Question: Which Data Science Predictive Modeling techniques are the most suitable to answer the questions given in the project assignment?

To select a model, we must first understand what factors help us determine whether or not a model is suitable for the given data and questions. Some of the factors include the size, quality, and nature of the data, the number of computational resources available to run analysis, and what you actually want to do with the data once it’s been analyzed.

After knowing this, we decided to look into some of the potential modeling options to help us answer questions about which demographics factors appear to affect customer engagement the most. Knowing we are working with big data primarily related to shoppers and their attributes, along with the fact that we have a limited number of computational resources available, we decided that the best models for analyzing this data would be:

1. Logistic Regression: We can use this binary classification method to compare different demographic attributes to overall customer engagement and see which attributes hold a high correlation to customer engagement

2. K-Nearest Neighbors: We also believe this classification algorithm would be a good selection because it’s a great way to find from a group of attributes which are most correlated to each other. This could be a great way to analyze patterns across the entire dataset.