Banco Federal de Finanças

Marketing Campain Analysis

presented by

Joaquin Solis

Scott Townsend

Zach Holcomb

Isaac Weyland

I. The Data

When reviewing the meeting notes from the clients, there were a lot of questions asked about the data in general. Before reviewing the model we created, we want to respond to some of those questions.

Clientele Groups

One question that was brought up was about single customers. "Should we only call single people on Saturdays?"
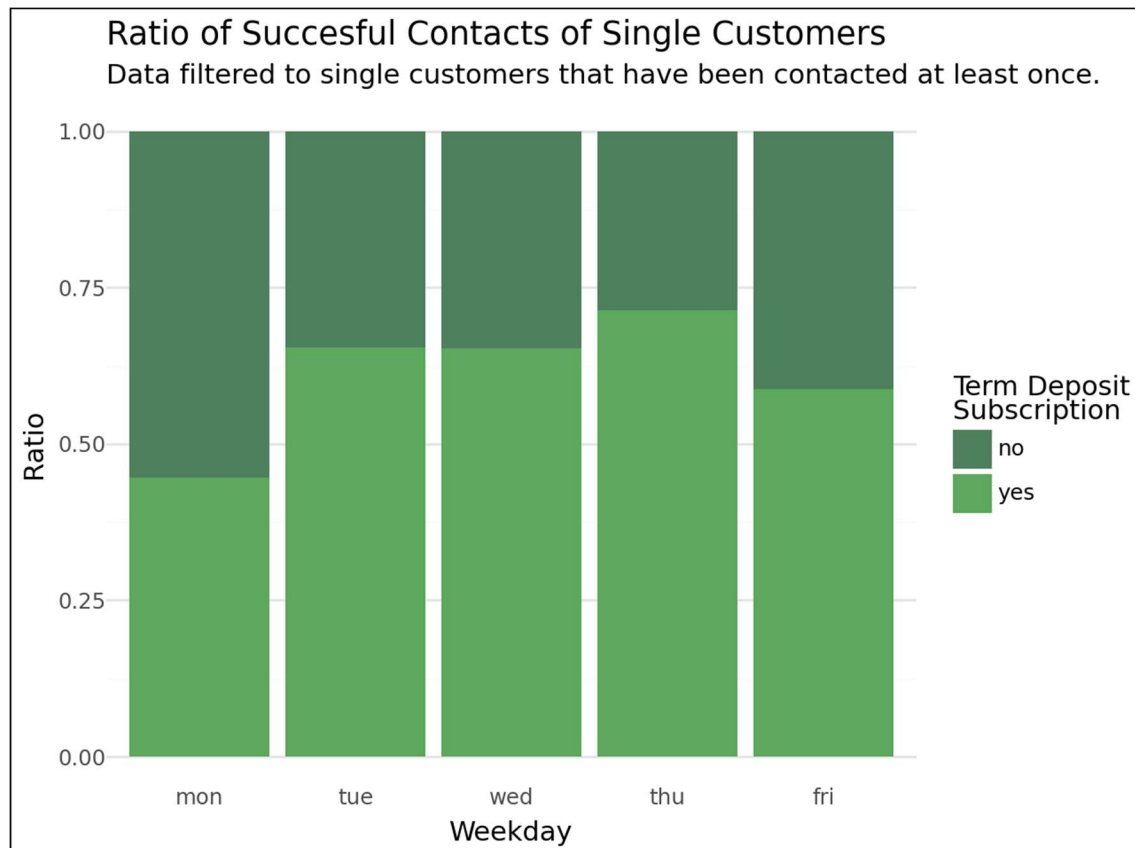
Hopefully no one is being contacted on Saturdays since it appears the call center doesn't work then. However, the data does give some insight in this area. Calling single clients on Monday is sure to guarantee lower subscription rates than if you were to call during the rest of the week.

Another question that was mentioned was if the bank should continue investing time and manpower into their student customers.

