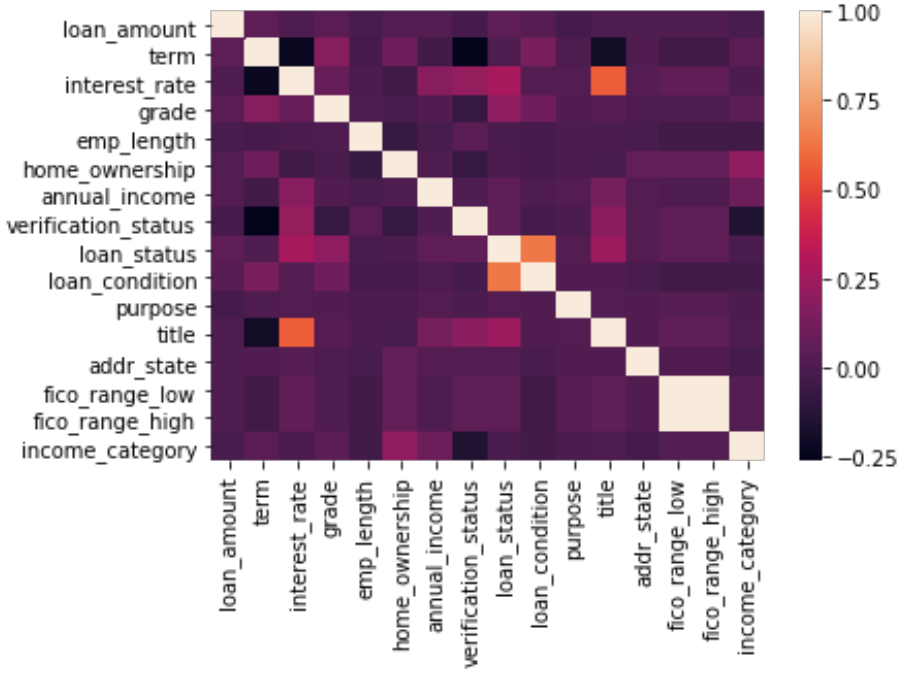
Capstone Project 1: Inferential Statistical Analysis Report

Lending Club Loan Data

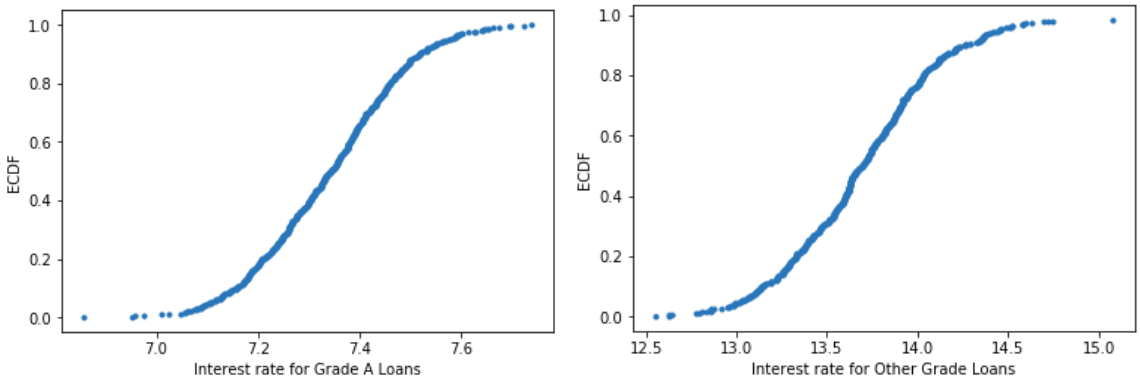
The direction to take the statistical analysis for this project was very clear once the data was obtained and visualised. It was very apparent that there was a positive correlation between the Loan Status and the following features:

- Interest rate
- employment length
- grade
- fico range
- income
- purpose etc.

Also the following heat map was plotted for the correlation among the features.



We also plotted the ecdf for interest rates for A grade loans and Interest rates for the other grades apart from a grade.



This observation made us to perform a Frequentist test on the interest rates.

The test was performed to check whether the interest rates offered for A grade loans were lesser than the other grades hence we did a single tail Welch's t-test as the variance is not equal.

$H0$: The interest rates offered for other grade loans is greater than the A grade loan.

H1: The interest rates offered for other grade loans is not greater than the A grade loan.

On running the test the following results were obtained:

- Ttest_indResult(statistic=324.96013788823643, pvalue=0.0)
- The confidence interval for A grade loans is: (7.339761200000001, 7.327202348671294, 7.352320051328707)
- The confidence interval for all other loan other than A grade loans is: (13.687281466395113, 13.651015426092169, 13.723547506698058)

As the p value was less than 0.05 hence the null hypothesis was rejected .