**Modeling a Customers Propensity to Join our Loyalty Program by Tate Kennedy**

My process starts by allowing the research question to guide my model-building and data exploration processes. In this case, the question is can we model a customer’s propensity to join our loyalty program. Questions about odds or likelihoods will use models specifically designed to deal with binary outcomes.

With the question in hand and the data loaded into the customers dataframe, I began my data exploration processes by looking at individual variables (univariate) and the pairwise relationship amongst variables (bivariate). I continued by cleaning and transforming the dataset, whereby one variable was removed altogether and others were transformed into numerical features for modeling purposes (binarized). From there, I looked to prioritize those features that would explain a large amount of variance in the outcome of interest. Due to the low volume of independent variables and results from the Principal Components Analysis, all available variables were utilized in the modeling process.

As the outcome of interest, loyalty, was an observation we had observed previously in this dataset, I was able to turn to a Supervised Learning approach. Furthermore, as loyalty was a categorical outcome variable (True or False), I looked to utilize Classification models. With my data split between training and testing sets, I was able to attempt several different classification algorithms. I used accuracy rates, confusion matrices and cross validation to evaluate each classifier.

Decision Tree produced the lowest accuracy rate and was the worst performer amongst those attempted. More than one algorithm produced similar accuracy rates at a peak of nearly 87%. This includes Logistic Regression (86.8%), Gradient Boosting Classifier (86.6%), Multi-Layer Perceptron Classifier (86.8%), Support Vector Machine Classifier with a RBF Kernel (86.6%) and K-Nearest Neighbors Classifier (86.7%). SVM took longer to run than the other top performers. Additional fine tuning of hyper-parameters and perhaps additional variables could help strengthen one or more of these models further in order to increase the number of customers predicted to join our loyalty program.