
# Bank Marketing Campaign

PRANEETHA RAJUPALEPU

SELAELO RAMOKGOPA

SURYA CHANDRA

07/03/2022

# Background

•ABC Bank wants to sell it's term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

•**Objective:** ABC Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

The analysis will provide information about the following:

•Business and Data Understanding

•Data Acquisition & Preparation

•Modelling

•Deployment

# Business Understanding

The plan is to help ABC company to provide a short list of customer that are more likely to buy their product based on their bank details information such as loan.

Marital status, account balance etc.

This goal will be achievable by using a sophisticated machine learning algorithm capable of using a customer record to predict their future action in a blink of an eye to reduce the company's time and resources.

## The success criteria:

The business problem would be based on how much maximum number of customers we are able to predict who have subscribed to the product

# Data Exploration

**Tabular data details:**

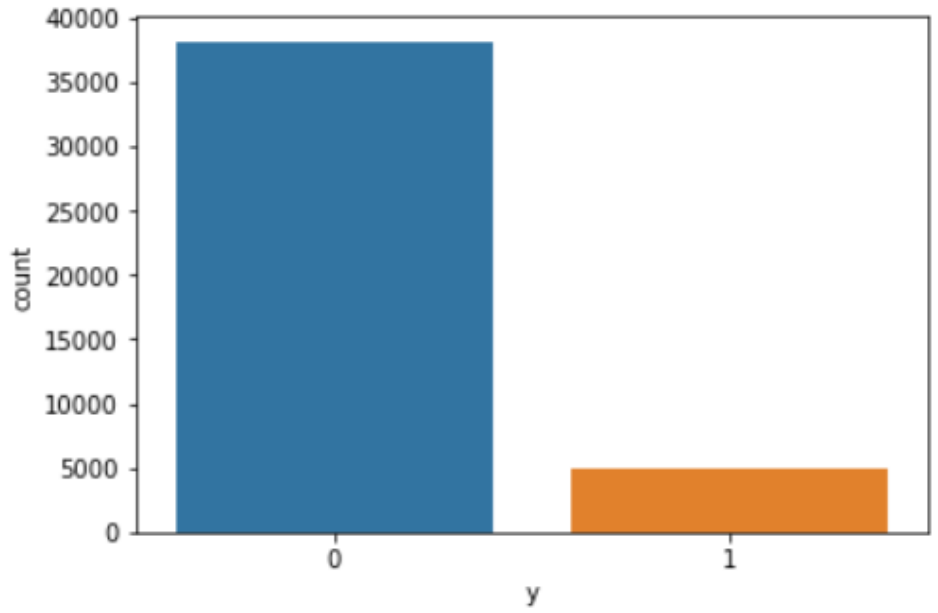
- Total number of observations 41188.
- Total number of files 1.
- Base format of the file .csv Size of the data 4.81 MB.

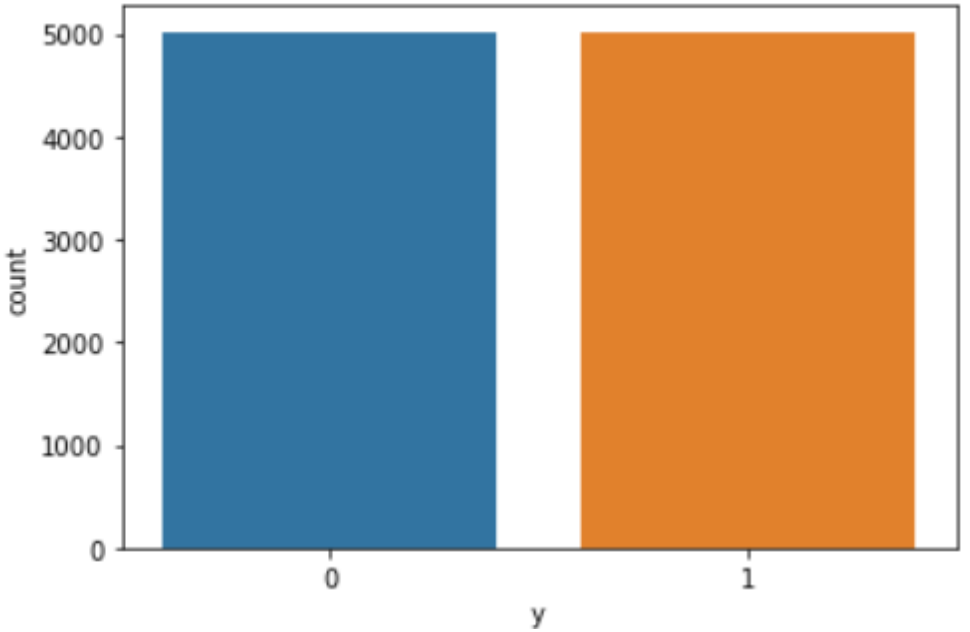
**Approaches:**

- Data looks pretty clean and we have unbalanced classes for our target and outliers.
- Categorical imbalance is reduced by eliminating unknown values and also replacing them with mean.
- The dataset is heavily skewed. We looked at a few of the numeric features, and the ones that are skewed will need to be transformed.
- Techniques like IQR score and Weight of Evidence(WOE) and Information Value(IV) are used to deal with outliers and skewness in the data.

# Handling Imbalance Target Data

- Target column→ Term Deposit subscription (y) is imbalanced.
- Minority class length: 5021  and Majority class length: 38172
- Imbalance is handled using Under sampling technique.



# Data Understanding

- The Relationship between the Number and Duration of Calls:
- According to graph:

