Karen Alvarado
Module 5 – Task 2

Lessons Learned Report

Did you learn anything of potential business value from this analysis?

Yes, actually there are a lot of items that represent potential business value; the tools explored in this task are both interesting and easy to use, which means they are also easy to understand from a third-party perspective. The graphs are very dynamic, and changes can be done easily to the data; for example, creating bins, creating new columns, and combining data.

In general, the data frame we were given is very easy to understand as for the demographic information, the payment information is also understandable, but we focused in the demographics as to understand if there are patterns that would explain the increasing default rates. We were able to identify and visualize this information through graphs that can be seen in the .ipynb files.

What are the main lessons you've learned from this experience?

The main lessons I learned is that this tool is versatile, to say the least, there are many ways of doing the same thing, but the good part is that there's always a way. For instance, I wanted to split the Age column into bins so it would be easier for me to graph and understand trends, there wasn't an explicit way in the task, but through a quick search I could find a way to do it and it worked perfectly.

I also learned that is very important to understand your data before doing any analysis since that will make your life easier, you can just input the data without having to go back and forth to the explanation sheet or to the excel file.

Finally, I learned that risk is a very complicated topic which needs to be analyzed carefully and in detail.

What recommendations would you give to the Guido regarding your findings?

I found some very interesting patterns, for starters, just in this data set we see a lot more women than men, which sometimes draw the results (negative and positive) towards women; we would need to assess if, in reality, women have more credit than men and if that affects the default rate and why.

In average the main clients are around 34 and 35 years of age; which is the most productive age of people; we would likely say that people within this range of age are more likely to get credit in order to buy homes, cars, and other material things that are expected from them at that age.

One of the distribution graphs show that the largest loans (in terms of money lent) are given to people within 41 and 60 years of age, this is expected as those individuals are probably very stable professionally and in high positions in the hierarchy, so we would recommend to focus more in that segment.

Finally, we performed an analysis based on age, marital status and default rate, and there's correlation between divorced people and a positive default; therefore, we would recommend gathering more data in that specific segment to verify this preliminary result. In general terms it is logical that divorced people will have a reduction in income as they'd have to split their belongings and their monthly income would be just theirs now; but we want to search for more information to validate this hypothesis.