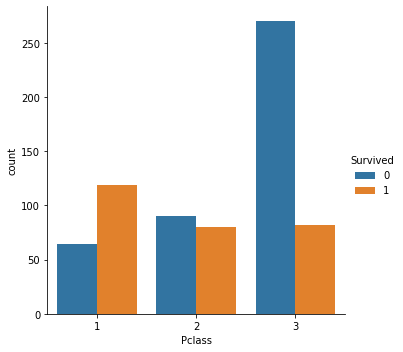
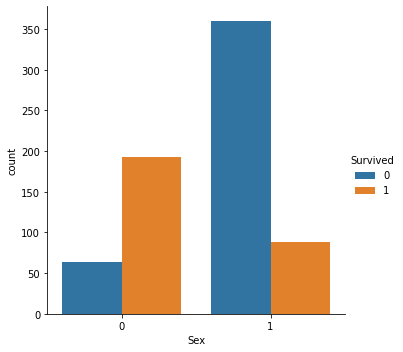
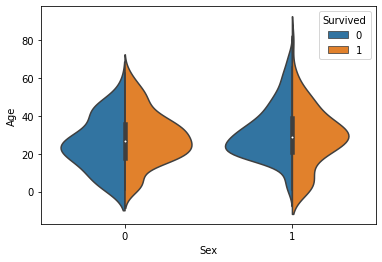
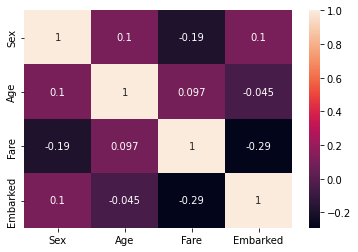
PRML LAB-3

Name: Shalin Jain

Roll No.:- B21CS070

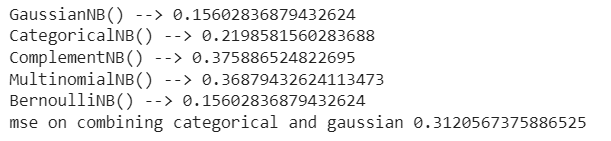
# Q-1)

## DATA VISUALIZATION

We mainly visualize the plots of “Pclass” and its count, “Sex” and its count, Relation between “Sex” and “Age”. All plots have “Survived” as the key. For the same plots are: -

## Identifying the naïve bayes

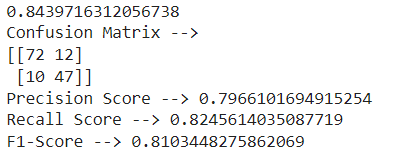
For identifying the naïve bayes variant we can find all the scores of different variants as shown:



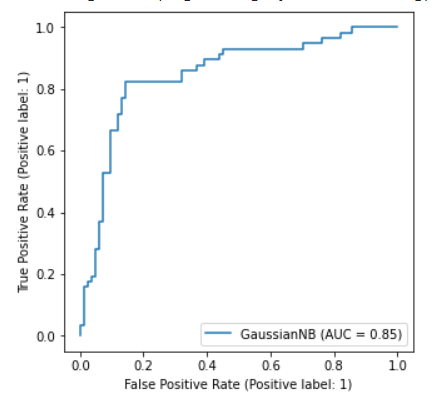
From the above observations of mean squared errors we can confirm that Gaussian naïve bayes variant should be used. We will also use the Heat map to rule out the highly correlated features.

## Performance based on metrics

Evaluation Metrics: -



Plotting of curve: -

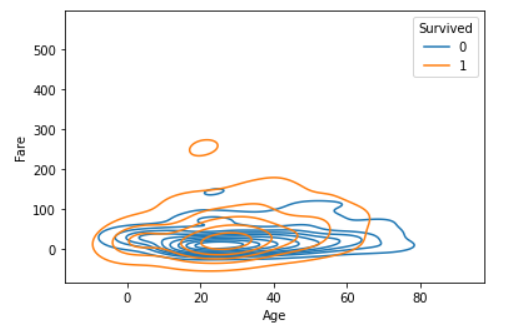


## 5-fold Cross Validation

On performing Cross Validation, we get 77% of score as mean score.

