Reddic Housing LLC

Marketing Campaign Machine Learning Analysis Report

## Presented By:

Becca Roeth

Ethan Houser

Kyle Bell

Kyle Parks

Makayla Whipple

Chantal Borchardt

# I. Introduction Summary

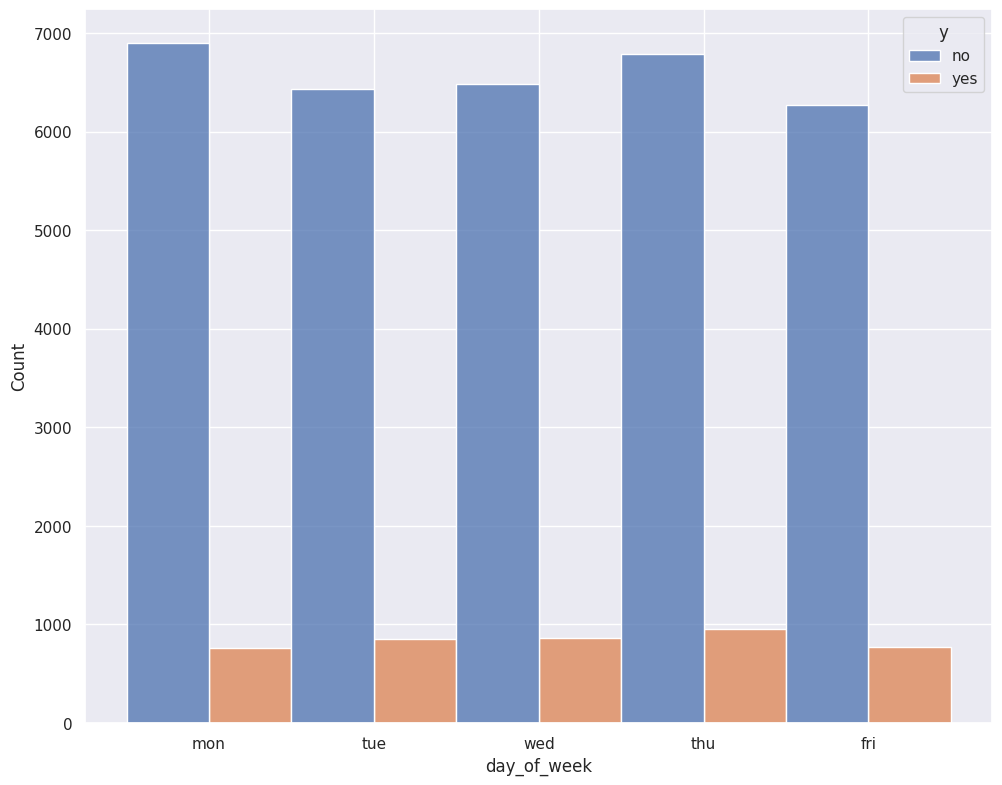
Our model helps Banco Federal de Finanças (BFF) by focusing their calls towards those that will more likely subscribe to a *term deposit*. However, there is a trade off that needs to happen between calling people that decline the offer, and not calling people that would have accepted the offer. Comparing the cost benefit of this trade off, it is safe to assume that it is better to call someone that declines, than to not call someone that would have accepted. This is because it is more expensive for the bank to not call someone that would have accepted, than it is to call someone that declined. So this model that we have created allows for more calls to people that decline, rather than not calling those that would have accepted.

# II. Questions

1. *The dataset has approximately 37,000 records. How much of that data will you use to train your model?*

For a 70:15:15 split (70% of total records is training data, 15% testing data, and 15% validation data), we are using 25,900 of the records for training and 5,550 for training and validation each.

1. *What do you think we might need to do for this project in order to be compliant with GDPR regulations?*

The GDPR doesn’t apply in this situation since we are only building a model and not selling the data. If we decide that we want to use the data under GDPR then we would have to get consent from the customer involved in the data. 

1. *Do some days of the week produce better results? Do some types of customers respond better on certain days than others?*

There is an even distribution of customers that said yes and customers that said no for each day of the week.