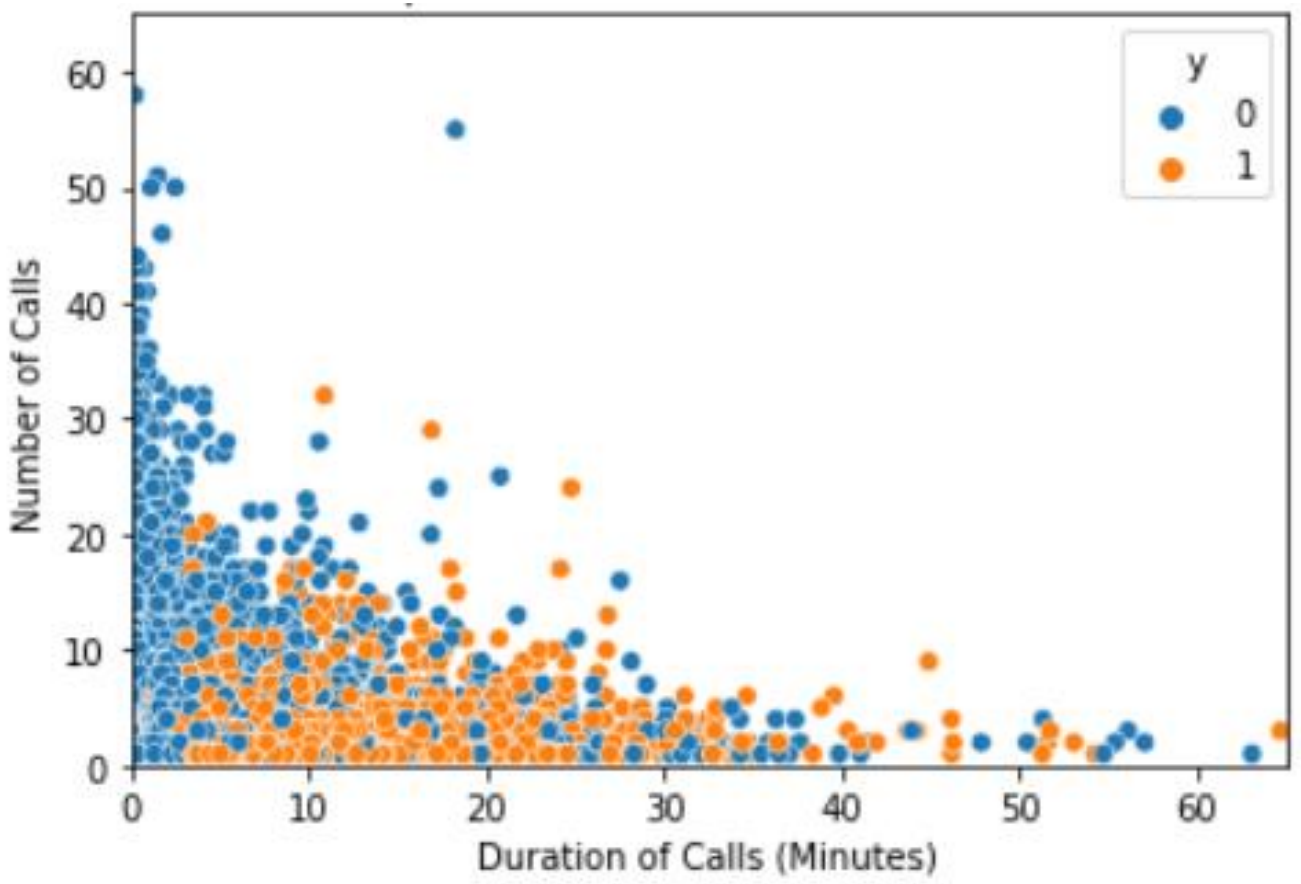
Duration and number of calls

depends on term subscription:

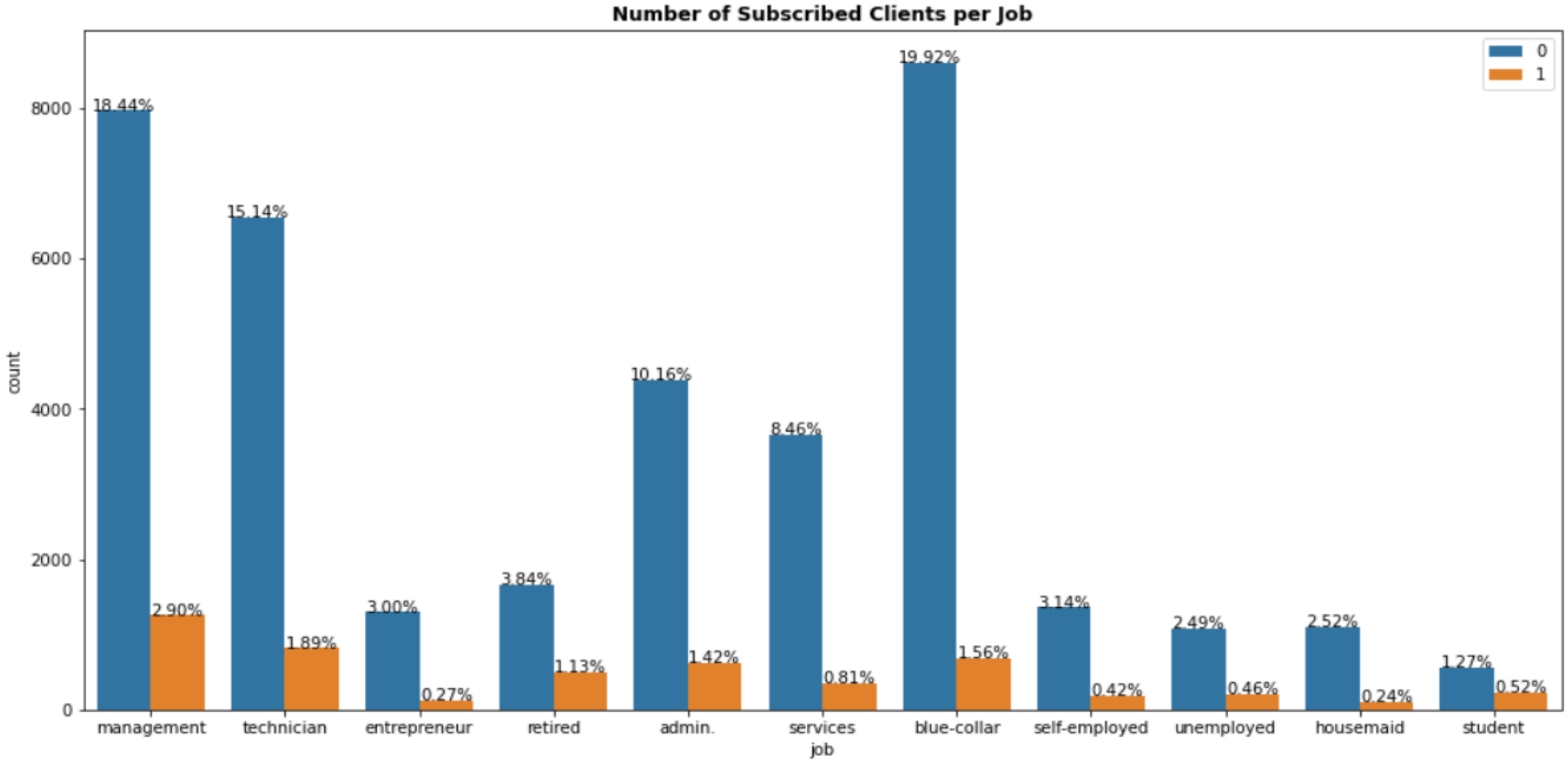
1. Less duration with no subscription
2. 10-20 mins range duration with