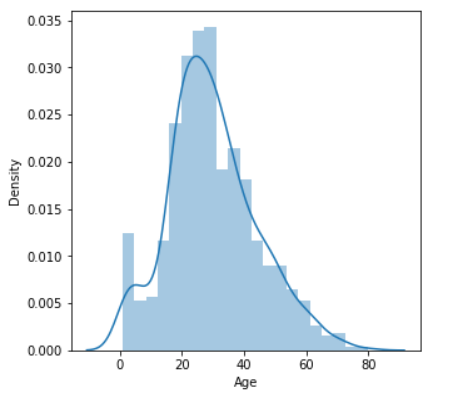


## Plotting the Contour Plots between different Features



## Comparing With the Decision Tree Classifier

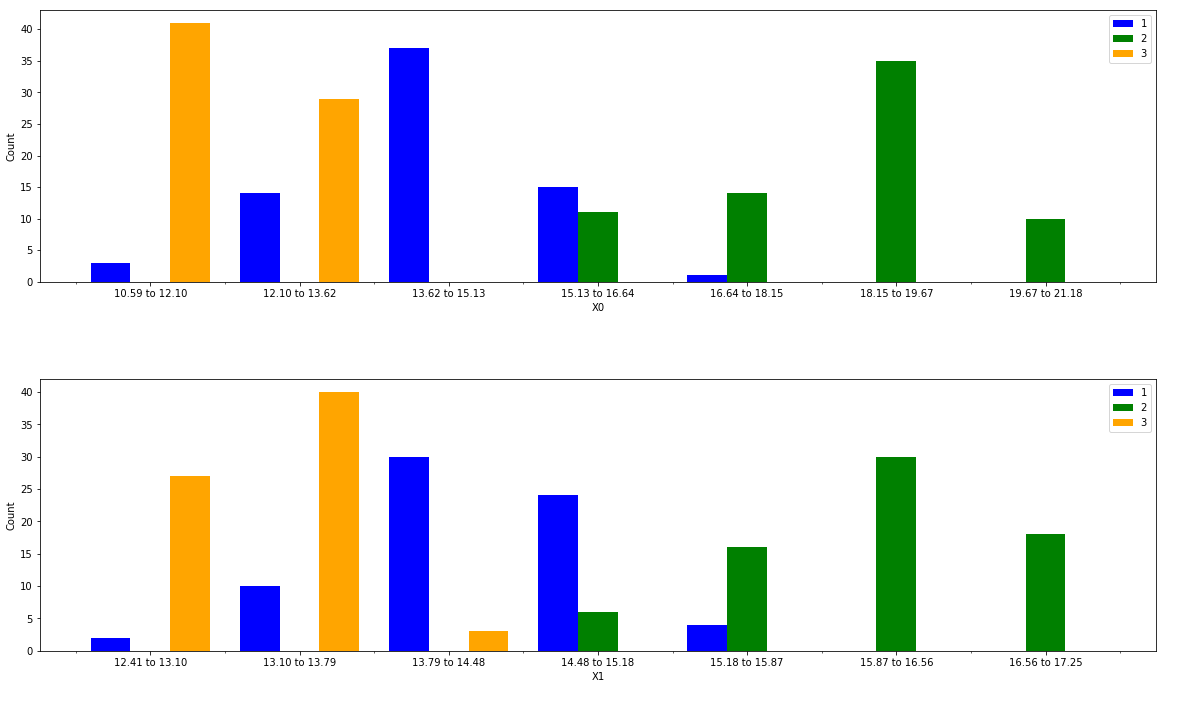