Students are some of the best clients despite being the smallest client group.
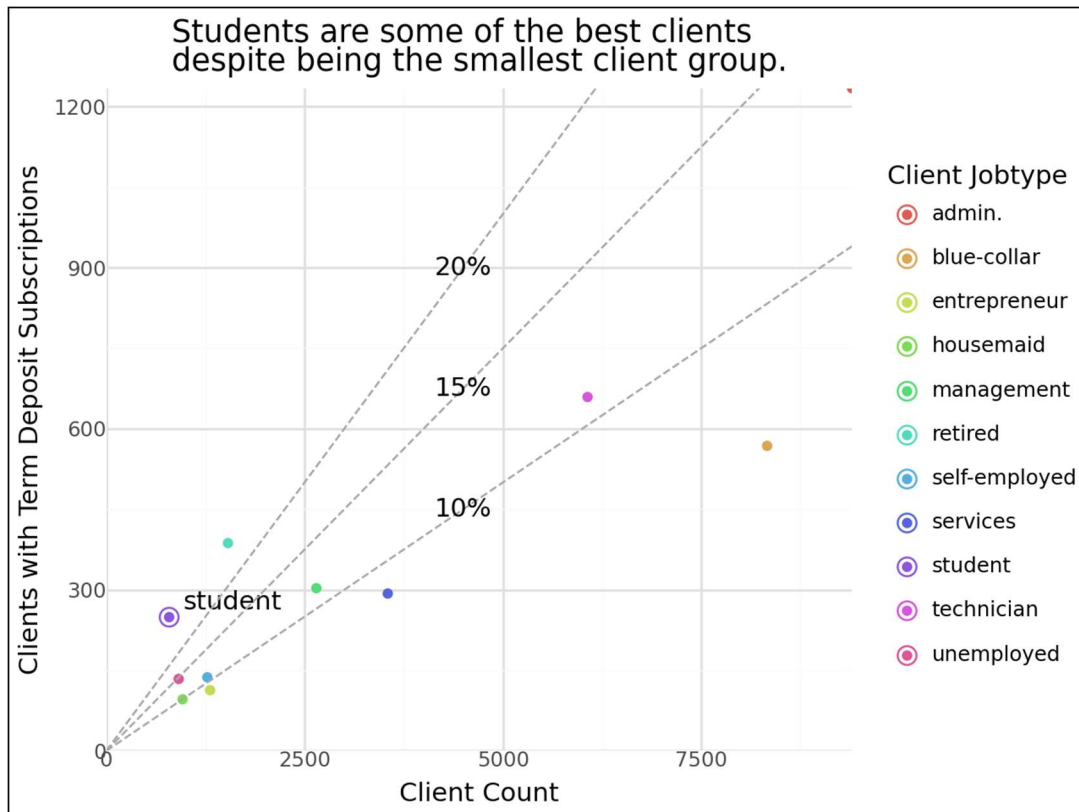
Although students account for the smallest customer group, the subscription rate of students is much higher than normal. This means that students actually hold more term deposit subscriptions than other job types like entrepreneurs, admin or retirees.

Effective Calling Practices

One crucial question that was asked was if calling clients too frequently negatively impacted success rates.

