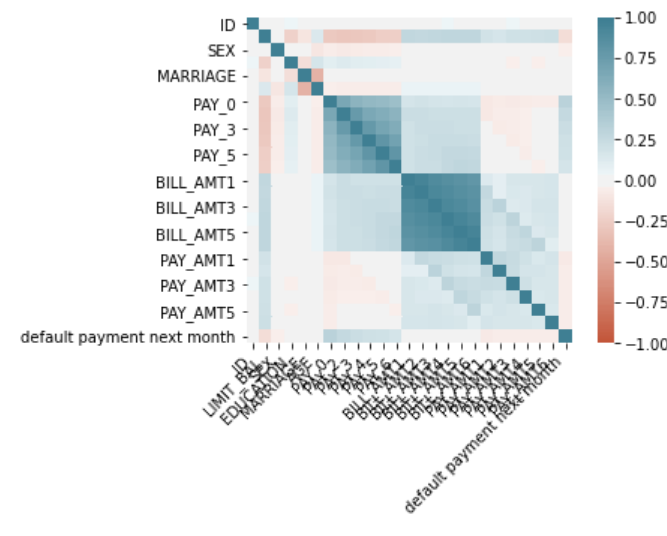
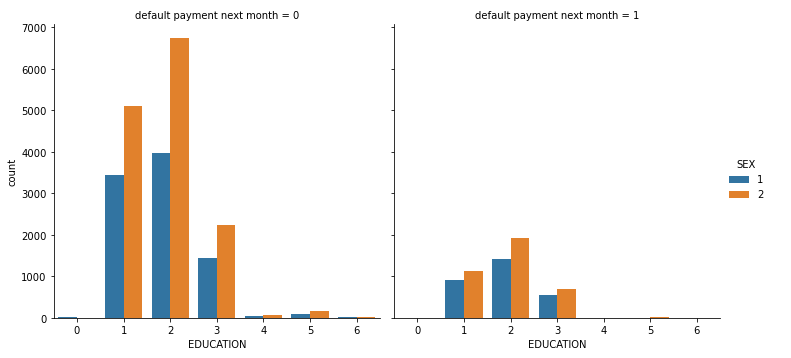
**Credit One Report**

Credit One currently has an issue where there has been an increase in the amount of customer default rates. This issue impacts Credit Ones business as it may result in a loss of Credit One’s business customers since Credit One is responsible for approving the loans in the first place.

In order to resolve Credit One’s issue, we attempt to develop a model that can predict a customer’s chance to default based on different variables. To start off Credit One has provided a detailed excel sheet that contains customer info which may prove essential when determining if a customer will default or not. The data provided may also contain data that may prove inconsequential to the end results, so in order to create a more efficient process the data will need to be preprocessed in a process known as EDA. Based on this EDA or exploratory data analysis, several important factors were discovered and detailed below.

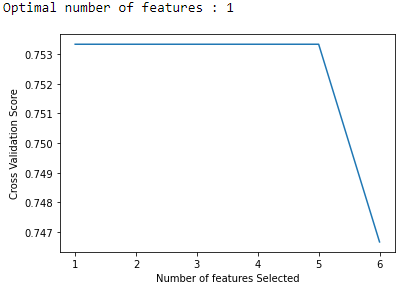


The figure above represents a correlation matrix between all the variables in the data set. Variables with high correlation and aren’t the dependent variable are usually removed as they may cause overfitting in the resulting model predictions. However, with new customers we might not have Pay, Bill Amount, or Pay Amount so they will be removed from consideration.



Taking a look at some other variables in the data set, we can see that females with lower education levels have a higher default rate then males of similar education levels. However, Females also represent a higher portion of those who took out loans.

With the exploratory data analysis behind us we begin the process of building and testing our models. The data provided was separated into a training set and a test set then fed into three different models support vector regression, random forest and linear regression. Before testing however, a recursive feature elimination with cross validation was ran on a part of the data set to determine the best features to use when training the models. The result of that process can be seen below.



The optimal feature selected was Education.

For each model training was performed for the default data set, with the selected feature, and with at least two tuned parameters. The best performing model was Linear Regression with an RMSE of 0.415 and an Accuracy of 0.770 this is with a fitted model and no selected feature. The results of all models can be found below.

Random Forest:

Default: RMSE: 0.438 Accuracy: 0.739

Selected Feature: RMSE: 0.420 Accuracy: 0.770

Parameters Tuned: RMSE: 0.420 Accuracy: 0.766

Parameters Tuned with Selected Feature: RMSE: 0.420 Accuracy: 0.770

Support Vector Regression:

Default: RMSE: 0.440 Accuracy: 0.770

Selected Feature: RMSE: 0.440 Accuracy: 0.770

Parameters Tuned: RMSE: 0.467 Accuracy: 0.712

Parameters Tuned with Selected Feature: RMSE: 0.467 Accuracy: 0.712

Linear Regression:

Default: RMSE: 0.415 Accuracy: 0.770

Selected Feature: RMSE: 0.421 Accuracy: 0.770

Parameters Tuned: RMSE: 0.415 Accuracy: 0.770

Parameters Tuned with Selected Feature: RMSE: 0.421 Accuracy: 0.770

Based on the selected model we can predict if a customer will default with 77% accuracy, which may help in determining if Credit One should approve a customers loan.