Name: Muley, Tushar

Title: Week 6 Exercise 6-3 Term Project Milestone 1

Date: April 25, 2021

The idea I had was based off credit card applications. Since all those applications ask for marital status, income, whether you rent or own your home and other information. I wanted to understand how much of that is information is correlated to a person being approved for a credit card? In the data set I found hopefully works to create a model that show correlation to whether an individual is low risk (approved for a credit card) or high risk in which case they are not approved.

I did some graphic analysis of the data and below is what I found. The first few where histograms of the data.

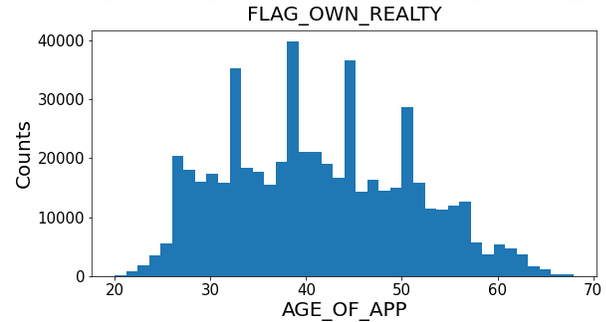


Figure 1

Figure 1 show the age of the applicant and the count of how many there are. There are few spikes in the data. The data is trails off to the ends and I would say is as expected. You have four modes or high points so it is multimodal. Majority of your applicants are between 25 to 55 years of age. You have peaks at different age groups like mid 30’s and late 30’s. You have a peak at mid 40’s and early 50’s. This tells me the histogram of the applicant is platykurtic. Light in the tails and heavy in the middle. Not perfectly distributed.

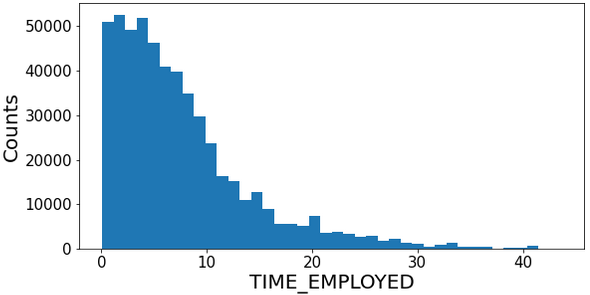


Figure 2

Figure 2 histogram of Time Employed is positively skewed. Telling me people are employed between zero and 10year before applying for a loan. Possible a few indicators of willingness to take on debt after getting a new job or being comfortable in an existing job.

