**Part 3: Regression Analysis (125 points)**

Louise Card left her job at the fifth third bank to become the director of alumni relations with College of Business, her alma mater. She is looking for a $20,000 loan to build her dream outdoor kitchen so that she can entertain alums at her house on football game days. She would like to pay the loan off over 36 months. Her income is verified at $150,000. Her credit is very good (A) and credit history goes back to 2003. Having graduated not too long ago, she does have an outstanding mortgage on it. As of now, she has open credit cards with a revolving balance of $4000. You now took the job of the fifth third bank loan officer.

***FROM PAPREGRAPH ABOVE:*** *Loan\_amnt (20,000), long\_term (0), annual\_inc (150,000), grade (A=1, all others =0),* ***credit history (back as far as 2003, 18 years),*** *mort\_acc (1), revol\_bal (4000)*

1. **Develop a regression model to predict interest rate for Louise Card. To see how your model works, please estimate the interest rate of a loan Louise would like for her outdoor kitchen.** 
   1. *Annual\_inc and loan\_amnt are not significant- according to our correlation matrix and the significant figures in our regression model including these variables. We will build a model that can accurately obtain and predict her interest rate without those variables.* 
      1. *Her estimated interest rate is: 29.252+2.940\*10-6 X1+0.003X2+(-3.306\*10-6)X3+(-0.29)X4+0.182X5+(-22.078)X6+(-18.130)X7+(-14.298)X8+(-9.673)X9+(-3.988)X10  = 6.93% is the interest rate*
2. **It may not be possible to incorporate all of the factors provided by Louise into the regression model. You just need to choose based on your regression model.** 
   1. *We could not incorporate in her credit history. There’s no variable for that in the dataset provided.*
3. **You may not have all the information you need for your regression model. What are you going to do in that instance?**
   1. *We decided to find the average value for both all\_util and tot\_rev\_hi\_lim since we found those variables to be highly correlated to the interest rate. The average all\_utilis 53.93 and the average tot\_rev\_hi\_lim is 39191.08.*
4. **In addition to estimating the interest rate, you might want to address the following questions:** 
   1. **If Louise decided to take out a $40,000 loan instead of $20,000, how would that effect the interest rate?**

*It does not have any effect on the interest rate as we removed the loan\_amnt from our regression model. While we correlated the interest rate with the loan\_amnt, there is a low correlation and found out it is insignificant in our regression model.*

* 1. **If Louise instead had a B credit rating, how much more would she pay?**

*Louise would pay 10.88% interest instead of the 6.93% interest rate we calculated with her Grade A credit rating.*

*Using a simple interest formula, she would pay $2370 more over the course of her loan with that increase in interest rate.*

**What is the most important determinant of interest rate?**

*Credit grade has proven to be the most important at determining interest rate. An grade A credit rating has an especially great effect on the interest rate given to a borrower.*

**In your report:**

* 1. **Explain which variables you chose for analyses and why.**

***The variables we chose are :***

* *Revol\_bal*
* *All\_util*
* *Tot\_rev\_hi\_lim*
* *Mort\_acc*
* *Long\_term*
* *Grade A through F (F as our base dummy variable for grade)*

*The correlations between the variables against the interest rate are high and they all are significant in our regression model.*

**Provide the output for each analysis.**

* 1. **If you conclude that Louise is NOT able to get a higher loan, what advices would you give to her to increase her loan amount? Perhaps, increasing the loan terms to 60 months instead of 36 months? What else?**

*We think that the loan\_amnt is not significant in our model based on the significance level and low correlation in the matrix.*

* 1. **Summarize your findings. Clearly outline what Louis should do for her outdoor kitchen.**

*The interest rate (6.93%) for Louise loan is lower than the average interest rate (12.62%) given by the Lending Club and it is a decent interest rate. We think she should be good to take a loan for herself and go forward for her outdoor kitchen plans.*

Louise really wants to show her Cardinal spirit. She would like to spend as much money on the outdoor kitchen as she can get a loan for. She is not sure how much loan can she get approved and is looking at you for help.

