Lessons Learned Report

1. Learning R overall has been a difficult and pretty daunting task. Trying to use multiple different models for the same data without using caret proved to be pretty difficult and produced many errors. Dummifying the data also seemed to complicate things, as it produced several additional variables that were hard to account for when creating the models. The most troublesome portion of R is troubleshooting, though. It is very difficult to troubleshoot R if you do not know what a certain error message means or if you do not know what you’re looking for. Something that worked really well and seemed to make most of the tasks easier was using the Caret package. It simplifies several tasks and functions within R and makes building models easier.
2. Some advice for future data scientists tackling a similar project would be to really get to know the Caret package and learn how to apply it to the project.
3. Lessons Learned:
   1. **Task 1- Getting Started with R.**
      1. Preprocessing is important- Without proper preprocessing, the results may come out skewed or misrepresented.
      2. Practice training models using outside datasets- It really helps to have practice training models on several different data sets. This will help get different perspectives on any issues that may happen while training a model.
      3. Don’t be afraid to play around with the data- Changing the raw data is not proper. However, changing the way you process the data and tweaking the data during preprocessing is important to maximizing the model and making sure of not overfitting the model.
   2. **Task 2- Predicting Customer Brand Preference.**
      1. Take time to understand the data- Understanding the data completely will help with knowing what to do to preprocess the data and get it ready for training.
      2. Caret makes things easier- Caret is a package that will make certain processes such as training a model more convenient.
      3. Test the data on different models- Testing the data on different models will help the understanding of which model will work best for the task at hand.
   3. **Task 3- Multiple Regression in R**
      1. Visualization helps- Using plots, charts and histograms for the data or certain aspects of the data can help to find patterns and correlations that can be used to analyze it.
      2. Output data sets and predictions- Making output files of the data sets and predictions can help with building the final report.
   4. **Task 4- Discovering Association Between Products.**
      1. Use as many different ways to analyze the data as possible- There are many ways to analyze data. From visualization (plots, histograms, etc.), to using functions such as summary() and inspect(). All of these will give good insight into the data and using different methods will get you different perspectives and results.