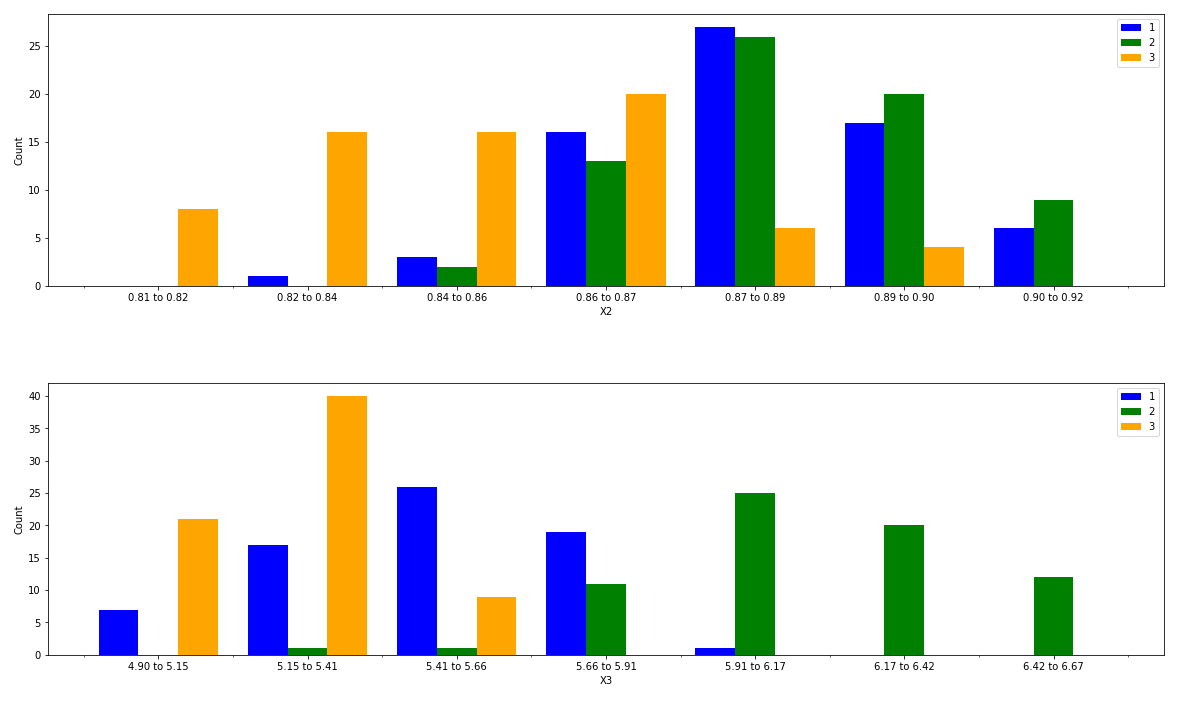


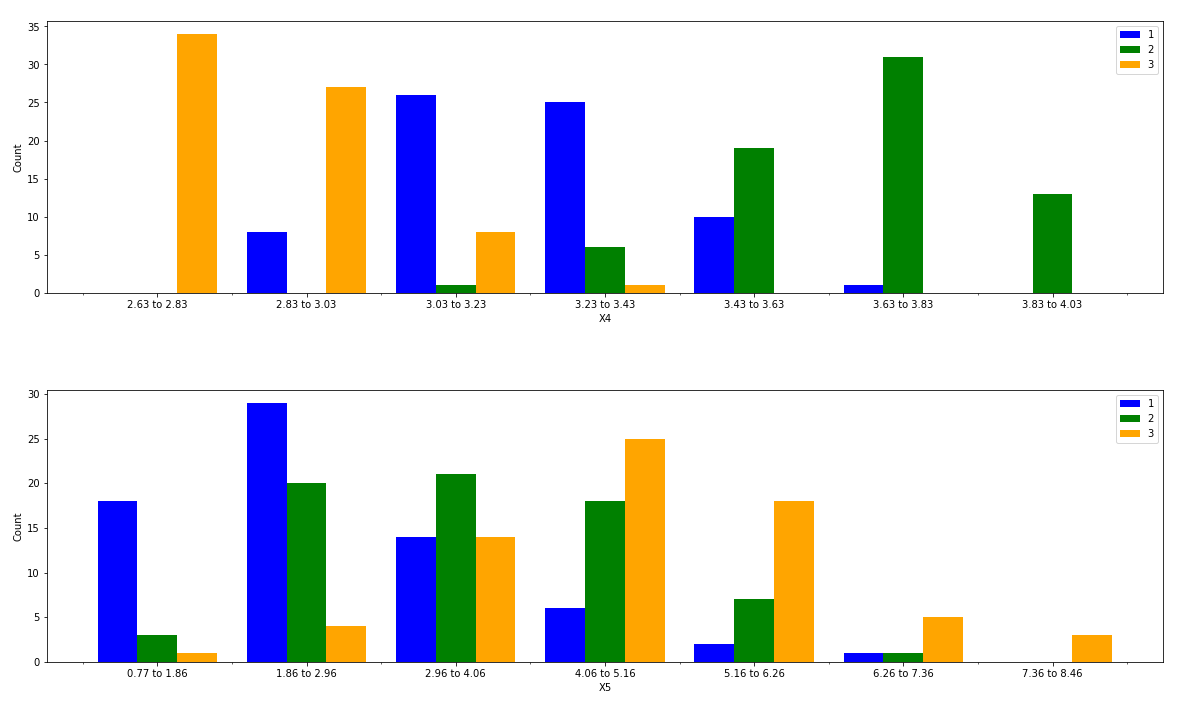
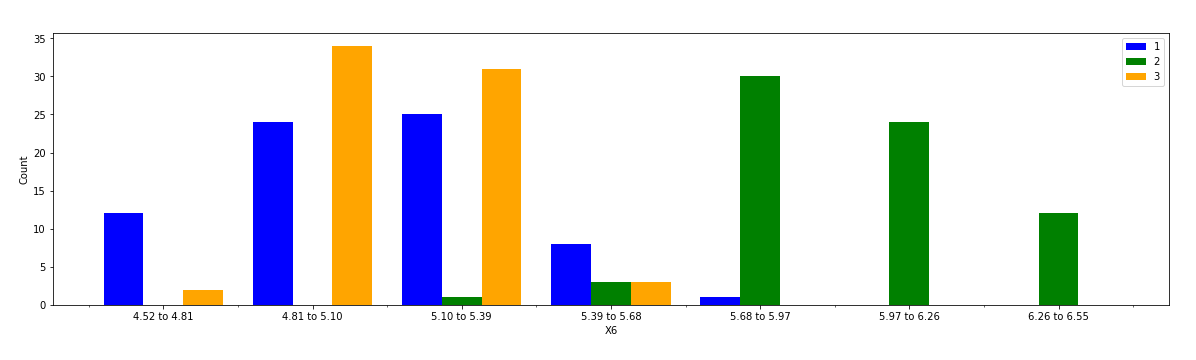
We can see that it is normal distribution hence Gaussian Variant of Naïve Bayes is better than Decision Tree Classifier. In Naive Bayes we assume that features are independent of each other. A major advantage of this variant is that it rarely produces a overfitted model.

