Banco Federal de Finanças

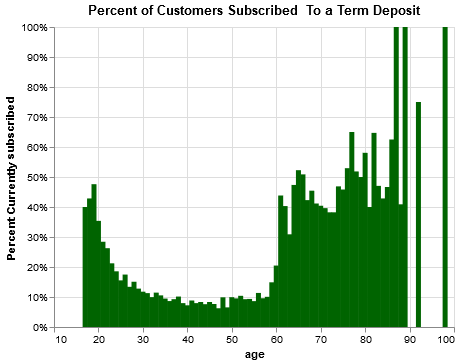
Marketing Campaign Analysis

## presented by

Jordan Carlson

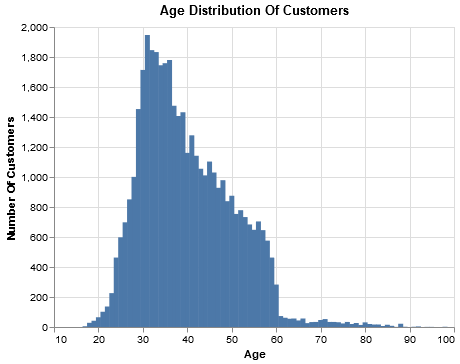
Cody Overholt

Kyle Mueller

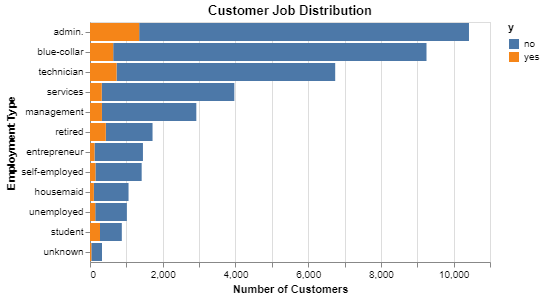
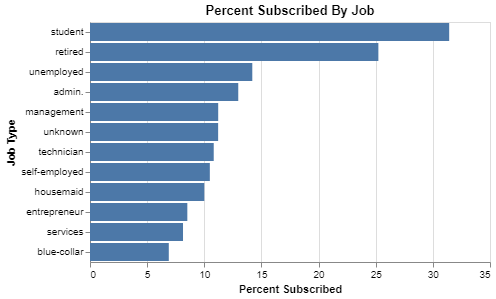
1. Target Demographic
   1. Age

Roughly 45% of people 25 and younger and 50% of people 60 and older subscribe to a term deposit. It is important to note that the population of people 25 and younger is greater than those 60 and older.

Target Demographic - Age (continued)

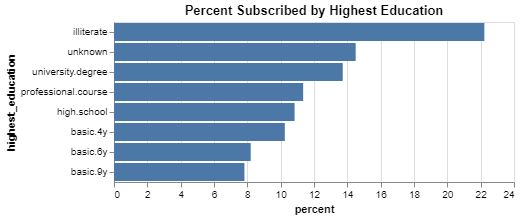
The largest demographic of customers are also the majority of customers who haven’t signed up for a term deposit and noted in previous graphs.

* 1. Job

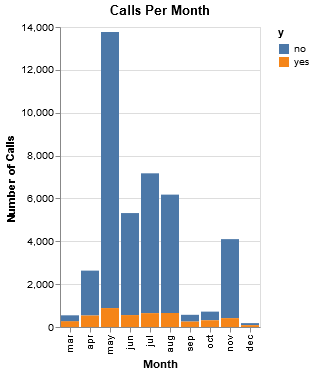
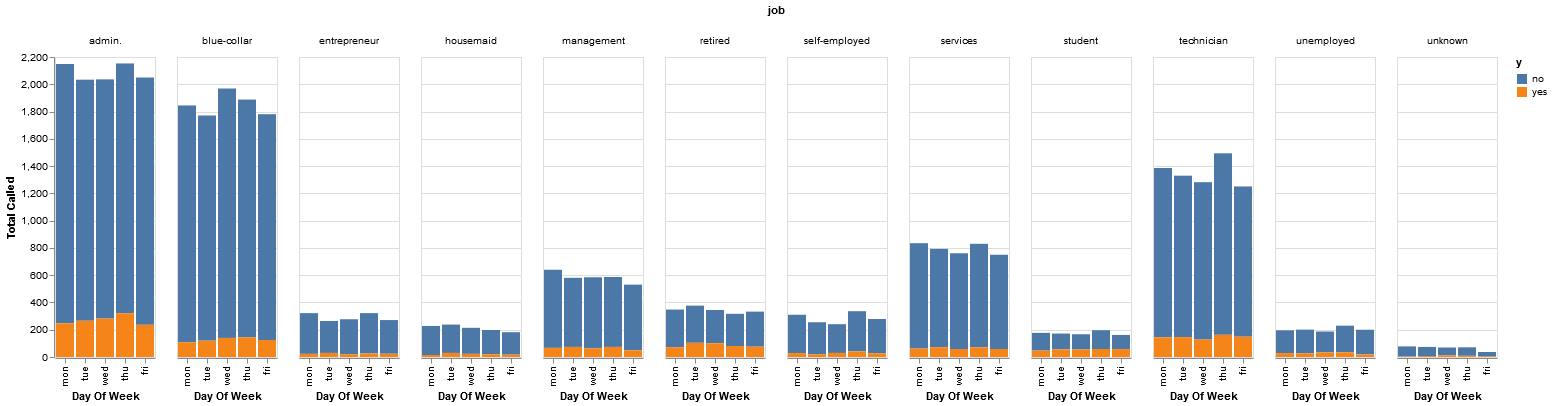