1. *Does contacting people too frequently for these marketing campaigns have an adverse effect on the outcome?*

While it may seem cold calling is effective, it has a 9% likelihood of a positive call (3,322 positive calls / 35,696 total calls in no previous calls) but repeat calling has an 11% likelihood of a positive call (4,208 positive calls / 37,069 of total calls in less than 27 days).

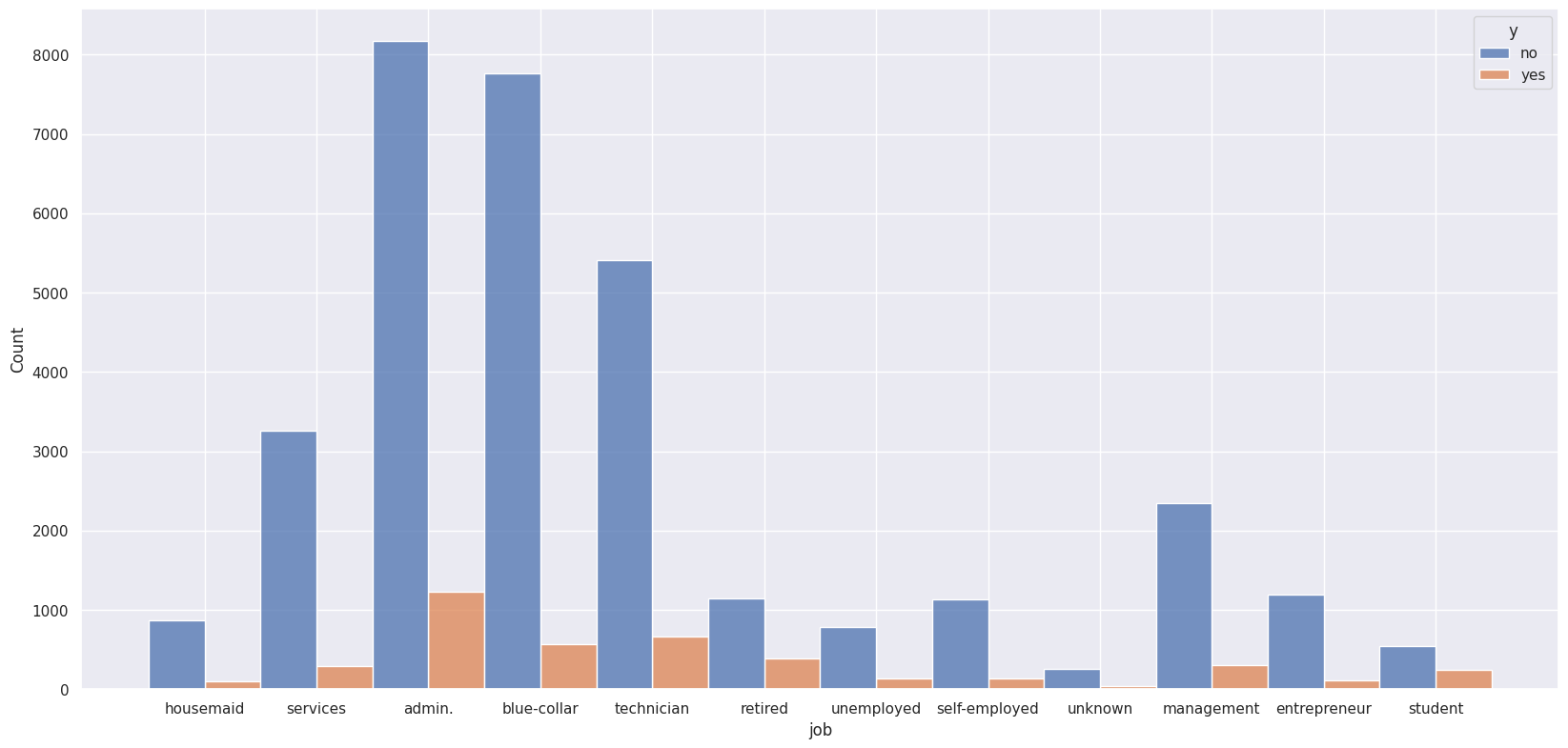
1. *Do you think a supervised or unsupervised approach would work best for this situation?*

A supervised approach would be the best type of marketing approach for the company to take following into the future. Several types of data analysis shown in this paper support this conclusion. Marketing strategy until now has lacked targeting, and the data shows that it is costing the company much more money than it is generating.

