ISM 370
Jalyn Buthman

Business Analytics Project 2

The bank client dataset offers data on the characteristics of the bank's clients and whether or not they have subscribed to a new deposit account. Within the dataset there are far more clients who are subscribed compared to those that are not subscribed(Figure 1). We can see from this disparity that there is opportunity for the bank to get more of its existing clients subscribed.

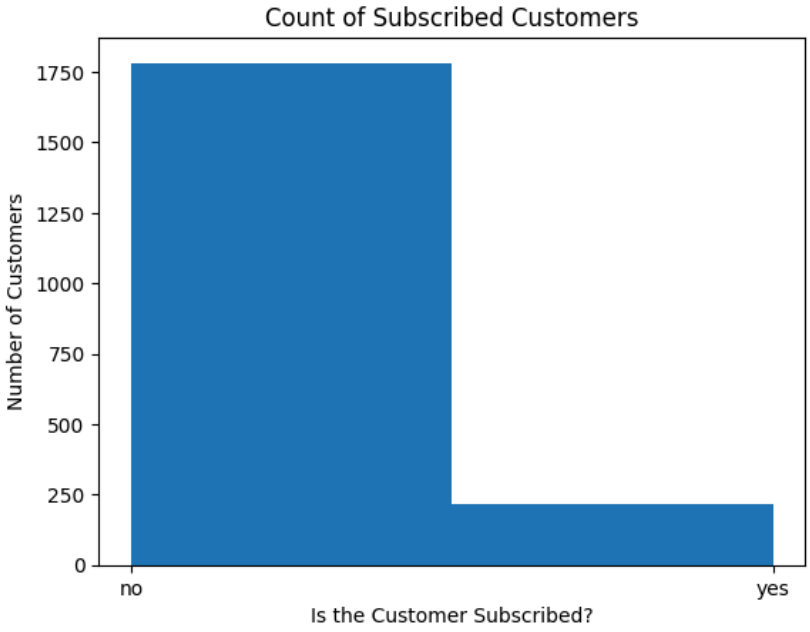

Figure 1

The dataset shows that the average age of subscribed clients is lower than that of non-subscribed clients(Figure 2). The median age of subscribed clients is 36 years old, while the median age of non-subscribed clients is 38 years old. This could indicate that the bank needs to focus more of their subscription messaging towards their older clients in order to raise the number of subscribed clients. Meanwhile, the younger demographic should be targeted less, as there are more likely to already be subscribed.

Figure 2

To test if there was a significant difference in age between the two groups I used the following hypotheses:

H1: There is a significant difference in the age of subscribed clients compared to that of non-subscribed clients.
H0: There is no significant difference in the age of subscribed clients compared to that of non-subscribed clients.

I conducted a t-test for the difference in the age between subscribed clients and non-subscribed clients. The resulting p value was 0.69, which is significant as this test used a significance level of 0.05. In conclusion, we can accept the null hypothesis and conclude that there is not a significant difference in the age of subscribed clients compared to non-subscribed clients.

In order to predict whether or not clients would be likely to subscribe I created a model that uses the number of times that the client was contacted in previous campaigns along with the number of times that the client has been contacted during the current campaign as independent variables, and the client's subscription status as the outcome variable. With these variables the model has a precision of 0.925 and a recall of 0.541. This model could be used to test hypothetical current campaign contact counts for various previous contact counts to see how many contacts are needed before the client is predicted to subscribe.

In conclusion, I would suggest that the bank check how many contacts from their current campaign are needed by testing hypothetical numbers along with real previous contact count numbers. From this testing they can gather the ideal number of contacts that are needed for each client based on how many times they were contacted in the last campaign. This will help the bank to increase the likelihood of their client's subscribing, as well as help save time and money by only increasing the number of contacts for the client's that are more likely to respond.