30% of student customers and 25% of retired customers subscribed to a term deposit. This is a major demographic to focus on in relation to the rest of the customer demographics.

* 1. Education

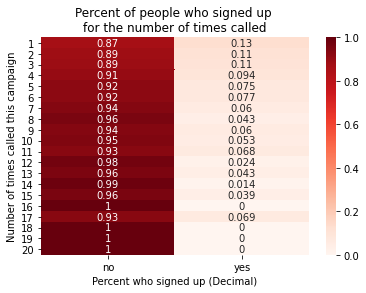
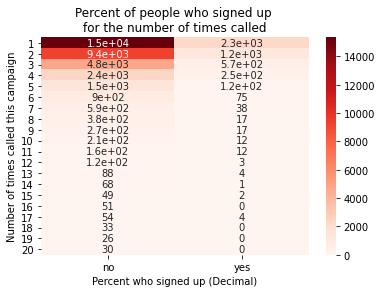
While customers from both the illiterate and unknown categories have the highest percentages in the bar chart above, they are represented by miniscule amounts of data and are thus outliers (unreliable data). However, customers with a university degree have a high amount of data and have a high probability of signing up for a term deposit. They should be the target audience when contacting customers.

1. Actionable Patterns

When It comes to contacting customers, the two most populated professions appear to be more willing to sign up later in the week as opposed to earlier in the week. In order to tap into those customer bases, it may be worth trying to contact customers later in the week.

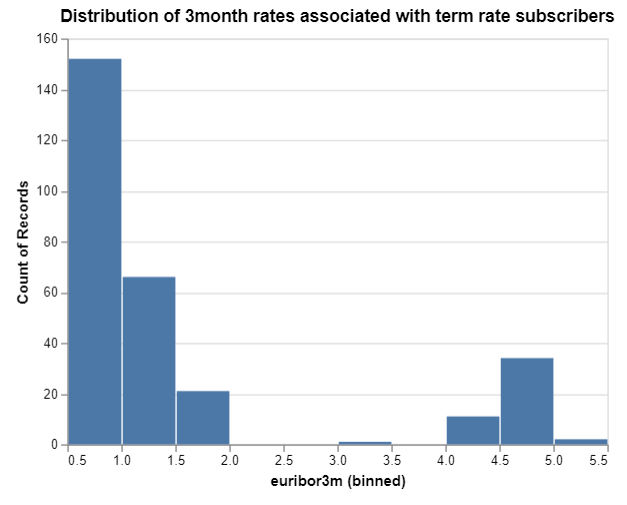
Based on the above graph, there seems to be a big difference between the number of calls going out each month. The total number of new subscribers remains relatively consistent month to month, except where there were significantly fewer calls going out. One reason this could be happening is because certain customers are receiving more calls than others.

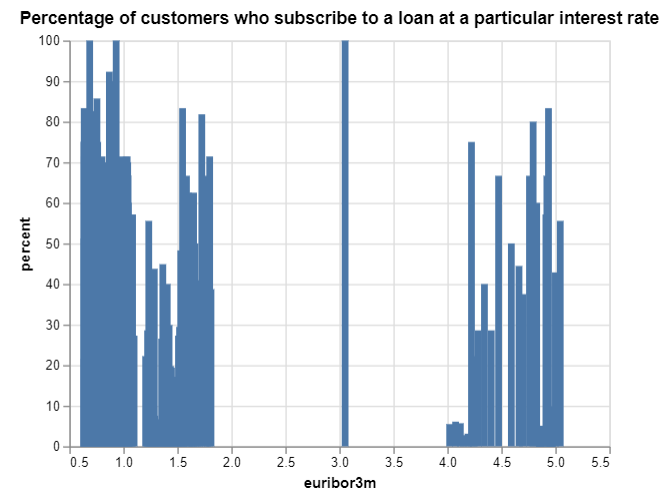
People can become annoyed if they are contacted too frequently, turning them away from subscribing to a term deposit. Based on our analysis, contacting people about our services only remains effective for so long.



Calling people about products seems to remain effective for the first 5 calls, at which point it begins to be less productive than contacting new customers. After around 17 calls, contacting ceases to be effective at all.

