Term Deposit Customer Marketing Analysis

By Sophia Wagner
Apziva Project 2

Problem Scope

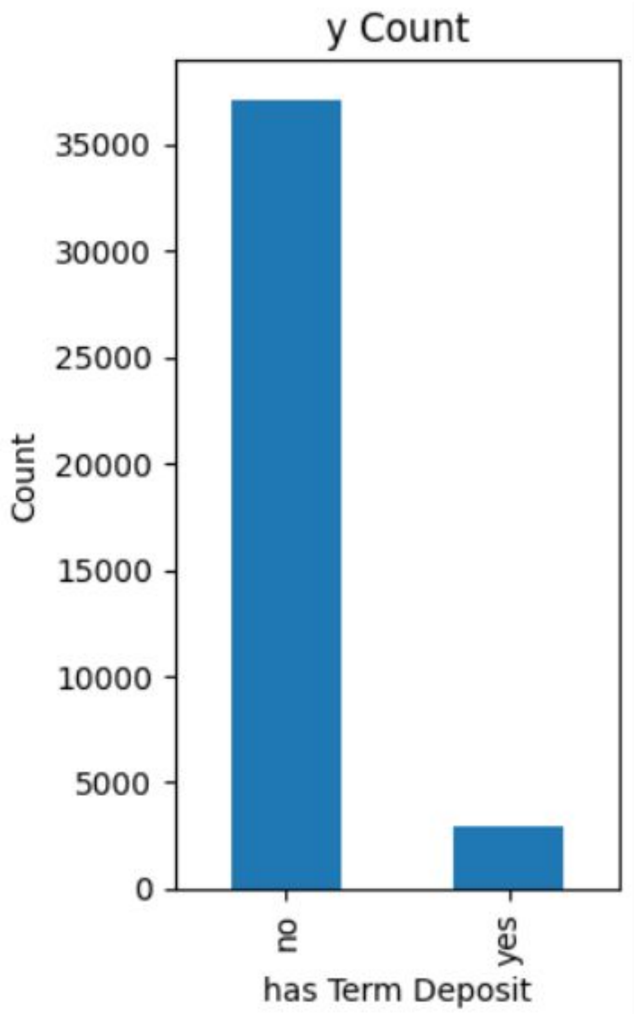
Scope: This dataset comes from direct marketing efforts of a European banking institution. The Marketing campaign is trying ensure subscriptions to a Term Deposit product. Term deposits are usually short-term deposits with maturities ranging from one month to a few years. When buying a term deposit the customer cannot withdraw their fund until after the term ends.

Goal: create a robust machine learning system that leverages information coming from call center data to predict potential subscribing customers

Peek at the Dataset (EDA)

Features in the dataset:

- Age (min: 19 | max: 95 | mean: 40.5 | std: 9.6 | 75% entries <= 48)
- Balance (min: -8019 | max: 102127 | mean: 1274.28 | std: 2903.77 | 75% entries <= 1319)
- Job (12 categories | most common: blue collar)
- Marital (3 categories | most common: married)
- Education (4 categories | most common: Secondary)
- Default (2 categories, ‘yes’ or ‘no’ | most common: no)
- Housing (2 categories, ‘yes’ or ‘no’ | most common: yes)
- Loan (2 categories, ‘yes’ or ‘no’ | most common: no)
- Contact (3 categories | most common: cellular)
- Duration (min: 0 | max: 4918 | mean: 254.82 | std: 259.37 | 75% entries <= 313)
- Campaign (min: 1 | max: 63 | mean: 2.88 | std: 3.24 | 75% entries <= 3)
- Day (min: 1 | max: 30 | mean: 16 | std: 8.3 | 75% entries <= 21)
- Month (11 categories | most common: may)



Target Variable Distribution:

7.24% of customers in dataset subscribed to a Term Deposit

92.76% of customers in dataset declined to subscribe

Overview of Models

Model 1:

- This Model would be implemented before calling any customers.
- Trained on only customer account related data: Campaign, Day, Month, Contact, and Duration columns were removed
- Use results to identify who to call
- Benefit of reducing time spent making unsuccessful calls
- Optimized Recall performance which minimized the number of false negatives

