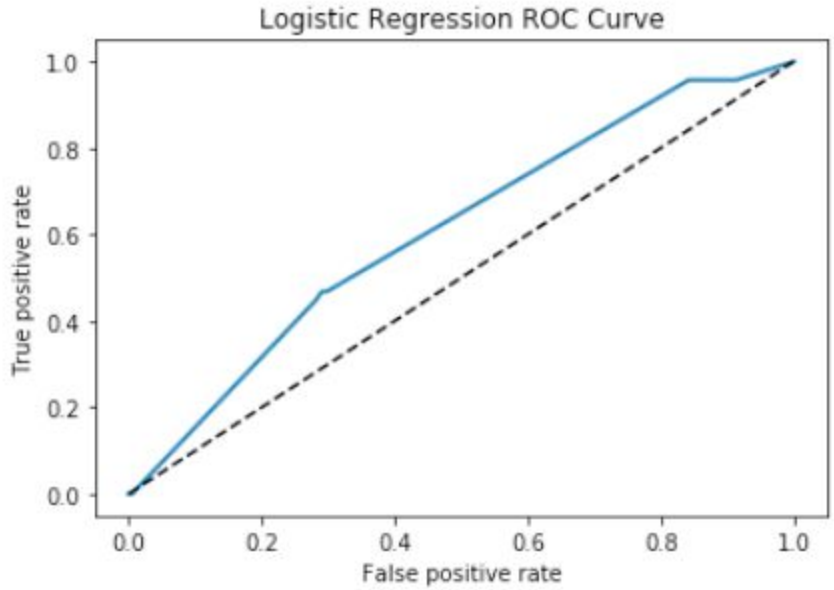
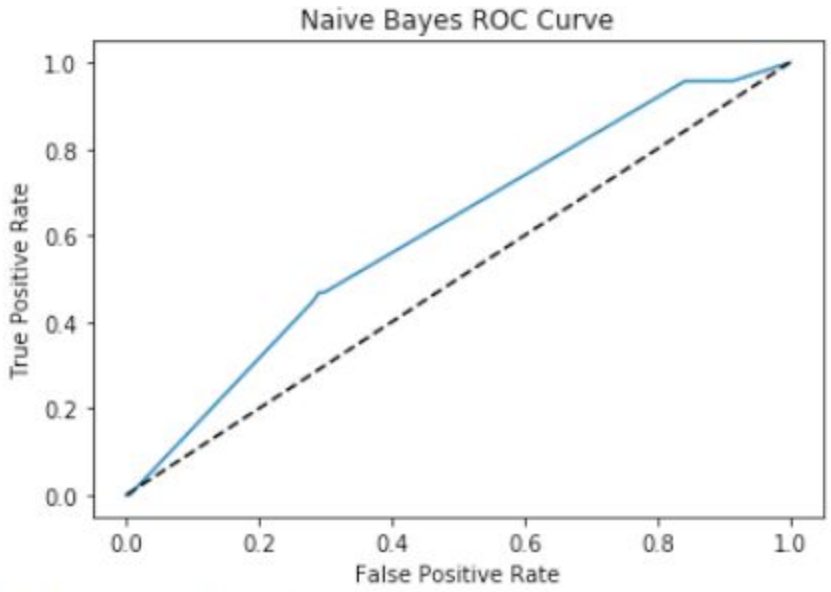
Problem Definition: The goal of this project is to predict the binary response variable which represents whether a client will subscribe to a term deposit. Three binary features including default, housing, and loan were used to generate a Naive Bayes and Logistic Regression Model. Model efficacy and accuracy were evaluated using ROC curve visualizations and AUC scores.

Research Design/Programming: Data was ingested and subsequently split into respective train/test sets using a 90/10 split. 3 - Fold cross validation was used to gauge model accuracy. Model accuracy was approximately 88% for each of the 3 folds. This accuracy was not a result of model efficiency; rather, it was due to the disproportionate number of 0 responses relative to 1 responses. Approximately 90% of the response values in the entire dataset were 0 and only 10% were 1. This means that by simply guessing 0, you have a 90% chance of choosing an accurate response.



AUC score of 0.6103395870453832

The Naive Bayes model displayed near identical results to the Logistic Regression Model. Any differences that did exist between the two models were negligible and would have little to no practical impact.



0.6103395870453832

Recommendations for Management: Both the Logistic Regression and Naive Bayes model performed with AUC scores of approximately 0.61. This is slightly better than guessing. I recommend the bank collects more data, specifically more positive response cases (1). This will provide more learning cases for the algorithm to be trained thus producing a more robust model that can more accurately predict 1's. In the future we can adjust the threshold value to increase the recall rate. This will lower precision but will increase the prediction rate of 1's.