minimal calls with subscription-based

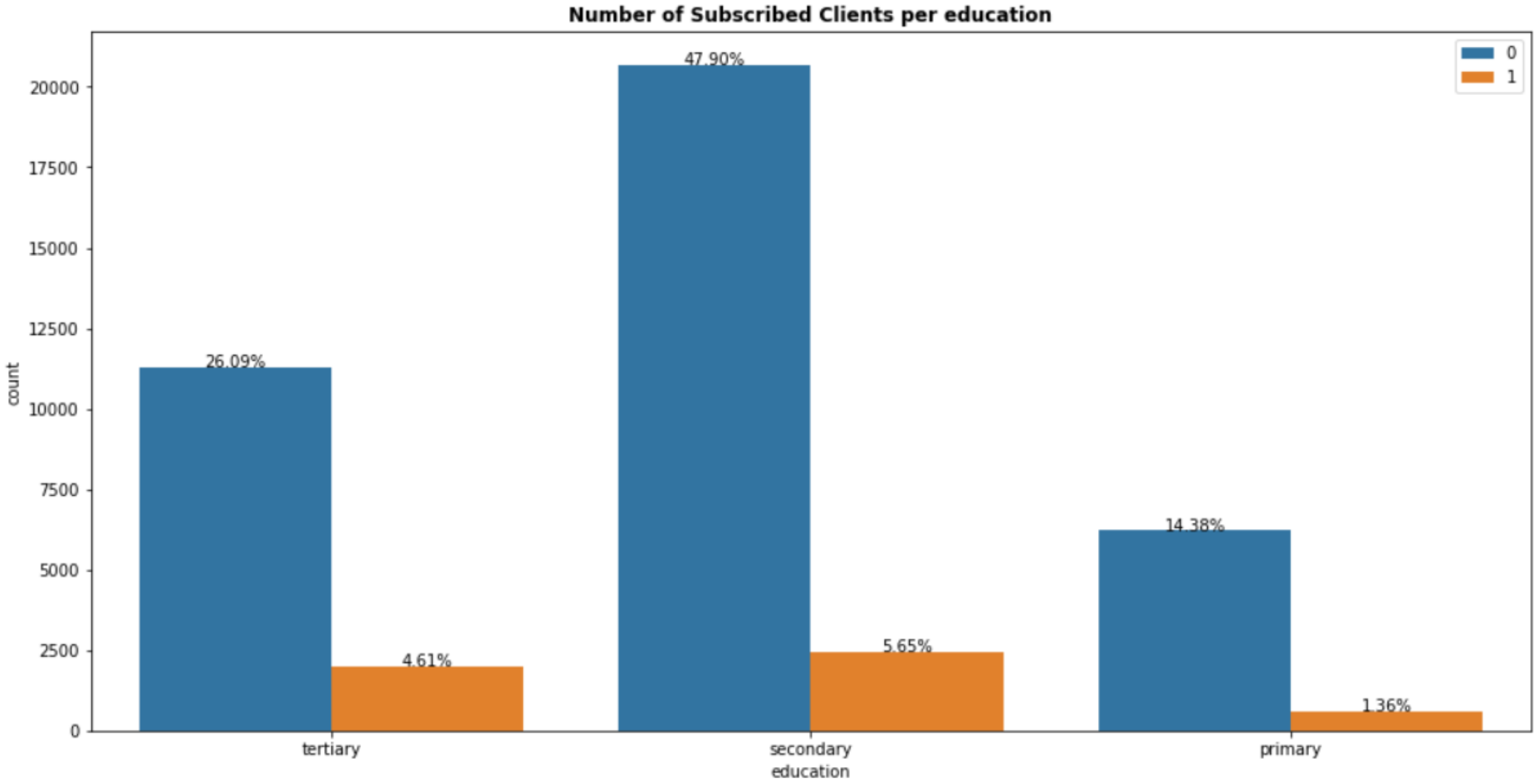
customers.



