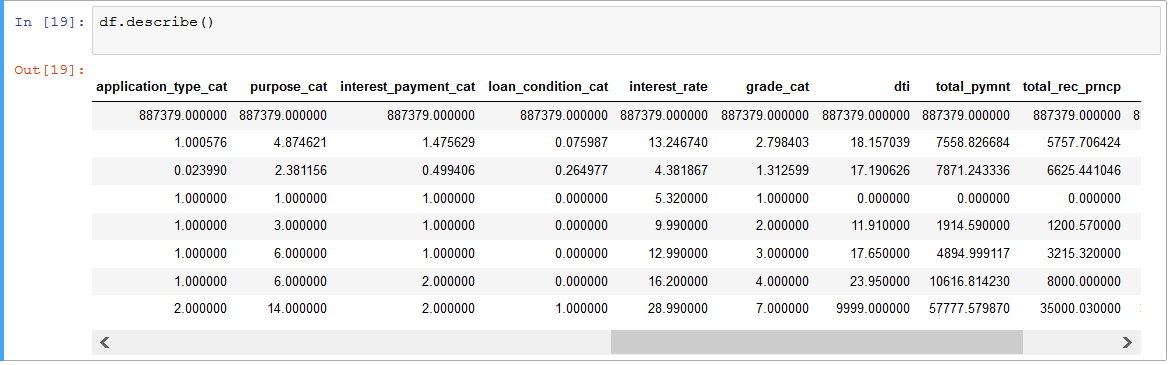
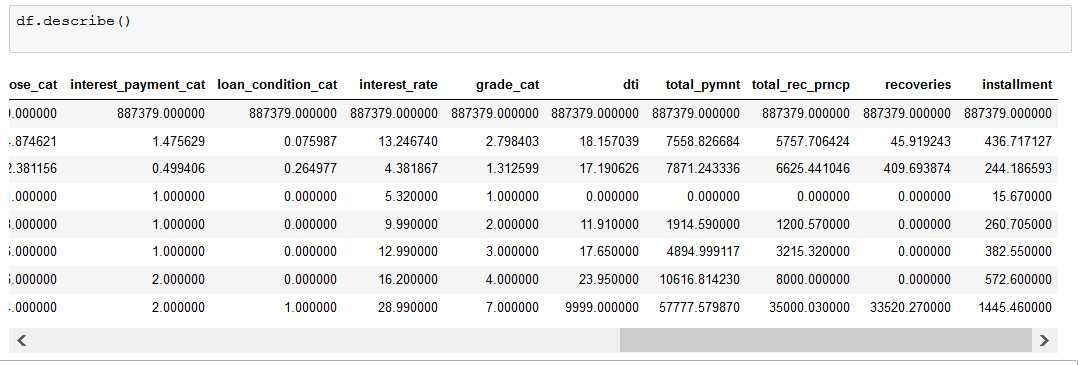
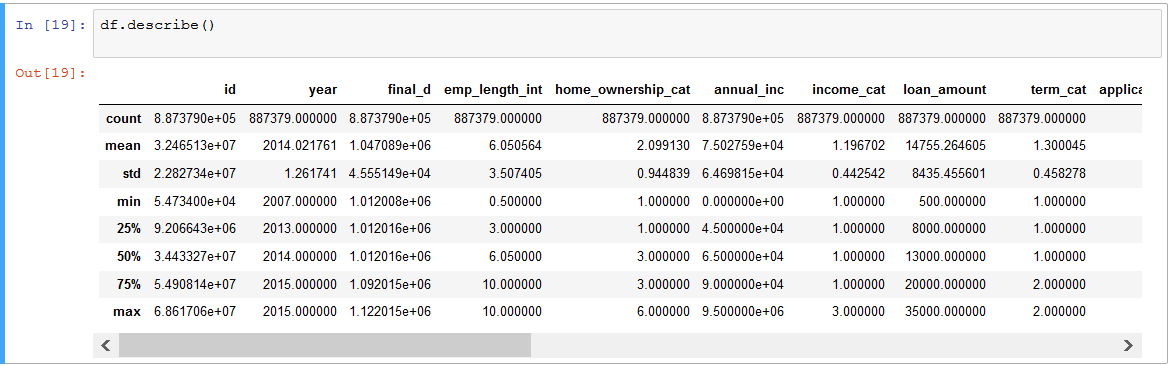
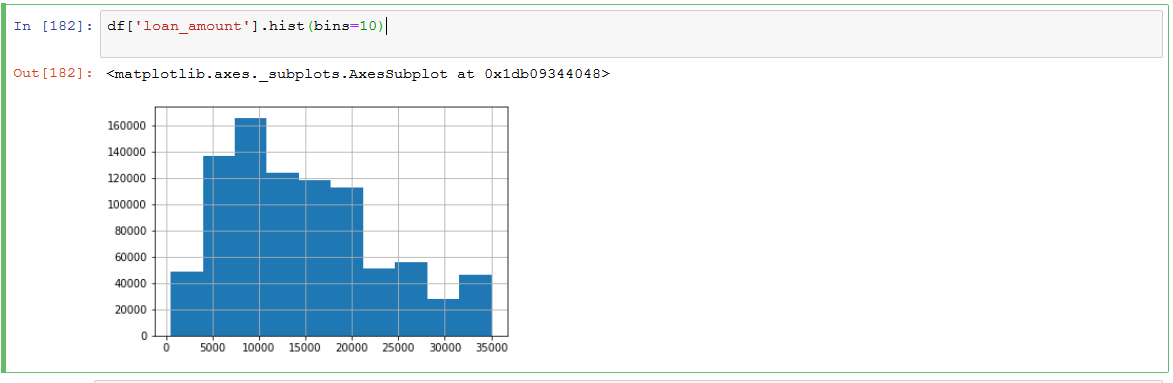
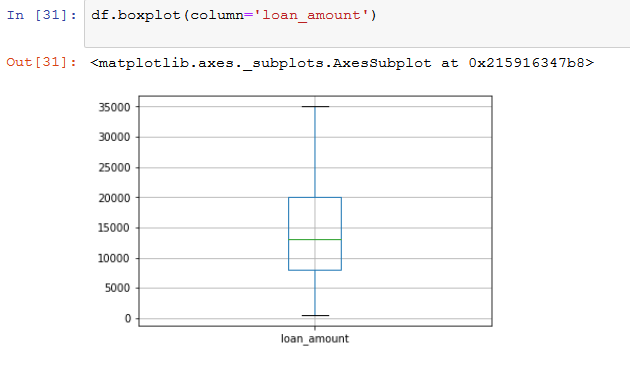
Section A -Sean Wright

