. Thus, model is working quite good with low bias and a small variance.