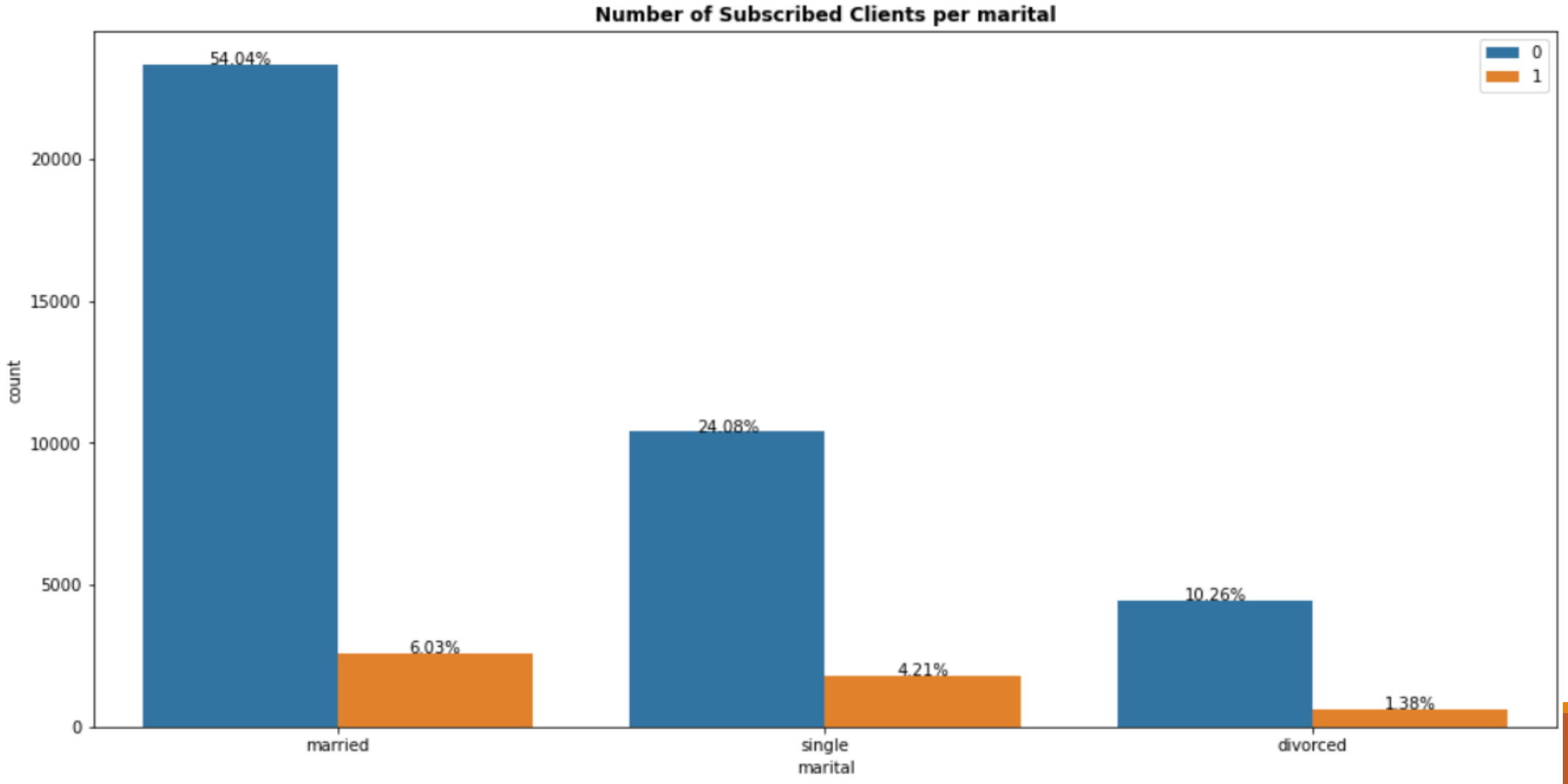
# Client subscriptions  vs Job



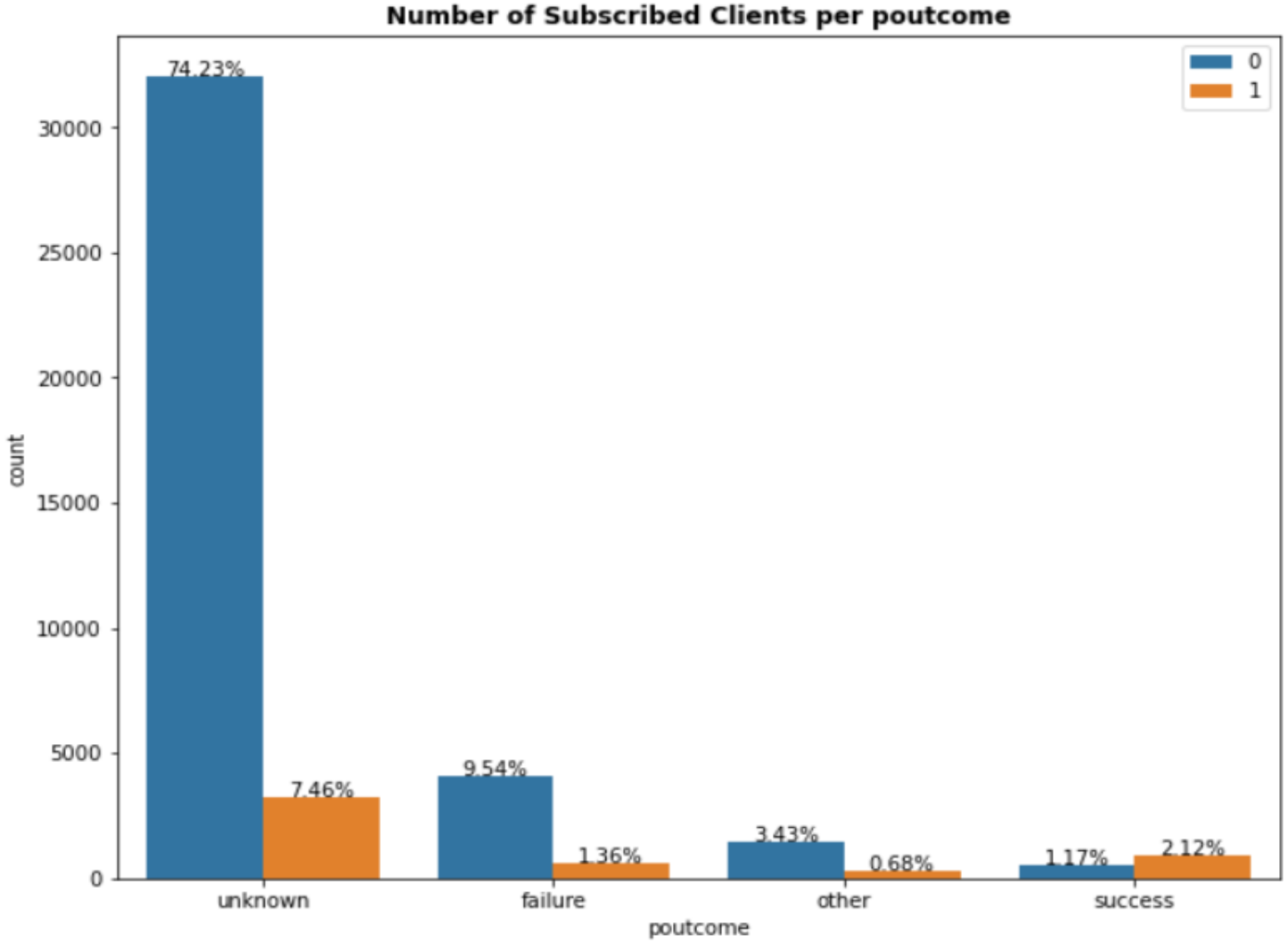
# Client subscription vs Education



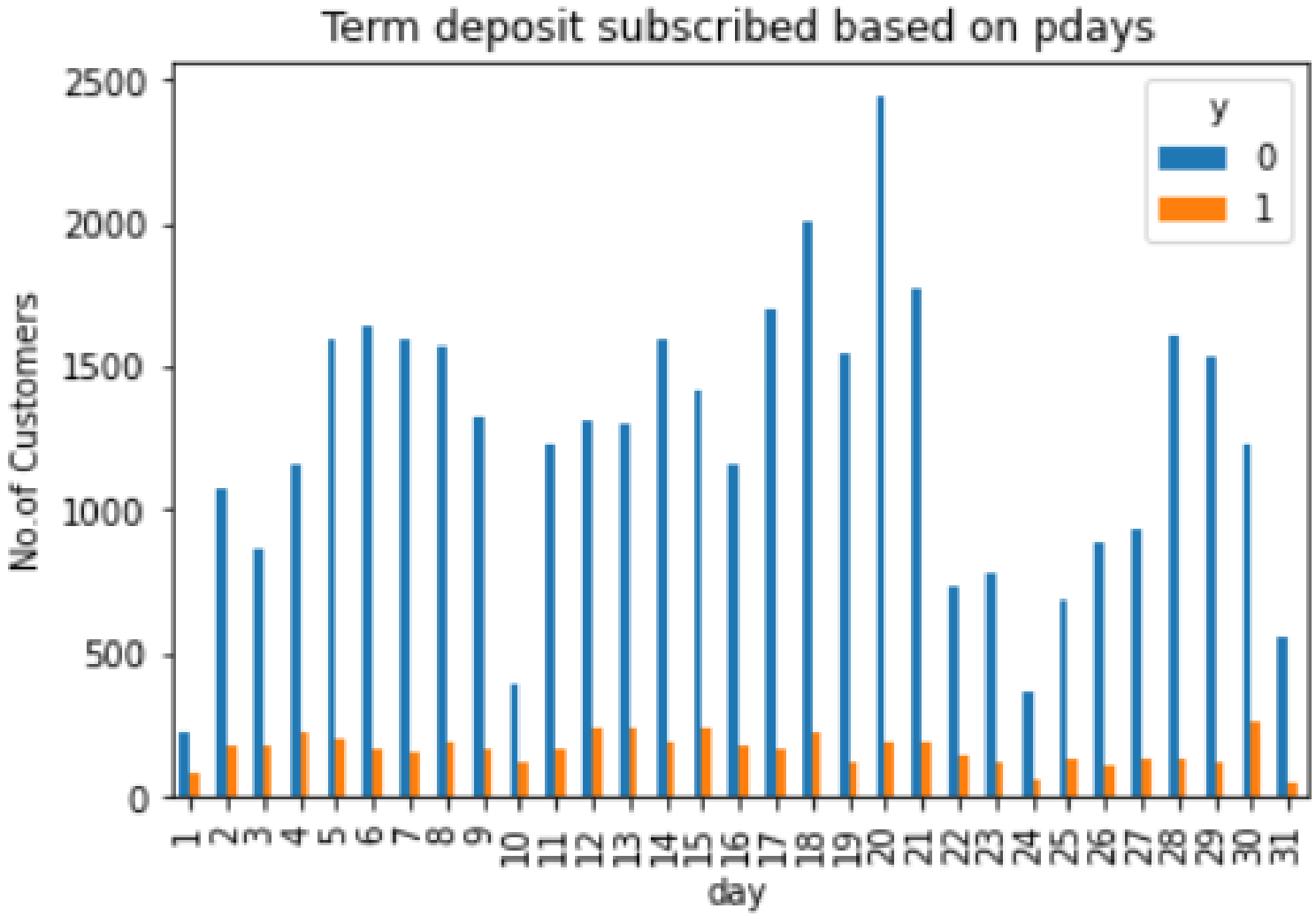
# Client subscription vs Marital Status



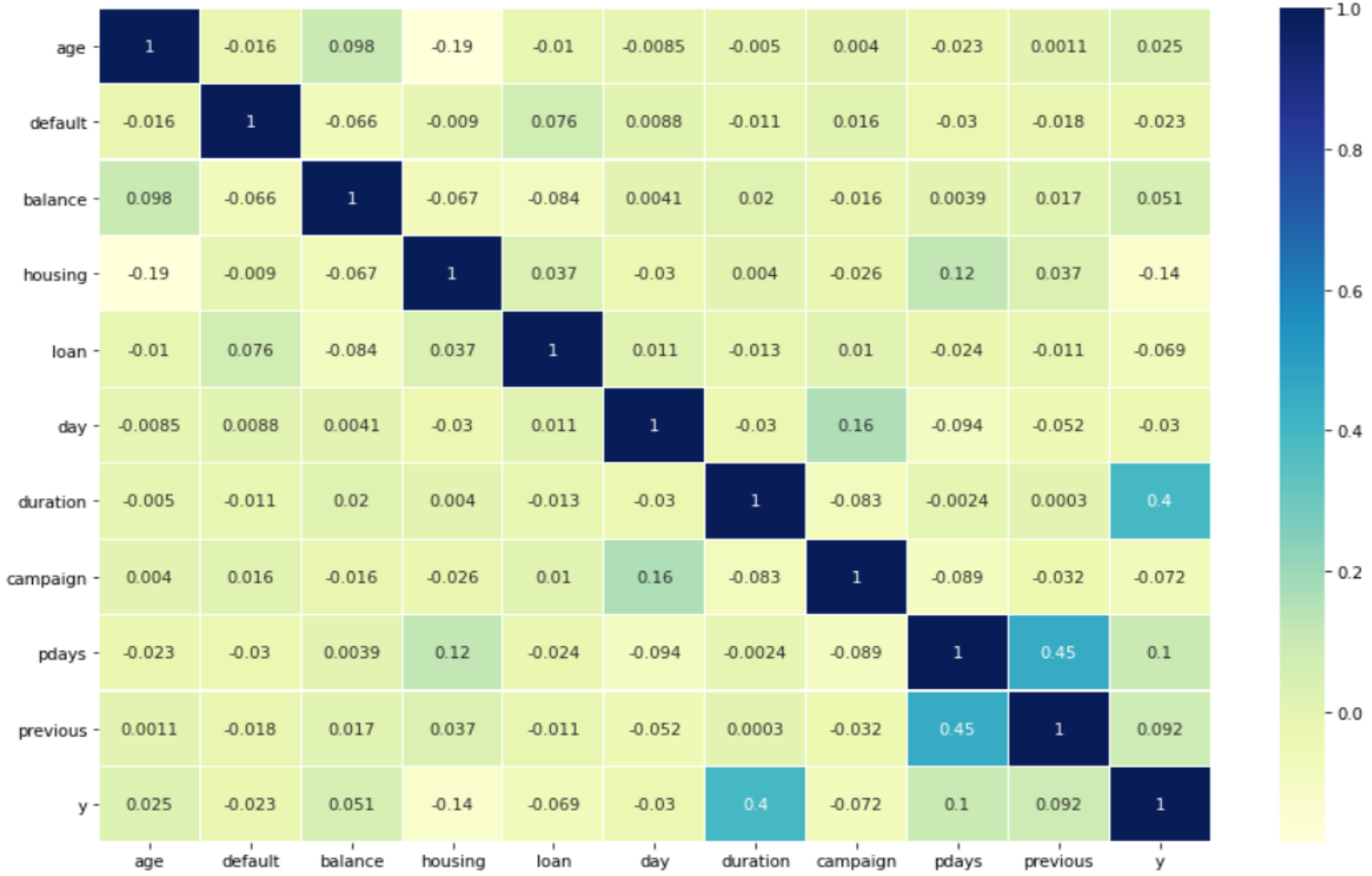
# Client subscriptions vs poutcome



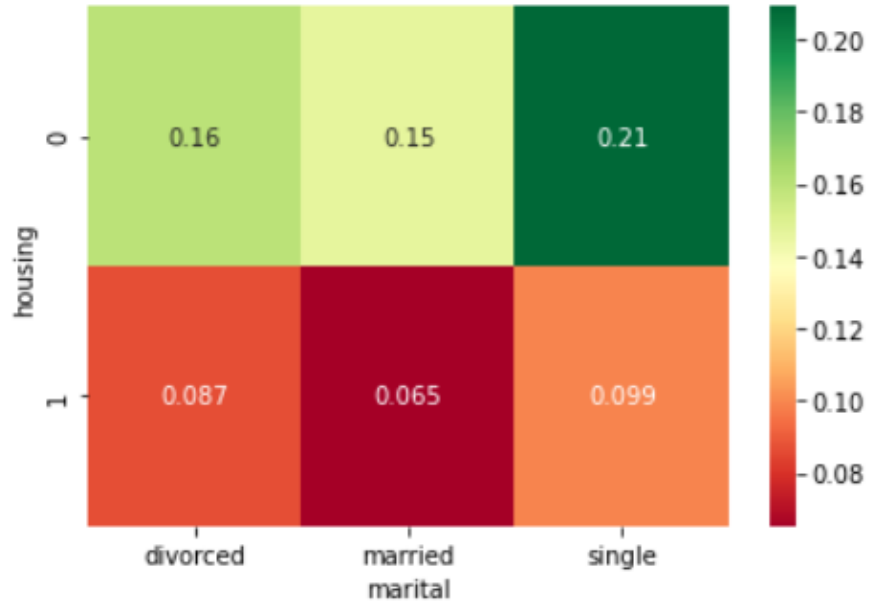