Model 2:

- This Model would be implemented after calling customers for the first time
- Trained on full dataset
- Use to identify who to continue calling after the initial campaign
- Continued optimization of marketing call time
- Optimized Precision performance to minimize the number of false positives
- Model can be used independently or together with model 1 as a two layer system

Model 1

Random Forest Classifier with 5 fold cross validation

Random Under Sampling applied to training data, model trained on 8000 data points

Optimized Recall performance score

Compared and evaluated Gradient Boosting Classifier, Random Forest Classifier, and Decision Tree Classifier

Test Set macro avg Recall Score: 0.61

Full dataset macro avg Recall Score: 0.76

<u>Using rfc model, it is predicted to miss only 8.77% of subscribers</u>

<u>Using rfc model, it is predicted to save at least 50.68% of call time previously spent</u>

```
[[22600 14504]
 [  254  2642]]
              precision    recall  f1-score   support

           0       0.99      0.61      0.75     37104
           1       0.15      0.91      0.26      2896

    accuracy                           0.63     40000
   macro avg       0.57      0.76      0.51     40000
weighted avg       0.93      0.63      0.72     40000
```

Model 2

Gradient Boosting Classifier with 5-fold Cross Validation

Random Under Sampling applied to training data, model trained on 8000 data points

Optimized Precision performance Score

Compared Gradient Boosting Classifier, Random Forest Classifier, and Extreme Gradient Boosting Classifier

Test Set macro avg Precision Score: 0.79

Full dataset macro avg Precision Score: 0.86

<u>Using gbc model, it is predicted to miss no more than 48.86% of subscribers</u>

<u>Using gbc model, it is predicted to save up to 84.03% of call time previously spent</u>

```
[[36624   480]
 [ 1415  1481]]
              precision    recall  f1-score   support

           0       0.96      0.99      0.97     37104
           1       0.76      0.51      0.61      2896

    accuracy                           0.95     40000
   macro avg       0.86      0.75      0.79     40000
weighted avg       0.95      0.95      0.95     40000
```

Customer Segmentation

2 cluster KMeans clustering on all ‘yes’ Term Deposit customers with a campaign number less than or equal to 15

- ‘Yes’ term deposit customers with over 15 campaign calls made up < 1% of dataset

The top job categories with the highest number of ‘yes’ term deposit customers are: Management, Blue-collar, Technician, and Admin   (listed in high to low order)

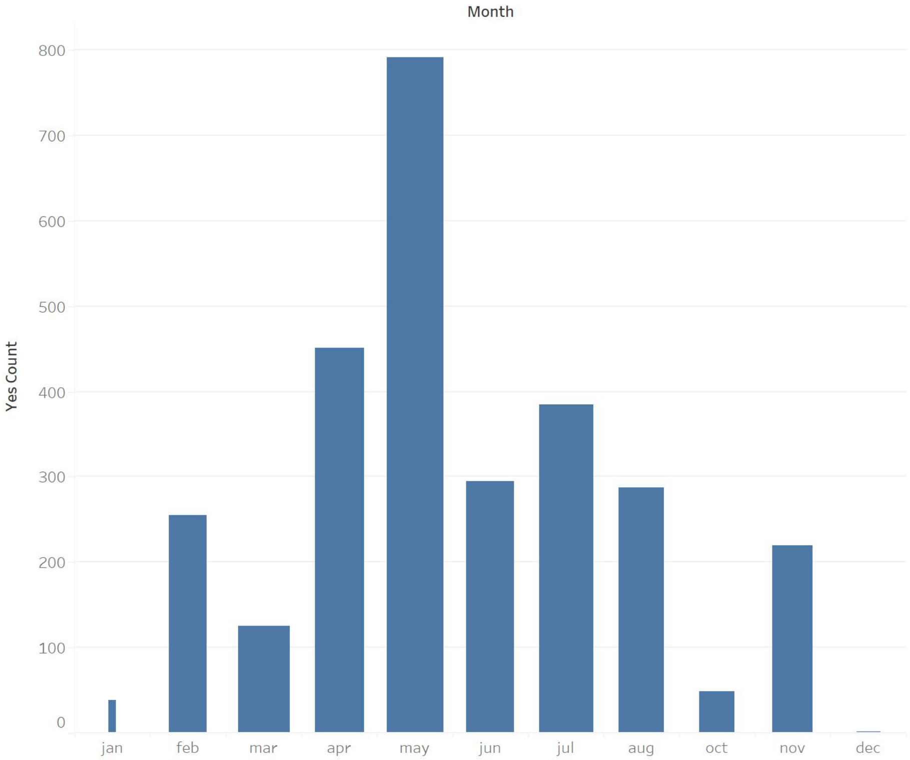
Most subscribers are Married or Divorced

