Abby Kannappan:**Unit 19 Supervising Machine Learning Homework**

The objective of this project is to Predict 2020 Q1 credit risk levels using data from 2019. Two source files were provided for the effort.

Part of this exercise was to use to use two different models to make predictions, logistic regression and random forest classifier.

I predict that both model will do a good job of predicting, and that scaling will help for both models.

For scaled data I predict that both models will provide high scores for accuracy on scaled data.

The following results were provided for unscaled data for logistic regression:

Training Data Score: 0.6508210180623973

Testing Data Score: 0.5168013611229264

These are not good results, but at least we know the model is providing prediction results.

After scaling data the logistic regression model provided the following results.

Training Data Score: 0.7078817733990148

Testing Data Score: 0.767333049766057

These results are improved from the unscaled data however they could provide more accuracy.

The results for the random forest classifier before scaling are as follows:

Training Score: 0.9999178981937603

Testing Score: 0.6412165036154828

The accuracy of the training score are good.

The results after scaling are as follows  
Training Score: 0.9999178981937603

Testing Score: 0.6420672054444917

These are very similar to the results before scaling, so we may conclude that in this case and data the scaling did not help improve accuracy for Random Forest Classifier. Perhaps decision tree analysis doe not need scaled data.

Overall the best model was Random Forest Classifier regardless of the scaling. Also scaling did with this data for Logistic regression.