So far, there are only 53 responses to the survey, so we are still waiting for more responses in order to have more data to run the model on. There also are about 7 subpar responses (unfinished, obviously botched, etc.) so they will be excluded as well. Once we gather all of the survey data, then we will be able to pin the sales data with the survey response data based on the client’s name. This is why no preliminary running of the model has been done.

There are three main types of variables that will be used in the attribution model. First, are the variables of the types of communication : email, phone call, virtual meetings, in person meetings, small social events, large social events, and formal business meetings. Second, are the variables of the frequency of these events : never, daily, weekly, monthly, quarterly, twice a year and yearly. The third type of “variable” are the open-ended questions that will be dummy coded using themes from ChatGPT for whether or not the theme was included in the response. There also are questions in the survey about the types of product training which will include variables: no training, hands-on training, distributor’s office training, contractor’s office training, large group, small group, and virtual training. I will initially run the models with those, but they most likely will not have an impact on sales performance.

The plan is to run logistic regression and other such models on the data to see which variables are most important in predicting sales performance. With more data, I will be able to see if the research question can be answered or if there is not a clear relationship. I guess we will see.