However, Single subscribers have a higher median balance

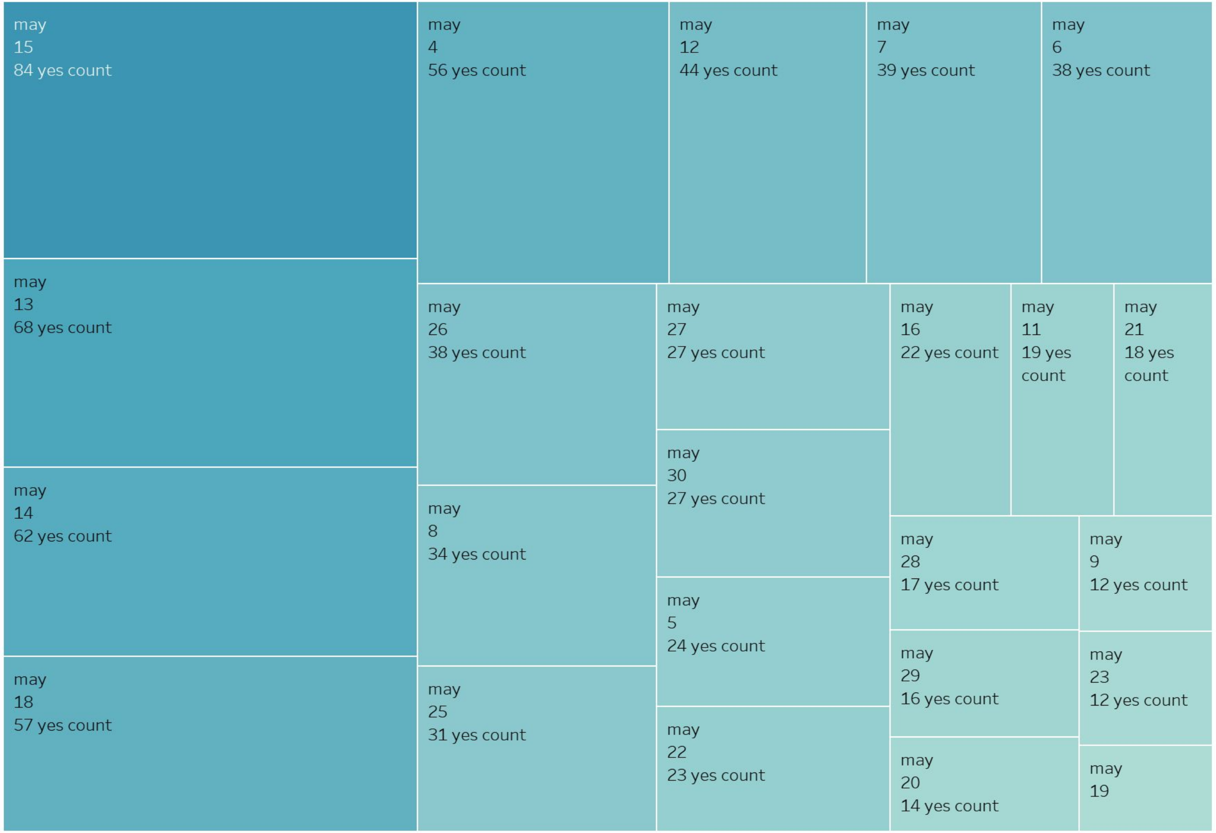
Most subscribers have a Secondary or Tertiary level education