# Client subscriptions vs days



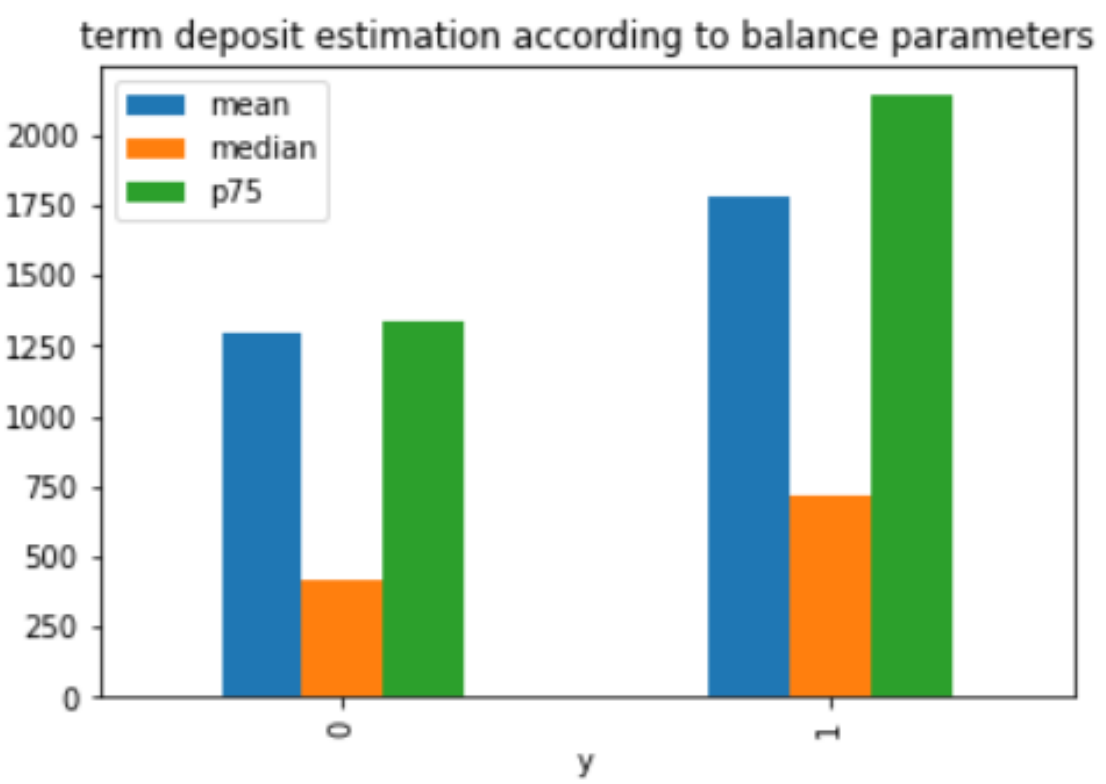
# Correlation Map of Data

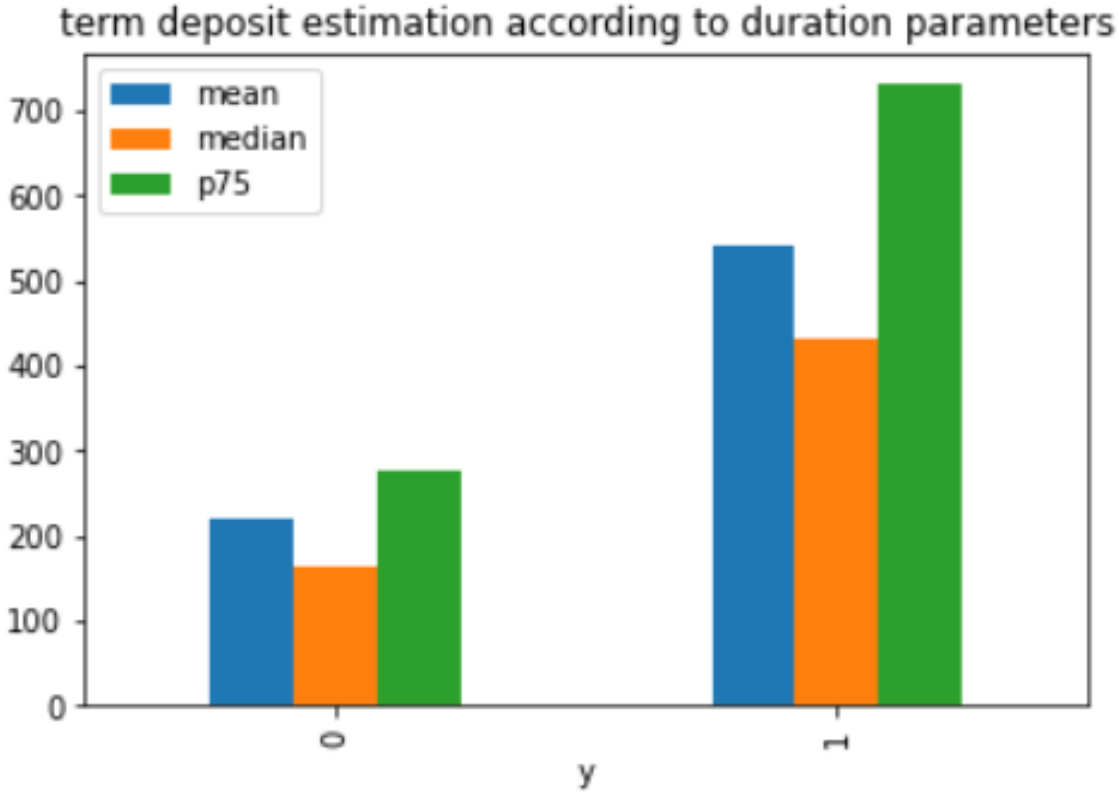


# Pivot Tables representation of client subscriptions between marital status and {education, housing, job}

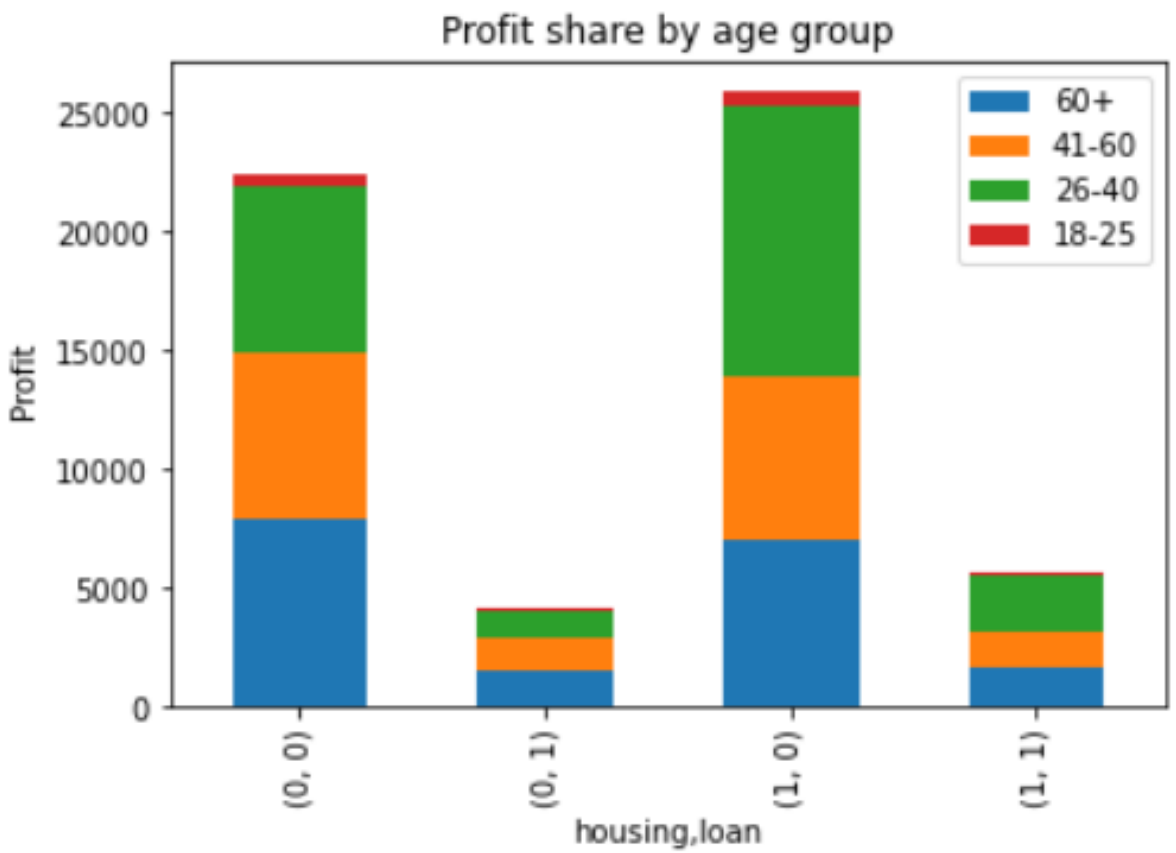



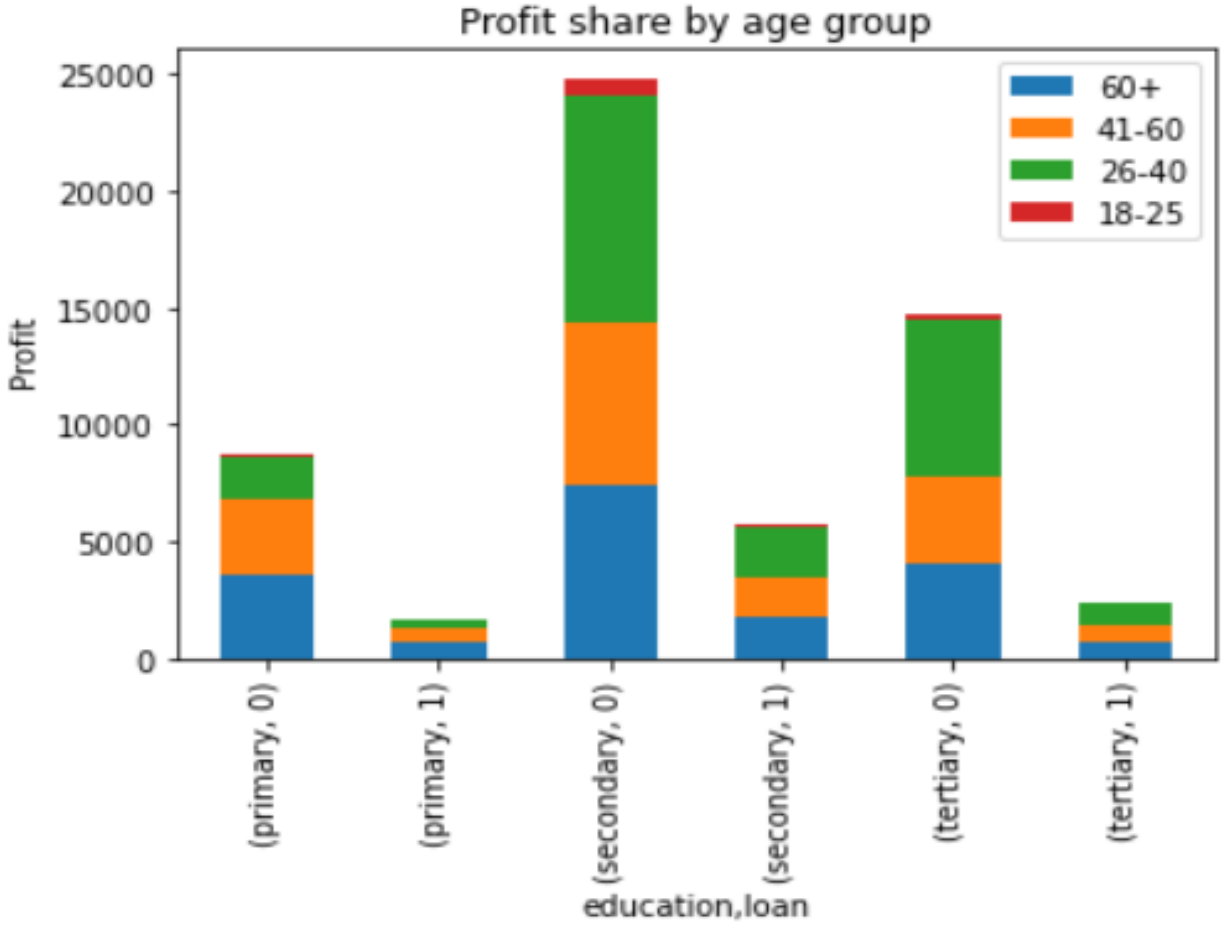

