Course 2 Task 2 Findings

Bailey Keel

Something of potential Business value that I learned is the odds of someone defaulting correspond strongly with their payment history. To look into this data, I used the payment records of September 2005. From analyzing the default rates, I found that the people most likely to default are those with their payment being delayed 7 months, followed by 3 months and 2 months. These results can be found in Figure 1 below. On the X axis below -2 is that they don't have a consumption record, -1 is they have paid their bill in full, 0 is they have a revolving credit and from 1 onwards that is the number months their payment was late by. For payments 2 months late and onward there is at least a 50% default rate, though one thing that should be noted, and I also found very interesting, is that on payments that are between 4 and 6 months late there is a decrease in chance of defaulting. This seems to go against one's assumption when first looking at the data. Also, should be noted that for data after 2 on the X axis there are less users in those categories. This means the results might not be as accurate as for the -2 to 2 range. I would recommend collecting more data in these ranges to allow for more accurate results. Recommendations I would give to the Data Science team is to make sure to look at your findings via a couple different methods and graphs. When first stumbling across these findings I looked at a density plot, this showed that the highest default rates were for people that have revolving credit. What I came to learn was that this really showed me that the most people defaulted with revolving credit because there were 3-5X as many users in this category compared to the others. Only after digging into those findings further did I realize my mistake of taking absolute defaults for the odds of defaulting that I came to find Figure 1.
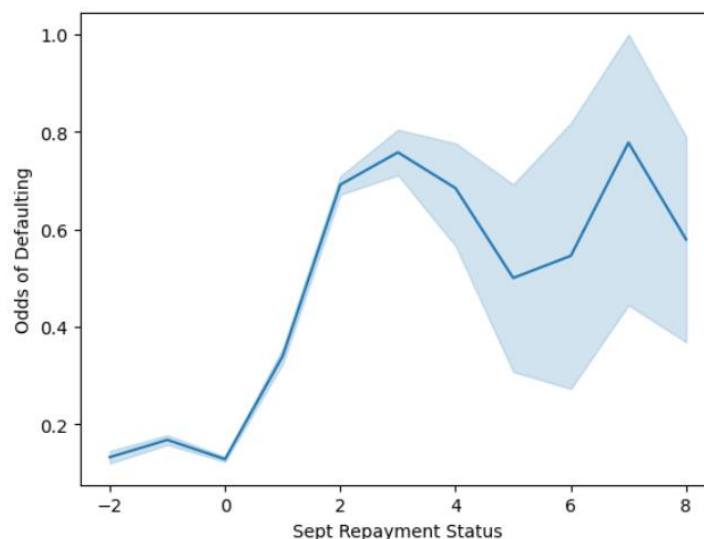

Figure 1: Odds of Defaulting based on Sept 2005 Repayment Status