Customer Segmentation

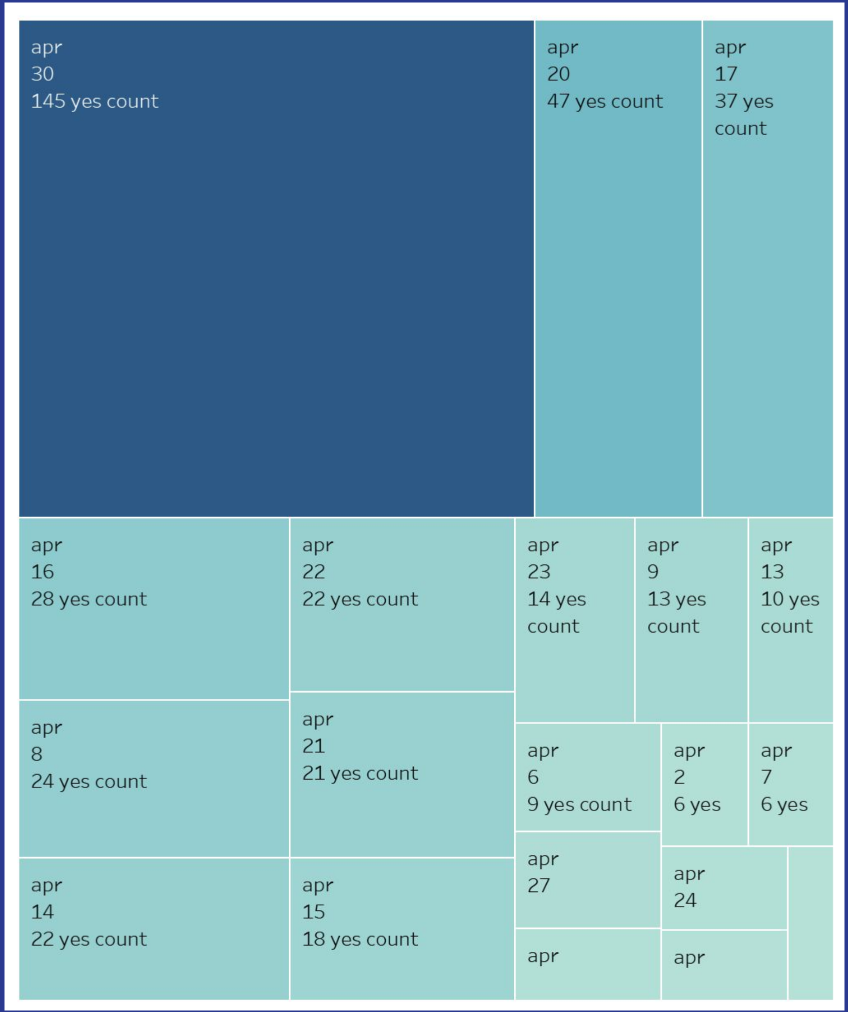
- No calls made to customers in September
- 1 customer subscribed in December
- **<u>May is the best month to call customers</u>**
    - 791 subscriptions in one month
- April is the second best month to call customers