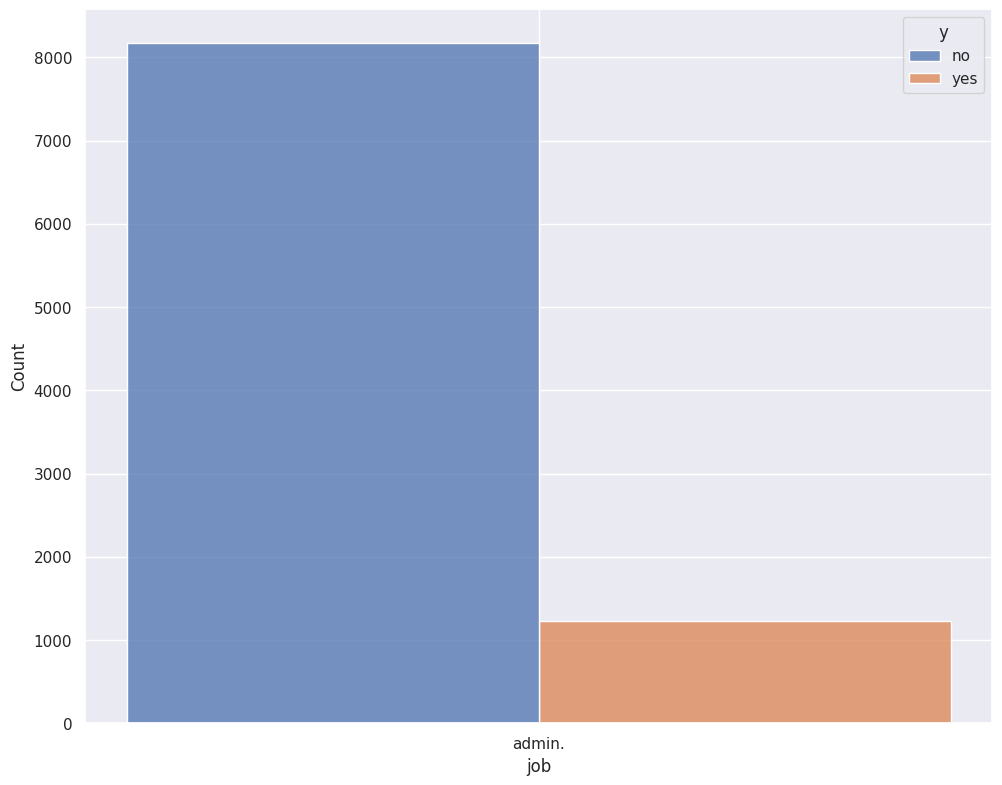
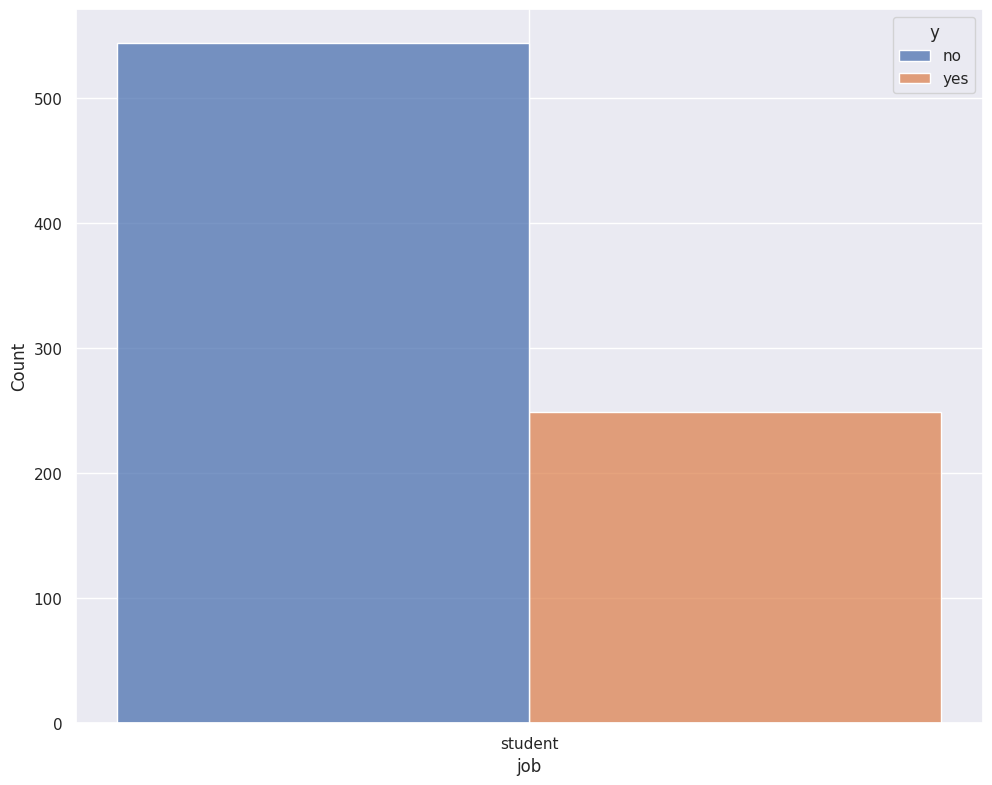
# III. Data Exploration

