

Lachlan Sinclair

2/2/2020

Week 4 Assignment 2

Reflections

- a. The margins on these two groups seems razor thin, most of the values in the categorical factor's graphs were within 1% of each other. The categorical factors I initially thought would be strong indicators of making a prediction were not significantly leaning to one side. Some even favored the opposing side for certain traits, I had expected people actively looking into changing would leave more frequently but that is not the case. Overall the graphs do follow the patterns expected, for example the customers who stayed were more frequently satisfied and less frequently unsatisfied than the ones who left. As the prompt mentions, leaving and staying are two very different outcomes so I had expected far larger margins in these graphs. But when you are dealing with problems such as churn it makes sense that these two groups would be closely related otherwise it would not be an ongoing problem.
- b. I think condensed representations like this could help data scientists narrow down what properties and relationships they should be focusing their development on. Essentially theses representations could guide the early stages of creating predictive models that use larger datasets. While it wasn't stark in our use case, modeling like this can help point out the important traits that are more useful for making accurate predictions or reveal some unexpected patterns. It also seems reasonable that representations such as these could identify potential issues that would cause problems for a predictive model built in a specific way. Compact descriptive representations like this can also be extremely useful when trying to explain what the plan is for the predictive model to a non-data scientists, these graphs are easily accessible to everyone and are quite presentable.