1. Develop another regression model to predict the loan amount (variable loan\_amnt). You may notice that the data does not suggest how much loan a customer is eligible for. In other words, loan amount refers to the particular dollar amount loan customers applied for. You may, however, use this as the maximum amount because if the loan requested was higher than the amount customer was eligible for, the loan amount was reduced by Lending Club (from data dictionary: loan\_amnt~The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value).
2. Some of the predictors for this model will be continuous variables and others would be categorical. Include grade as one of the variables.
3. **Treating Grade B as the baseline group**, what is the difference between the loan amount for Grade A, C, D, E and F.

*Putting everything else constant, as compared to grade B, the difference between the loan amount will be:*

* *A = 773.73$*
* *C = -1071.07$*
* *D = -2483.77$*
* *E = -4242.38$*
* *F = -6034.97$*

1. While conducting #3, would you include other variables in the model. Why? Why not?

*Yes, we included while conducting #3. To answer the above question, we held all the variables as constant to provide our insight. Including the other variables helps make the coefficients from Grade A through F significant against the loan\_amnt.*

1. **Louise is wondering if her loan amount would be higher if she pays more money in each installment at a higher interest rate. It makes theoretical sense to you, but you have to provide proof that this is actually true given the historical data.** 
   1. *We ran our model with installment and interest rate as independent variables against loan\_amnt and created an interaction variable. As loan installments increase, the lower interest rates receive higher loan amounts quicker than higher interest rates. Although her loan amount would be higher if she pays more money in each installment, it would be significantly higher if she maintained a lower interest rate.*

In your report:

* 1. Explain which variables you chose for analyses and why for each model.

**For Linear Regression Modelling:**

* *We ran two linear regression models, in part 1 of the assignment, we used int\_rate as dependent variable and in part 2, we used loan\_amnt as dependent variable.*
* *In part 1, when we used int\_rate as dependent variable, we used revol\_bal,long\_term, mort\_acc, all\_util, annual\_inc, loan\_amount, tot\_rev\_hi\_lim, Grade A to E. We used these variables as they had high correlation with the int\_rate. As we developed the linear model using these variables, we found that the model is significant and has a variability of 97%.*
* *This model has helped us to answer most of the part 1 questions.*
* *In part 2 linear regression model, we used the dependent variable as loan\_amnt and the independent variables are long\_term, installment, int\_rate, A,C,D,E,F as we found that the variability in this model is high at 97% and the model is a significant model. As we did the correlation between the continuous and categorical variables, we found that we have to remove the least correlated variables and we also removed revol\_bal as the correlation with the installment is high. The linear models that we ran with and without variables have differences in their significance levels, and once we removed the variables, our model looks better and helped us answer the part 2 questions.*

**For the Interaction Modelling:**

* *We used the int\_rate, installment, long\_term dummy variable, and the dummy variables for grade A-F, using B as the baseline as instructed in step 3.*
* *We also created and used the int\_install\_int\_rate variable by multiplying installment and interest rate together. We found installment had 93% positive correlation to loan\_amnt. We also found that loan term (long\_term) had a positive correlation of 38%.*
* *We included grade as the question required it for the first part and wanted to stay consistent in what we did for the second part. Although the int\_rate did not show a significant correlation to the dependent variable, we included it in our multiple regression with and without interaction to be consistent.*
* *It was a significant factor in determining the loan\_amnt. We found a significant correlation between revol\_bal and loan\_amnt as well but those two variables had a high correlation to each other, suggesting multicollinearity problems would be present in our model if we included both. We felt it was best not to include both in the final model.*
  1. **Provide the output for each analysis.**

***For the interaction model:*** *We ran our interaction model and transported key information from that report to the 2-way interaction spreadsheet to find the relationship for the increase in both installment and interest rate on the loan amount. We found that as installment amount increased, lower interest rates were approved for higher loan amount than a borrower with a higher interest rate. As the installment amount increase, both high and low interest rate borrowers would get higher loan amount offers but the borrower with the lower interest rate would enjoy a high loan amount offered to them.*

* 1. **Summarize your findings. Clearly outline the impact of your findings for Louise.**

*Looking at the interaction model, I’d urge Louise to reevaluate her installment amount as the standardized coefficients indicates it has the effect on the loan amount that she can receive from the bank. Her loan amount would be 17222.30 considering the joint effects of the installment and interest rate variables. Without interaction her loan amount with our multiple regression model would be 13723.79.*