## Issues with the Data

The more you call a group of people overall, the more they subscribe to a term deposit.

For example, the job someone holds affects whether they say “yes” or “no”, and this graph implies that those in administration should be called since there were more of them who said “yes”.

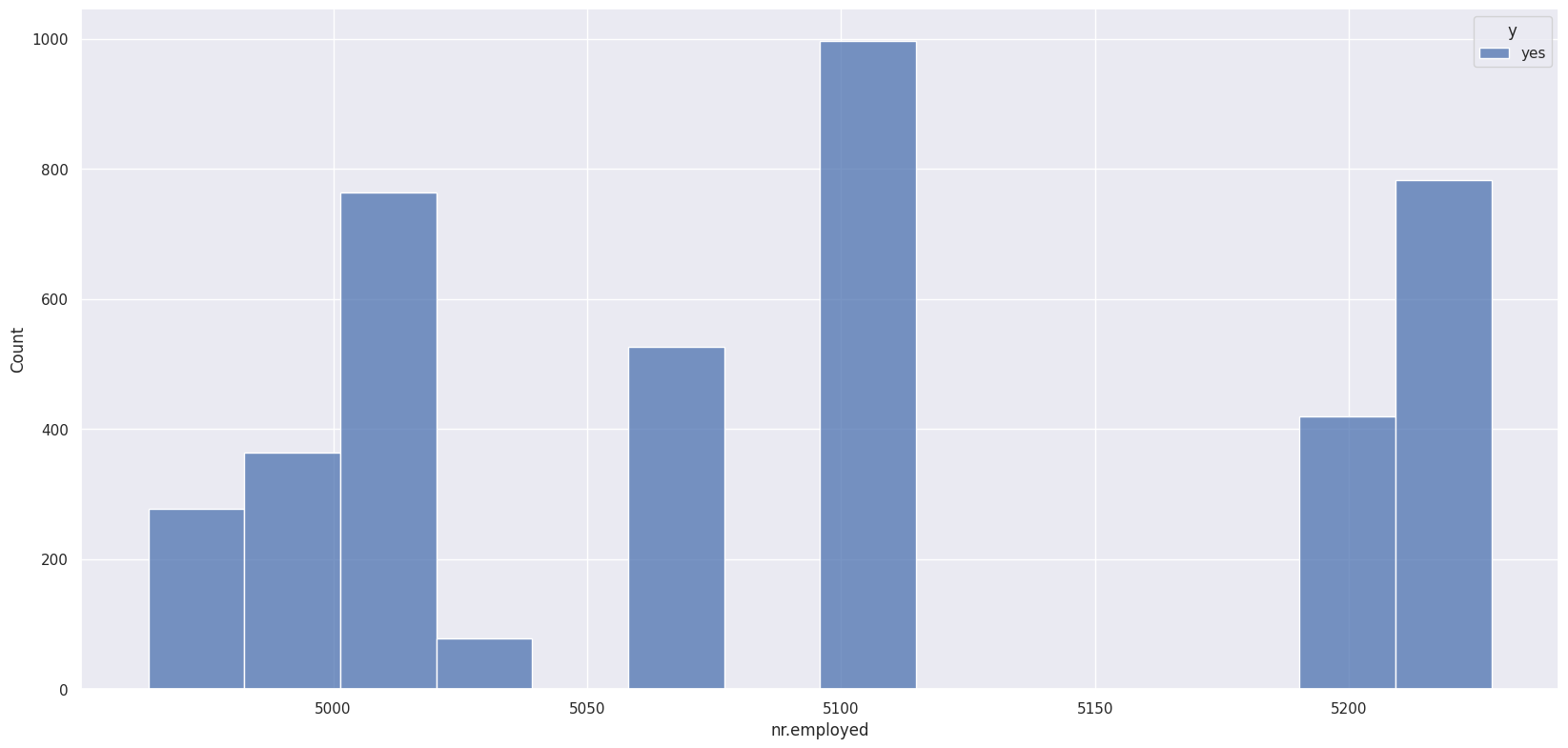
When adjusted for the number of people called, however, students have the highest ratio of subscriptions compared to administrators.

