**Data Exploration**

Upon initially receiving this data set, I took time to look at the structure of the data and to think critically about the features that would be best suited in a model. I notice data quality issues and addressed them, and I was able to look at the distribution of several features graphically as seen in block 7. I looked at them both graphically and in the csv file to get an understanding of how the data was written into the csv itself.

**Data Preparation**

After looking at the data, I addressed making sure that the data was all classified correctly. This included changing any Yes/No variable into a binary variable. I also did some binning of discrete variables as well as dropping the columns that I had already transformed. I then split my data into a train and test set. After I completed this task, I checked for interactions and did a chi-squared test and then included the ones with a significant p-value. I made sure that all the variables that needed to be in some form of bin or category, like income and sex were transformed.

**Build model**

I elected to do three different classification models. One logistic, one knn and one SVM. I trained and tested all three. I chose to include the significant interactions from my logistic model into my SVM model as well.

**Select**

Overall, I found that the knn preformed the best and used that to make my final prediction. I chose to use the model that had showed the lowest error rate.