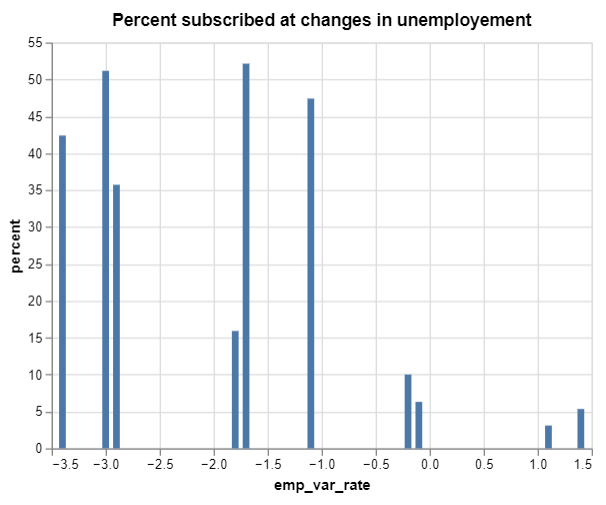
1. Economic Indicators
   1. Low rates

The majority of customers who subscribe to a term deposit subscribe during times where 3 month rates are relatively low.

Low Rates (continued)

Although the majority of customers subscribe when the interest rates are low, there are also some customers who are likely to subscribe to a term deposit at mid to high interest rates (i.e. 3.0, 4.3, and 5.0). However, the population density for those rates is much lower.

* 1. Unemployment vs term deposit subscription percent

When unemployment rates decrease, we see that the percentage of those who subscribe to a term deposit is high. Therefore, when the economy is doing well, we should expect an increase in term deposit subscribers.

1. Model
   1. Algorithm

The model uses supervised learning to discern which people are more likely to sign up for a term deposit. The algorithm uses all features available as X and the “y” feature as y. The algorithm uses oversampling to improve the accuracy of the model. The model uses a train size of 75% and a test size of 25%.

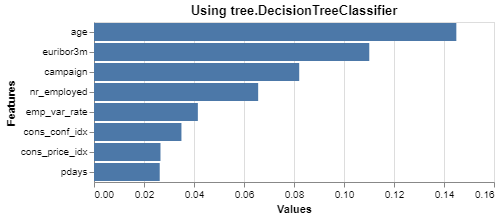
* 1. Accuracy

The model has an accuracy score of over 96%, which is incredibly high. The model has this level of accuracy due to oversampling and using the RandomForestClassifier from the sklearn.ensemble library. To further improve the accuracy of the model, more data (and more diverse data) would be needed.

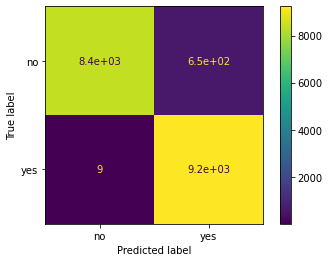
* 1. Decision Tree

The model makes use of the RandomForestClassifier from the sklearn.ensemble library. This offers higher accuracy than the RandomTreeClassifier.

* 1. Feature Importance

Below is a feature importance graph. The most important features for this model to accurately predict which people will sign up for a term deposit are their age, the interest rate (euribor3m), and how many times they are contacted.

* 1. Confusion Matrix

Below is an image of the confusion matrix for the model. As you can see from the matrix, the model is very accurate, only having 9 false negatives and approximately 650 false positives out of a testing dataset of 10297 people.

Note: The model can be found at the bottom of [this](https://colab.research.google.com/drive/17AiwEKeX62XeX-Jkfn9aoJj4wCqux2PY?usp=sharing) jupyter notebook. Using the last two blocks of code, you can persist and load the model.

Conclusion

We have created a model that is accurate for the data that was provided. According to our model, customers that are most likely to subscribe to a term deposit are:

* Customers younger than 25
* Customers older than 60
* Customers with a University Degree
* Customers that are students
* Customers that are retired

According to the data we collected, you should try to contact more customers at the end of the week, and you should avoid contacting customers more than 5 times. The consumer confidence index is not a very important feature according to our model, so you should focus more on the important features (age, interest rate, # times contacted, unemployment rate).

While the model is very accurate for the data provided, we are unsure whether it will remain accurate when more data is introduced. To create a model with a higher degree of accuracy, we will need more data and more diverse data. Much of the data collected is from people between the ages of 25 and 60. To verify that the model is accurate we will need more data from customers younger than 25 and customers older than 60. Let’s see how accurate the model is in your testing, and we can continue to improve the model as you gather more data. Thank you!

1. Python Notebooks

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

This file was the main file we used, but it was too large for a gist:

<https://colab.research.google.com/drive/17AiwEKeX62XeX-Jkfn9aoJj4wCqux2PY?usp=sharing>

This file was also too big for a gist:

<https://colab.research.google.com/drive/1hDQhbWBQiZM1rF_kmVju2lTIYeQ9DZHC?usp=sharing>

<https://gist.github.com/mueller14003/67868918840f7b9f67b6b0db48488621>