May 4, 12-15, and 18th were top subscription days with the month of May

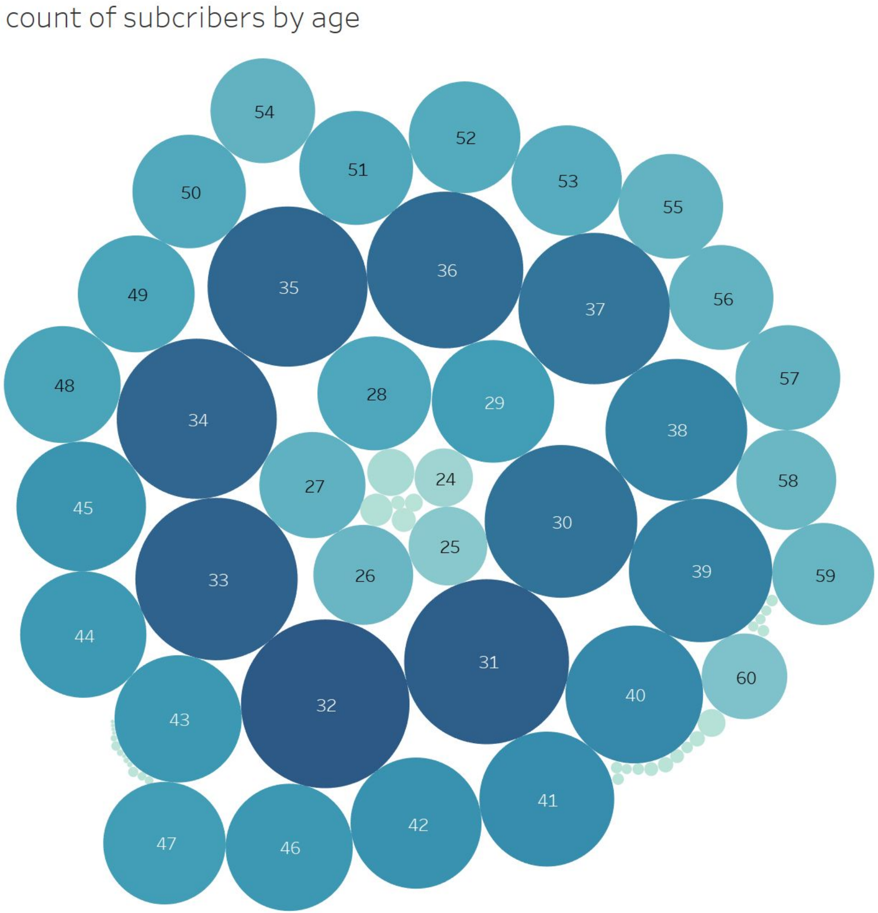


April 30th had the highest subscriber count for one day at 145 subscriptions



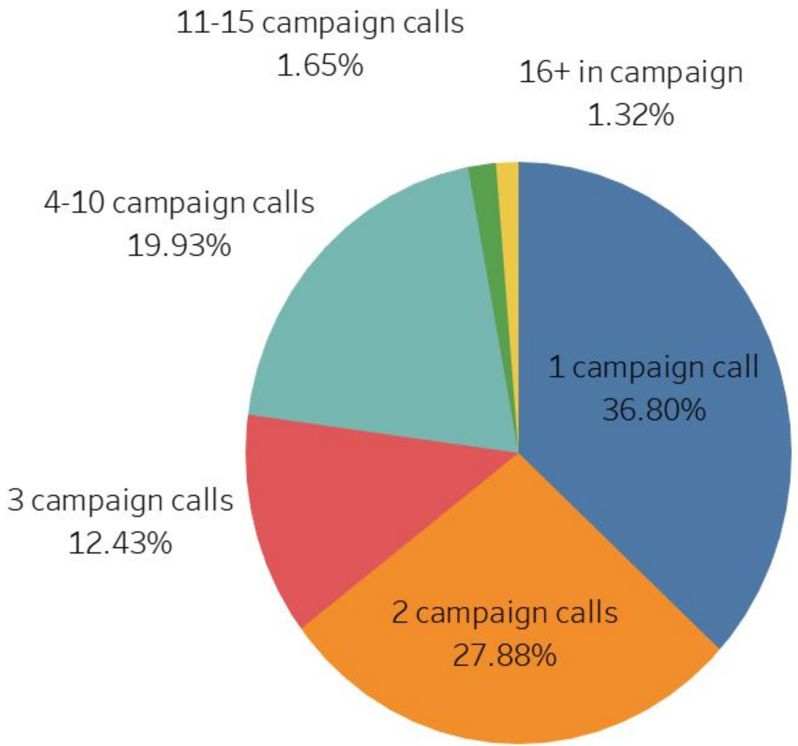
Customer Segmentation

Customers in the age rance 30 - 37 had the highest subscriber count



Other Suggestions/Findings

- Restrict future Campaigns to a <u>maximum</u> of 10 campaign calls
    - Projected to save at least 51 hours of call time
    - Only 2.97% of customers received over 10 campaign calls
        - The ‘Yes’ Term Deposit customers are 0.11% of that
- The majority of subscriber campaign calls lasted between 1 minute and 18 minutes



Experience Gained

- Using various data sampling techniques
- Visualization and Dashboard creation in Tableau
- Unsupervised model creation
- PyCaret
- TPOT and OPTUNA

Thank you!