# Q-2)

## Plotting Histogram



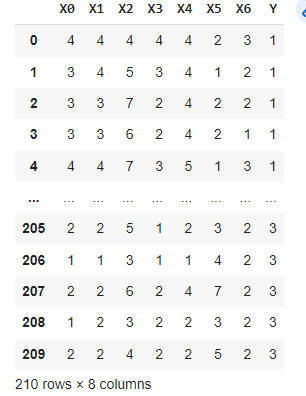




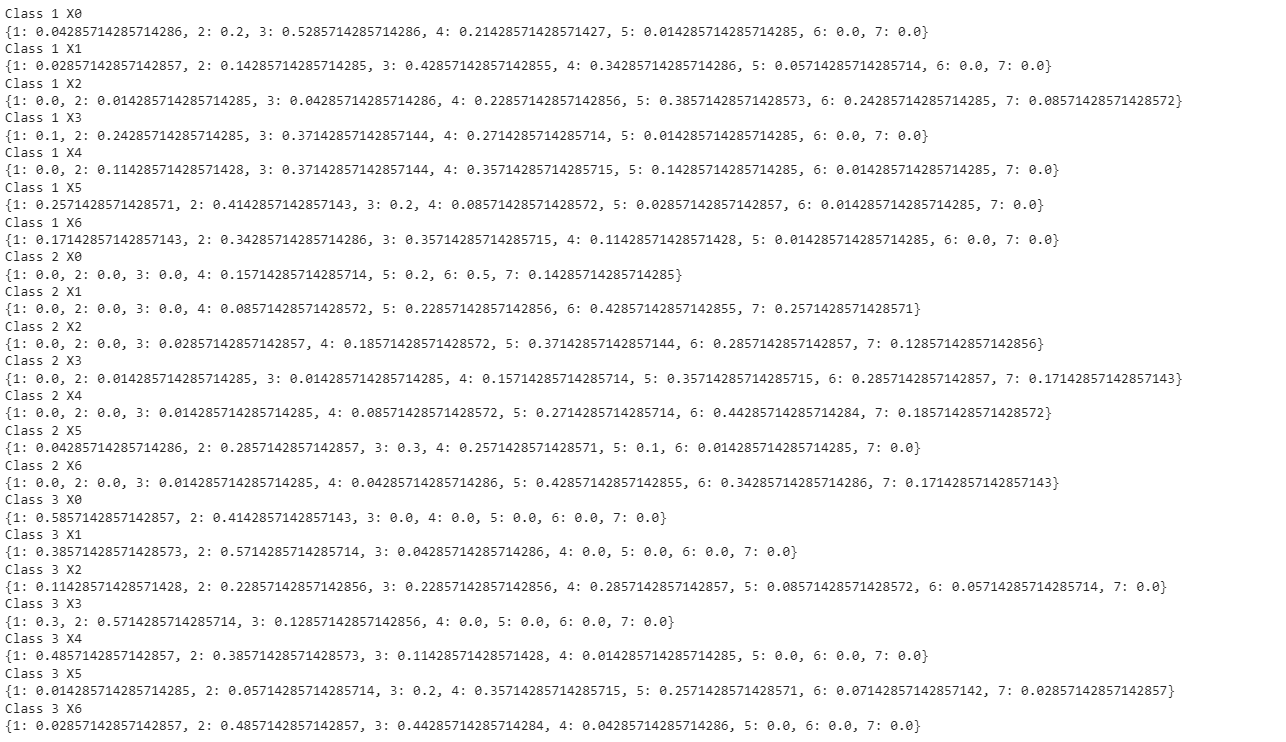
## Prior Probability of Classes



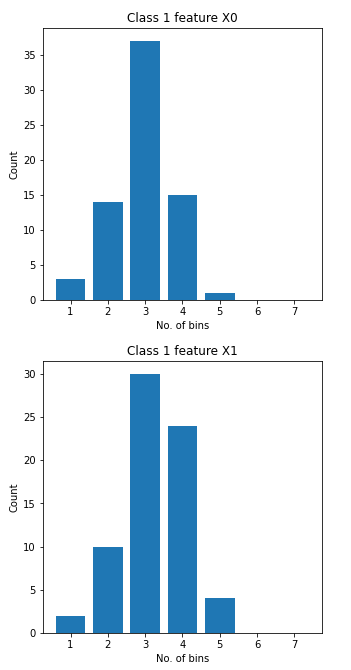
## Discretize the bins



## Likelihood/Class Conditional Probability

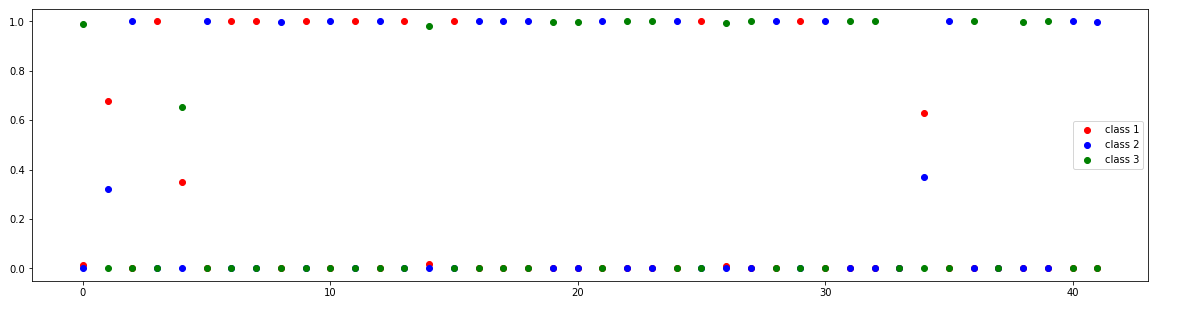


## Comparing the Plots



By comparing the plots, we can see that the histogram plotted for a particular feature and a particular class is similar to the count of unique elements after discretizing as bins.

## Posterior Probabilities



By analysing the plot we can find that posterior probabilities are nearly 1 or 0 for most of the cases.