Success Rate vs. Contact Frequency
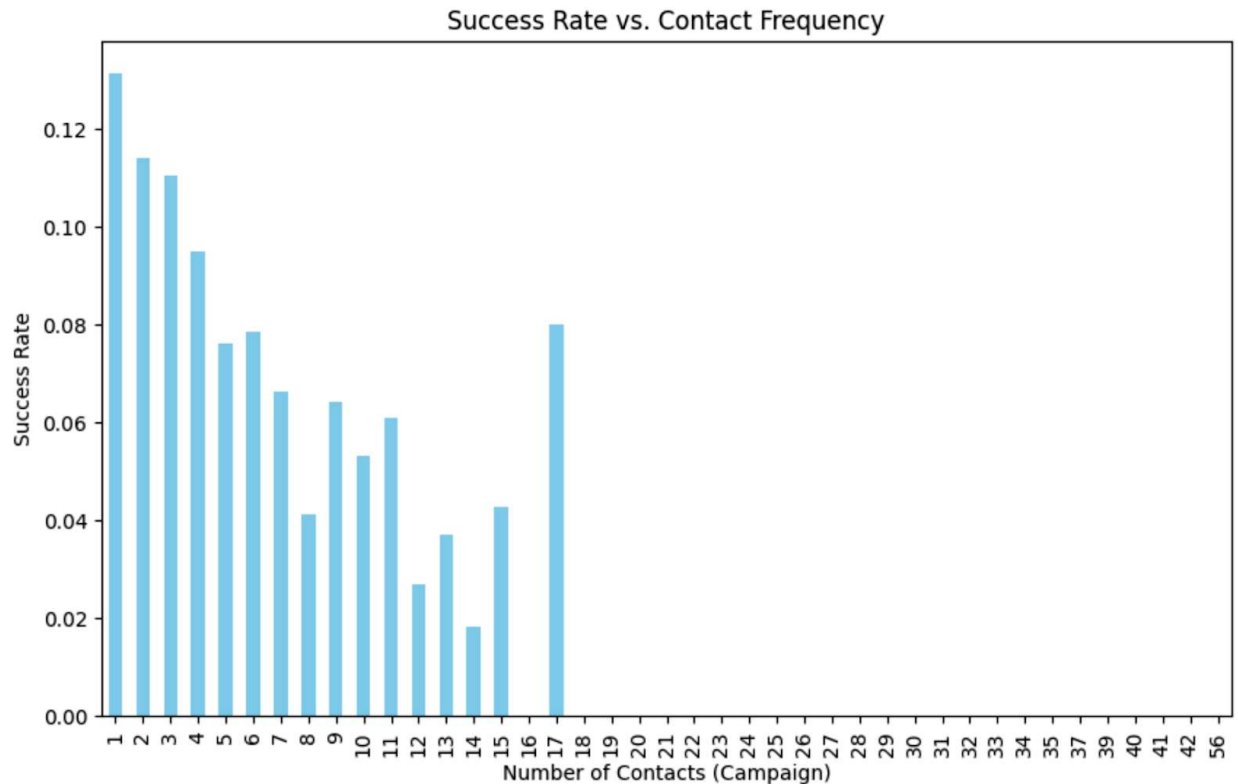
As the plot shows, success rate drops dramatically after the first several calls, dipping to half at around the 8 call mark. This suggests that repeated calls is not as effective as other strategies like seeking new customers.

II. ML Model

We trained our model on most of the data we were given after taking a sample out to test our model's accuracy. After training, we are able to predict if a customer will subscribe with 82% accuracy. The most difficult part for our model is when a customer *does* subscribe. This is largely due to the lack of data regarding subscribed customers vs unsubscribed.

Attached is the code necessary to generate a working model. Based on the training data, the best use for the model would be to shorten your call list and avoid waisting calls.

III. Python Notebooks

Below are Github Gist links to the notebooks we used during this case study:

https://colab.research.google.com/gist/holcombzv/35dd8eac5471d64d7dd573869c8246e9/w04m2.ipynb

IV. Discussion Responses

I. Supervised or Unsupervised?

We recommend using a supervised approach, as we have access to data related to our target variable so we might as well use it. Using this data set will allow us to create a predictive model and reach our target outcome.

II. Training Data Split

Using an 80/20 split would be most appropriate for this analysis. Allocating a large portion of our data towards training and testing while reserving the rest for validation should make our model more accurate. The target variable is very simple to use and interpret, so keeping a simple solution for the data split should work well for our needs.

III. Additional Insights

We believe that this data set will provide opportunities for further insights beyond the current objective of the analysis. We can explore additional features and relationships within the data and gain a deeper understanding of factors influencing the problem. There is plenty of data that we can use to give extra detail about various aspects of the problem.

IV. GDPR Requirements

In accordance with GDPR, the use of anonymous data is exempt from regulation, so we can proceed without needing any further compliance with GDPR regulations.