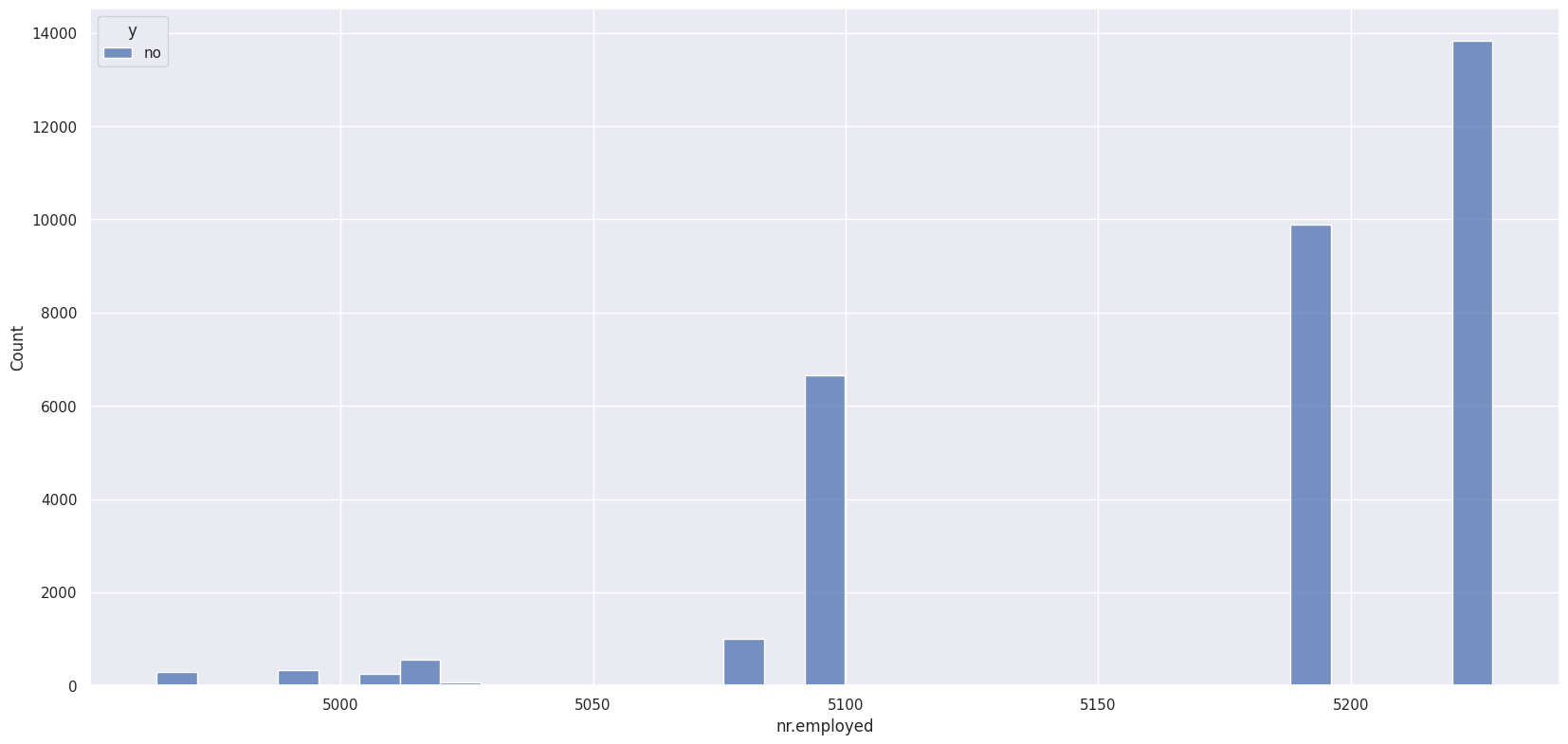


So if students were called more (and if the data scaled properly in the real world), they would make more money since more of them are saying “yes” compared to how many are saying “no”. However, that is assuming that the students who were called can represent the total population trend of all students, which we do not have a strong sample size for.

Currently, calling administrators may have produced more money even though the ratio (money earned vs money lost) is bad simply because there were more administrators called.

## Data Features Chosen

With this in mind, features like nr.employed (the number of people employed at BFF) are valuable because, -while the rate of those that said “yes” is relatively balanced for the amount of people who were employed:

-the rate of those that said “no” is not as balanced:

When there were more employees at BFF to make calls, more calls were made, but there was a worse ratio of “yes” answers and “no” answers.

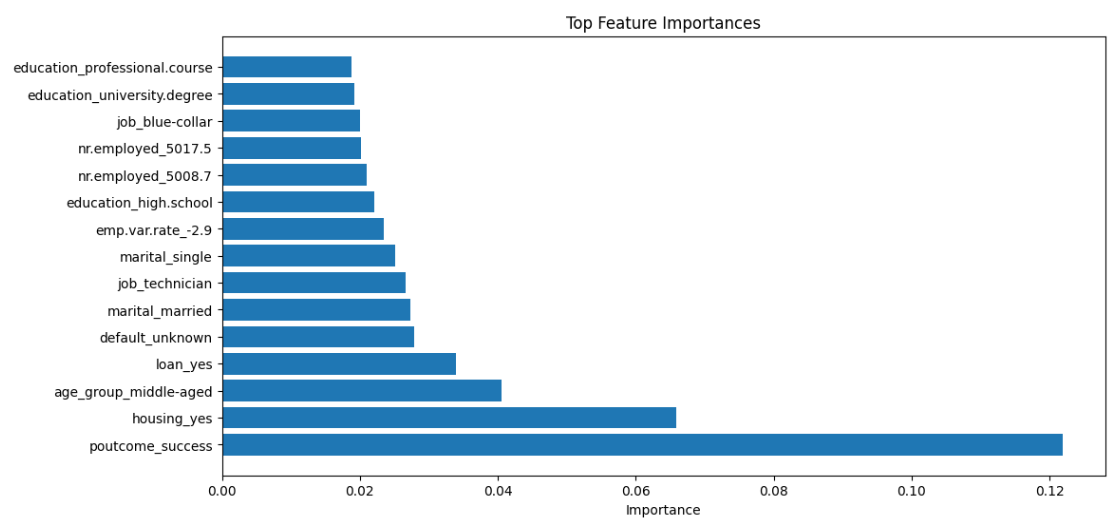
**NOTE**: We are not recommending that you fire people to reduce the difference between “yes” and “no” answers, we are pointing out that there is a correlation between employee numbers and good ratios. When the acceptance rate increases the number of employees hired can be increased.

To explore this specific instance of less employees-higher income ratio more, we should look at why there would be low employee rates.