# Term deposit estimation according to balance and duration parameters:



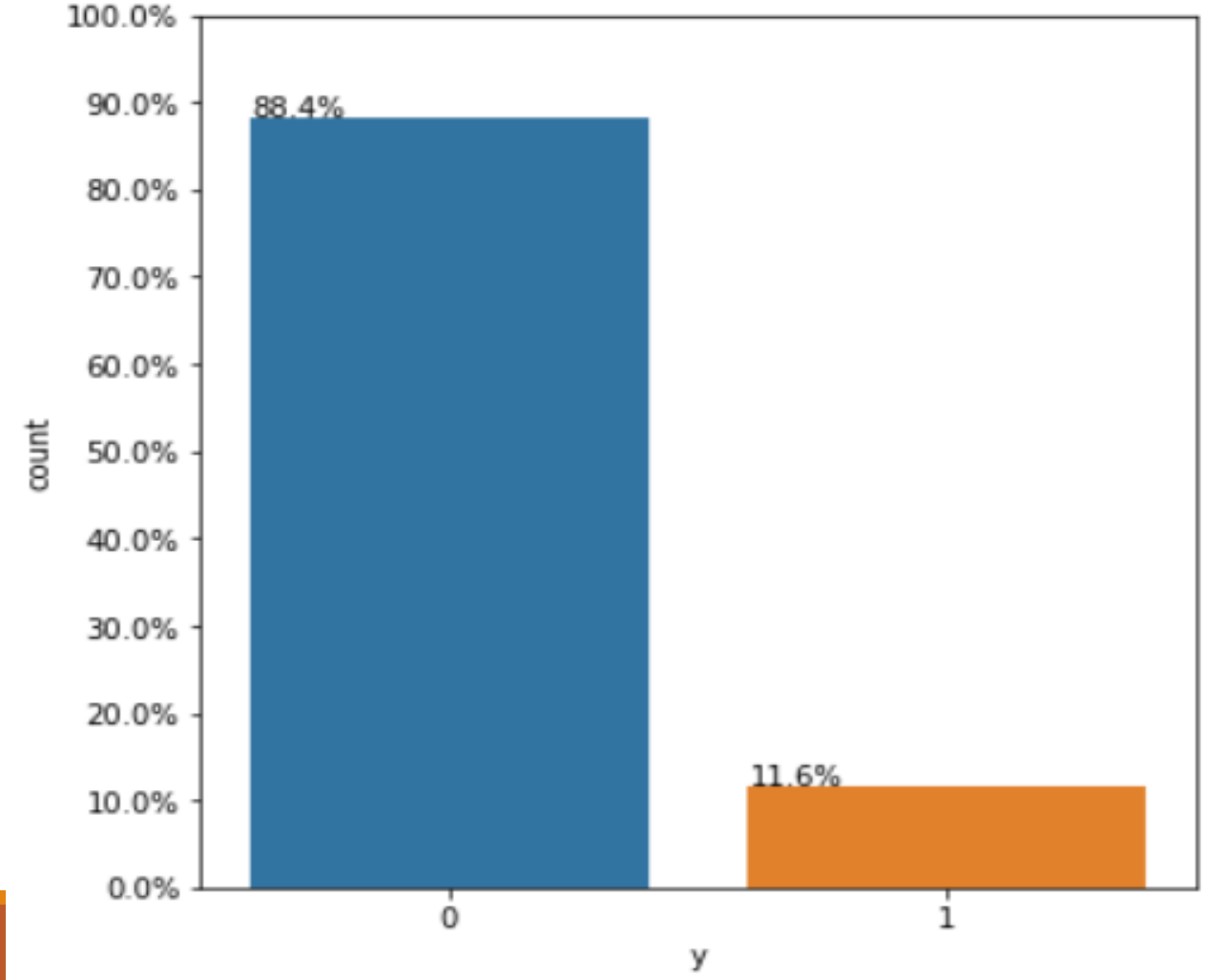



# Profit shares by age groups in housing, loan, and education



# Client subscriptions total count (in %)



# Models

❖Several models considered by our team members are:

1. Logistic Regression
2. Gradient Boosting classifier
3. XGBoosting Classifier
4. Cat Boosting Classifier
5. Voting Classifier

**Note:** Voting classifier is considered to get final ensemble accuracy of all algorithms.

**Key Point:** Due to overfitting issue we tried three different boosting algorithms a part from logistic regression to calculate final accuracy.