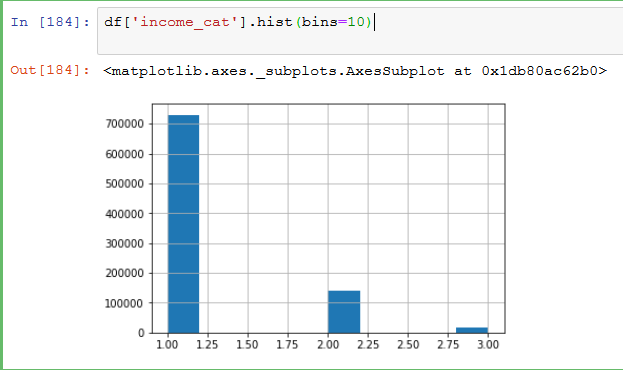
I used the df.describe() function to describe the data set were using this shows count, mean, standard deviation (std), min, quartiles and max in its output, the output of this can be seen below.



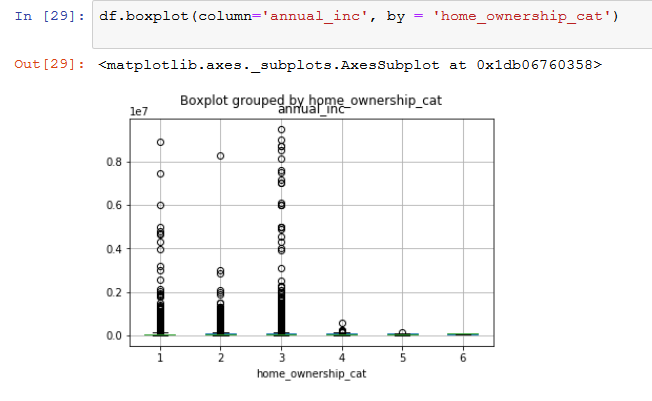
Shows the most frequency of the loan amounts in this data set we can see here that most loans are around the 10000.



