STAT 6012: Linear Models for Data Science
Class Activity 8

Due date: Thursday, July 25 by 10:50 am Via Canvas.

Complete the following questions an R Markdown file and submit your compiled HTML file. If you are working in a group, list the names (last, first) of the group members in alphabetical order of last names.

The attached dataset is a simulated data set containing information on ten thousand customers. The aim here is to predict which customers will default on their credit card debt, `Balance`.

We want to build a multiple linear regression model to predict `Balance` using:

`Income, Limit, Rating, Cards, Age, Education`.

1. [2] Data exploration: make a correlation plot to compare all the pairwise correlations between predictors. From the correlation plot, does there appear to be issues of multicollinearity?
2. [4] Build a multiple linear regression model and use it to further investigate issues of multicollinearity using the VIF values. If there are multiple multicollinearity issues, rebuild the model.
3. [4] Pick reasonable values of the predictors in your model from Question 2 to make prediction in terms of:
   (a) [2] Prediction interval. Also, interpret your prediction interval in context.
   (b) [2] Confidence interval for prediction. Also, interpret your confidence interval in context.
4. Homework/practice: Does your model meet the model assumptions? If not, can anything be done?