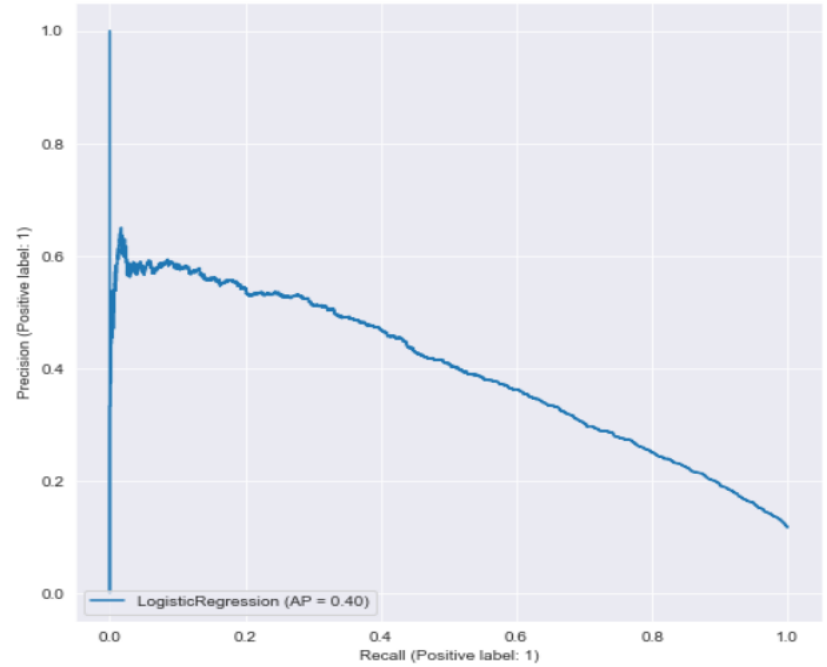
# Test Results

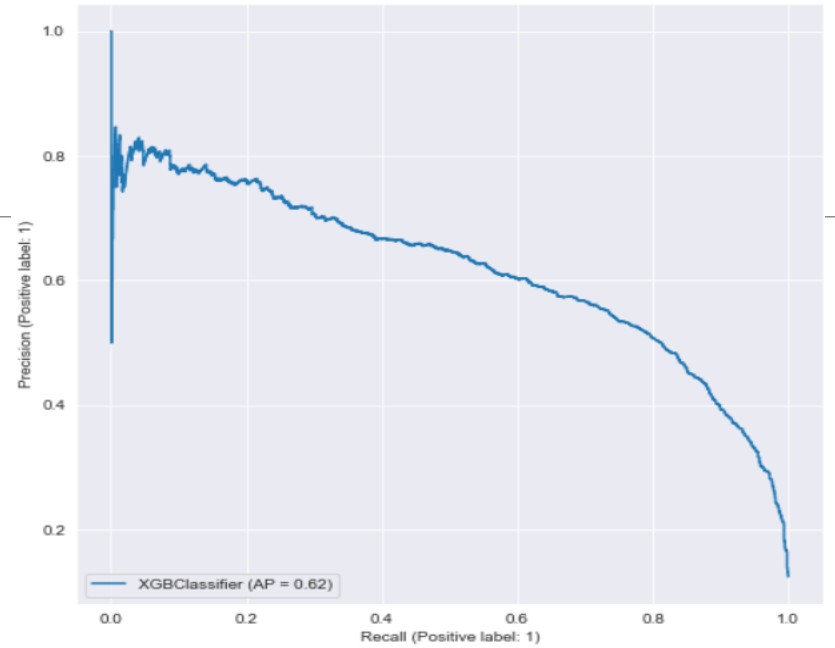
Accuracies:

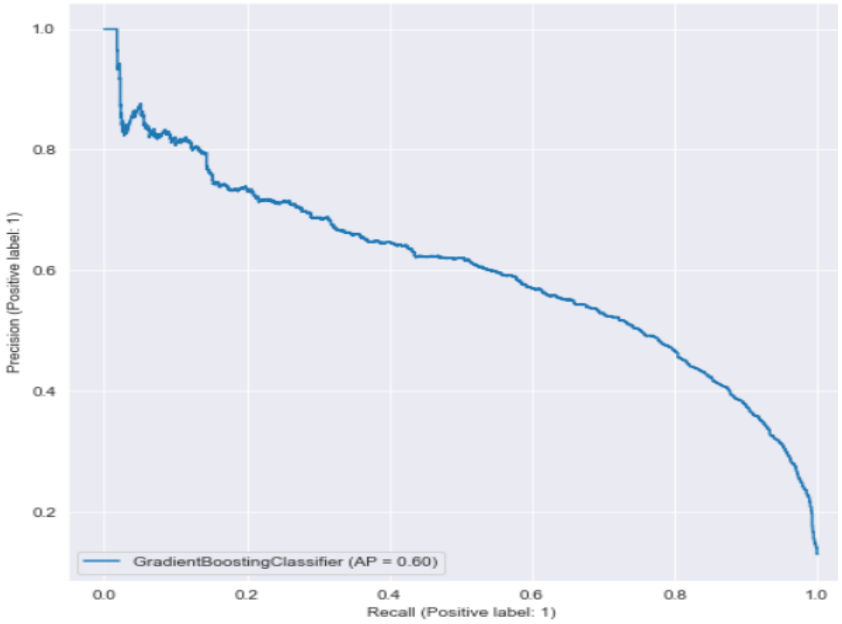
1. Logistic regression: 88.74
2. Gradient Boosting Classifier: 90.38
3. XG_Boosting Classifier: 90.93
4. Cat Boosting Classifier: 91.17
5. Voting classifier: 90.461

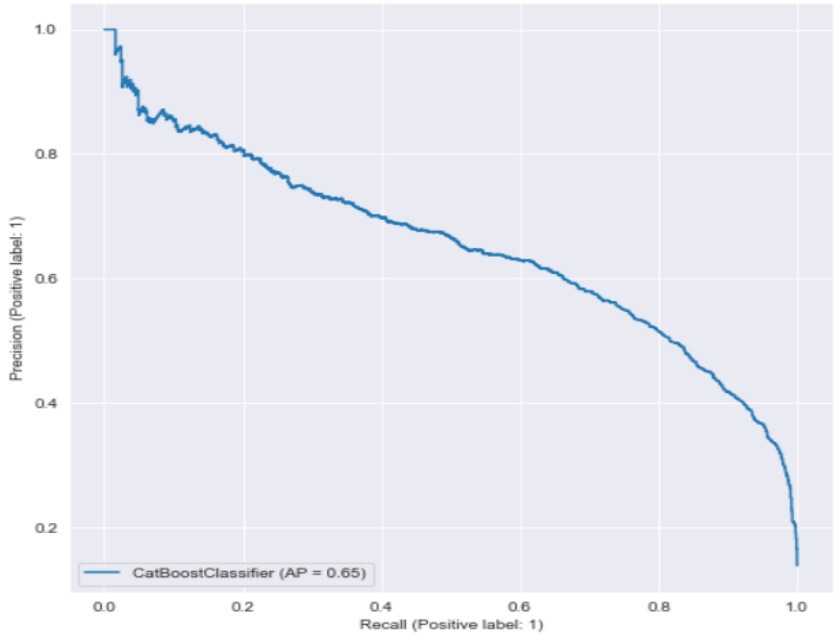
**Note:** Voting classifier final score is comparison best score of all the above algorithms combined.

# Precision-Recall curves




According to AP values of all models observed in the curves, the higher AP value is obtained for **CatBoost classifier.**

# Voting Classifier results

The voting classifier is an ensemble learning method that combines several base models to produce the final optimum solution. This brings diversity in the output, thus called Heterogeneous ensembling.

By applying hard and soft voting for classifier algorithms our group proposed:

Accuracies are:

- ✓Hard voting: 90.461
- ✓Soft voting: 90.816