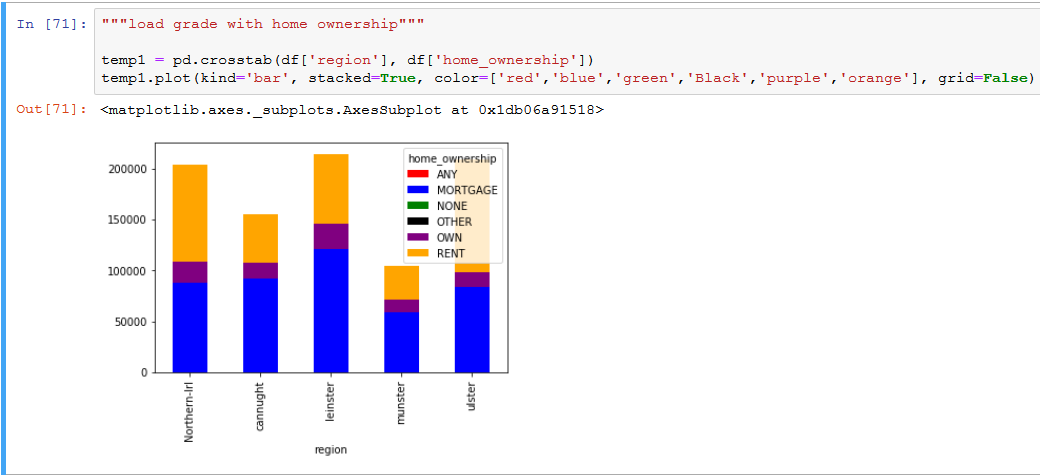
This boxplot reflects the same mean as what we can see in the histogram above also showing the upper and lower bound of the data set.



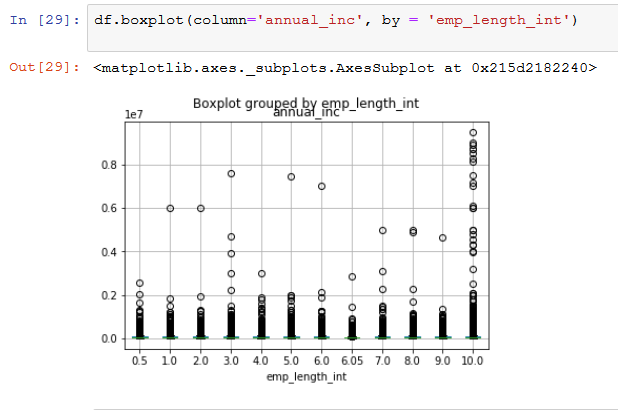
This histogram shows the frequency of the income of the loan holders we can see most are in the low income category .



This boxplot shows the how annual income effects home ownership in the data set.

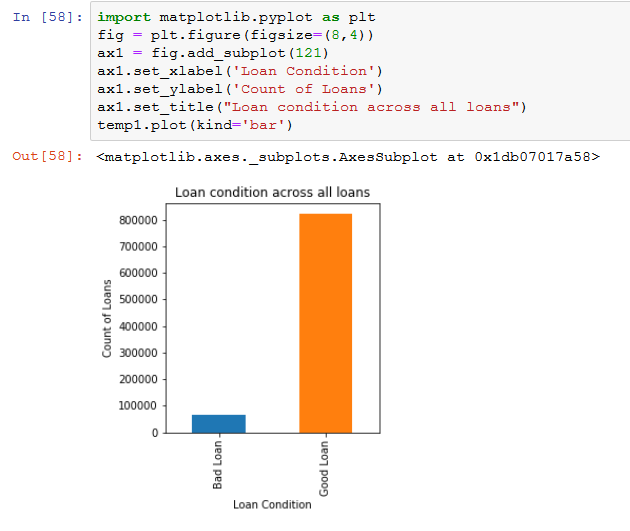


Here I wanted to continue to the description of home ownership, I wanted to examine if there is any correlation between home ownership and the region.

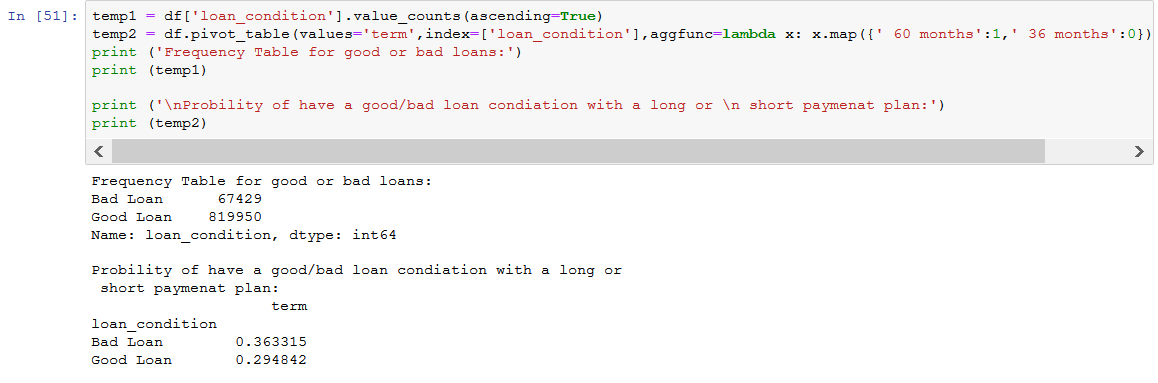


To continue getting some insight in this data set I wanted to examine the correlation between employment lent and annual income this boxplot shows a clear relationship between the two columns.