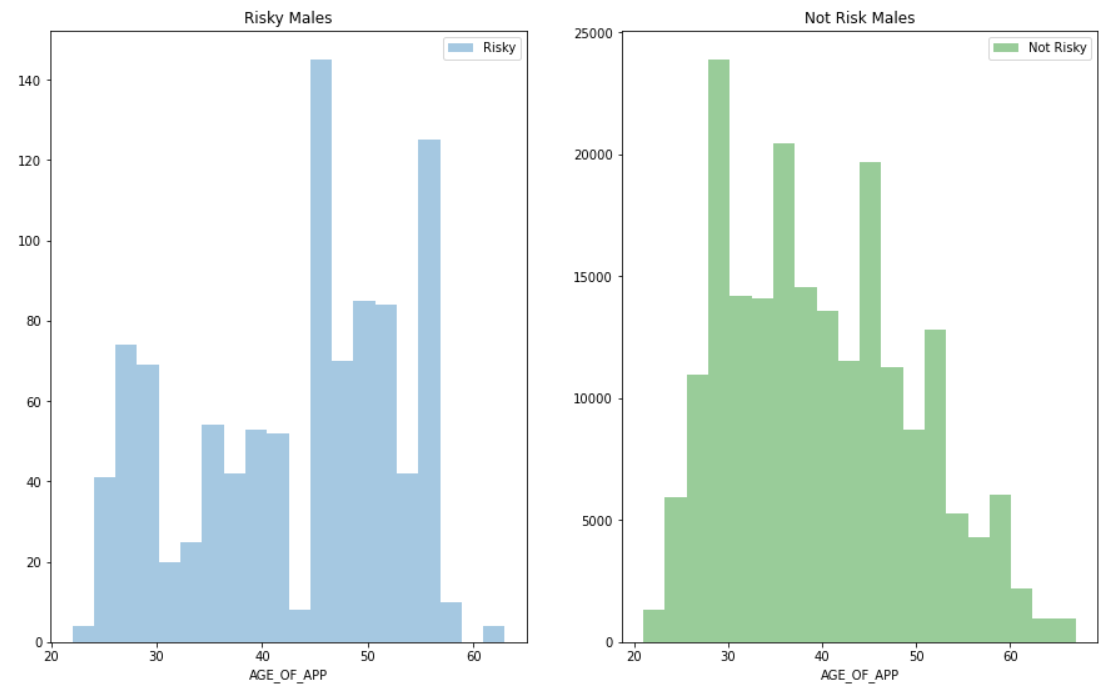


Figure 3

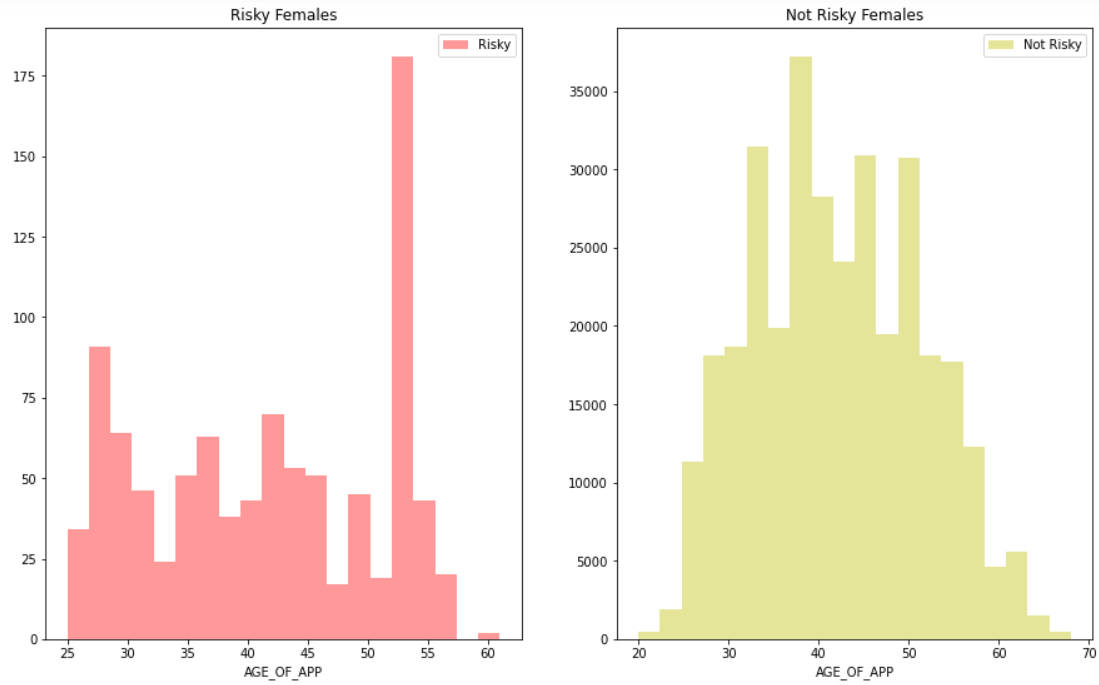


Figure 4

I bring up these four plots. They show more females applying than males. You also have more females showing up less risk compared to their male counter parts. Not sure of what this trend means. I would have thought the data would be equal.

This next plot (Figure 5) shows income levels by male and female. It might relate to why females have a higher request for credit compared to the males gender.

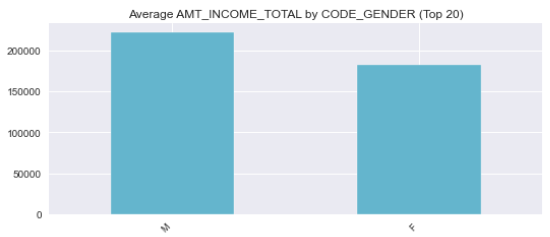


Figure 5

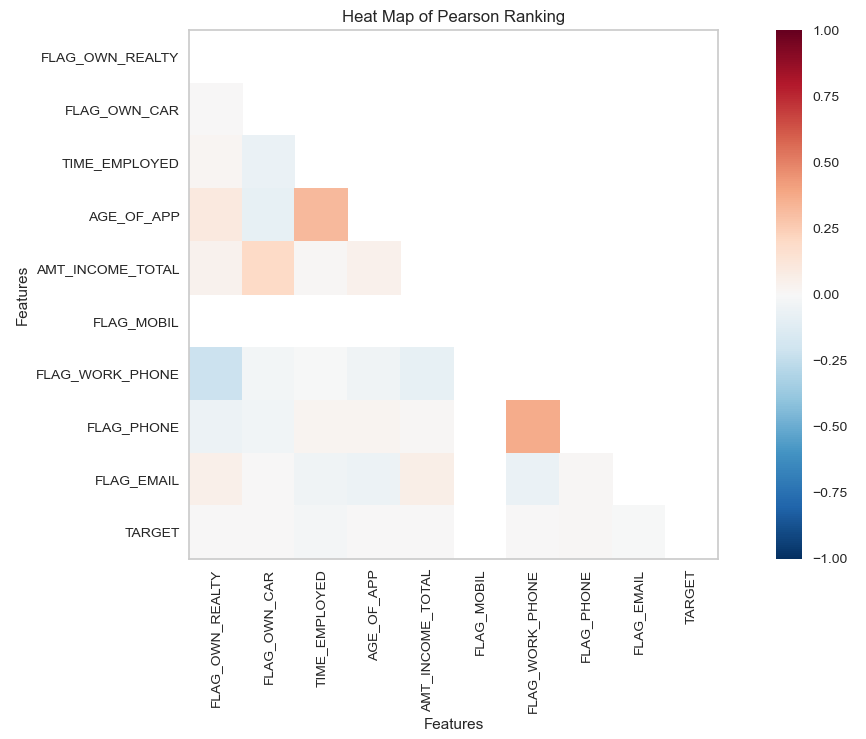


Figure 6



Figure 7

Figure 6 and Figure 7 shows a couple of heat maps. Figure 6 is Pearson’s ranking of select attributes, while Figure 7 is a correlation of all attributes available. Here Age and Time Employed are highly correlated. Total Income has a stronger correlation to car ownership compared to real estate ownership. The Target is the risk given to those applicants. Not a lot of correlation here between the other variables. But combination of these might play a big part in credit approval.

After reviewing the data, it seems these things on their own do not have any real correlation to each other outside of the normal things you would assume are associated like amount of income to age of applicant. It leads me to believe all the additional information asked for on an application do play a part in the model to determine approval of the individual for credit. I’ll need to figure out how these individual parts are correlated.