# IV. Machine Learning Model

## Feature Importance

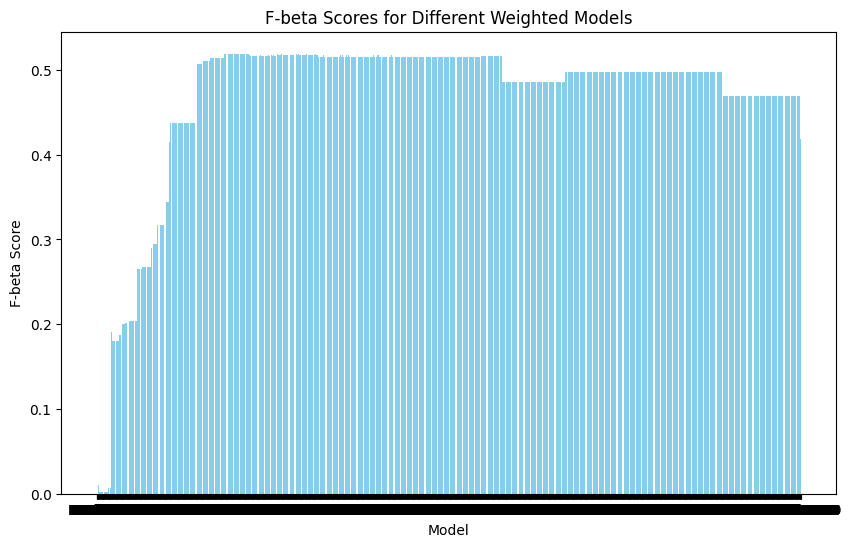
When it comes to the feature importance, we were able to train the data using the Decision Tree Classifier. The graph below demonstrates the significance of each feature and how it corresponds to the target value (whether the person subscribed). Through these graphs and the many different tests we performed we were able to conclude that the best features for our model were the features weeks, nr.employed, cons.pride.idx, emp.var.rate, and cons.conf.idx.



## Feature Engineering

Besides One Hot Encoding our chosen features, the only feature engineering we did was the “pdays” column. This column tells us how many days have passed since the last time that we contacted them. We converted this column into how many weeks have passed since we contacted them, and if we haven’t contacted them, set the value to -1.

We used a decision tree model. This gave us a higher recall score than the random forest classifier. We decided to emphasize recall because we would rather call someone that says no, then to not call someone that would have said yes. In order to put more emphasis on recall rather than precision, we needed to tell the classifier which to weight.



The F-beta score provides a single value that captures the trade-off between precision and recall, allowing you to evaluate the overall performance of a binary classification model based on your specific requirements and priorities. By adjusting the beta parameter, you can emphasize the importance of either precision or recall depending on the context of your problem.

This chart shows that the higher the f-beta score, the more cost efficient our model will be. We later checked this with the cost of making calls, and adjusted our decision tree to reflect that decision by changing how much "weight" we gave our target value.

# V. Results, Action Items, and Limitations

Based off these assumptions:

**Time on Call** = 30 seconds

**Minimum Wage** is $6.50 and typical teller wage is $11

**Call Cost** = wage\*time\_on\_call

**Average Savings** = $4,960

**Percent in Term Deposit** = 75%

**Net Interest Margin** = .012

We can use the following equation to calculate the total amount earned:

**Positive call benefit** = average savings\*percent in term deposit\*net interest margin

**Total earned** = incorrect calls\*call cost + correct calls\*call cost + correct calls\*positive call benefit

The end results of our model can help the bank make approximately $11,000.

Future marketing campaigns need to be more targeted. The data from the last one shows that the range of customer base was too general, and the data gathered overgeneralized the types of jobs potential customers had. This has created a situation where the Data Science team can tell where the last advertising efforts went wrong, but be limited in what type of feedback that can be effectively given. Continuing efforts need to be more mindful in the type of data being gathered.

The current customer profile needs to be revised. Current data shows that out of all the professions recorded students and retirees overall have the best “yes” to “no” ratios. It would benefit the bank if future telemarketing campaigns call more students and retired persons as a start. Focusing on calling the right type of customer first would fix a lot of the current problems with marketing campaigns.

# VI. Python Notebooks

[](https://colab.research.google.com/drive/1FS-vc4RSav5xalC42u9Ce9HClWcNSFmE?usp=sharing)

* <https://colab.research.google.com/drive/1FS-vc4RSav5xalC42u9Ce9HClWcNSFmE?usp=sharing>