From here on I wanted to start taking a more in depth look at some of the information in this data set, I choose to look at Loan Condition.

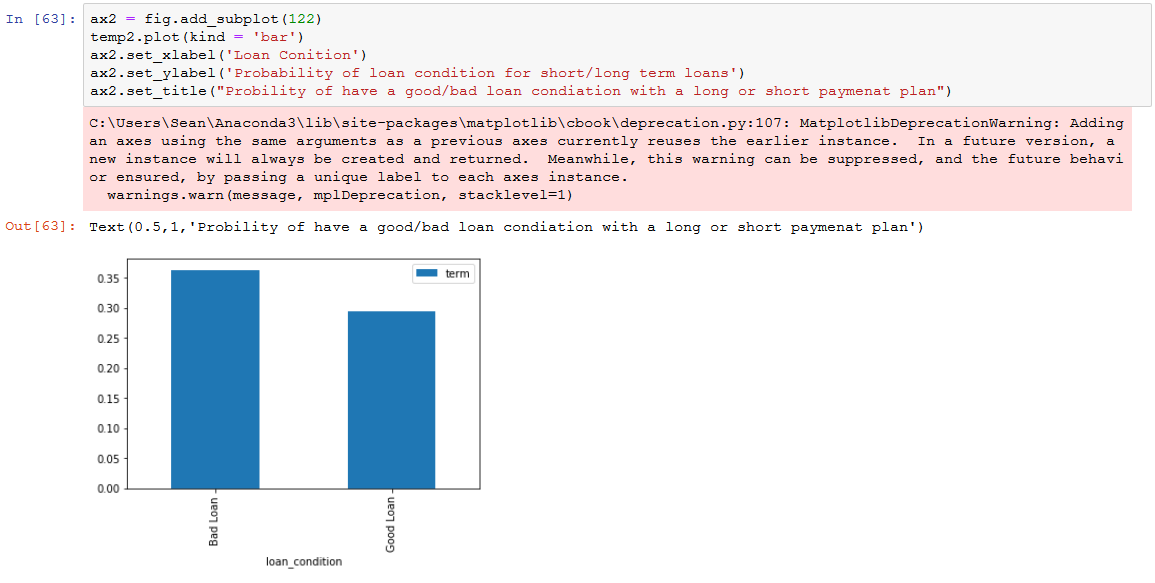


This bar chart shows the frequency of Good and Bad loans



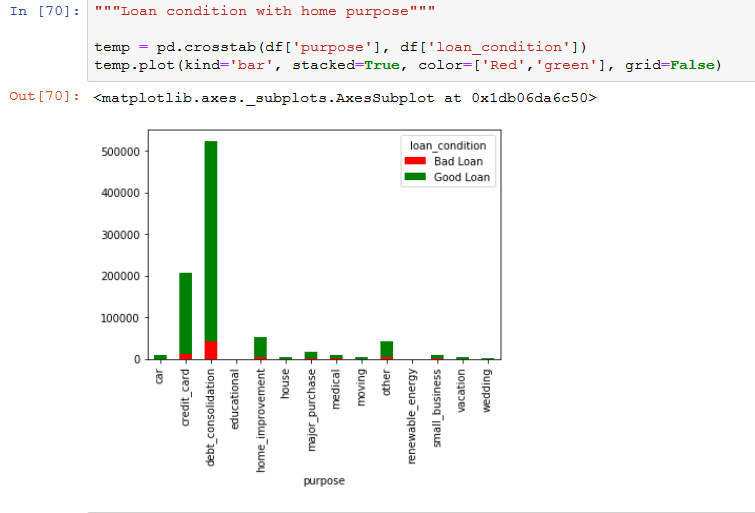
This is a more in-depth description of how a loans conidiation is effected by a loan being long or short term.

This is the above probability in a bar chart.



The last way I described loan condition is with the purpose of the loan to see what most of the loan were taken out for, we can clearly see the most common loan purpose was for debt consolidation.

We can also see that regardless of the purpose we had most loans had a good condition.



Although my main focus was on describing the data set I also worked closely with MJ on the ML algorithm, the result we found most interesting as what the Random Forest Classifier returned when we searched for the best predictors for loan\_condition\_cat, we found the same top five in a different order then when we used these as predictors on the loan condition we both had very high accuracy and high cross-validation score(see below).

