



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

BANK MARKETING ANALYSIS USING MACHINE
LEARNING

4/13/2022

NAME: CATHERINE SANDA

EMAIL: sandacate@gmail.com

COUNTRY: KENYA

SPECIALIZATION: DATA SCIENCE

Agenda

Executive Summary

Problem Statement

Approach

EDA

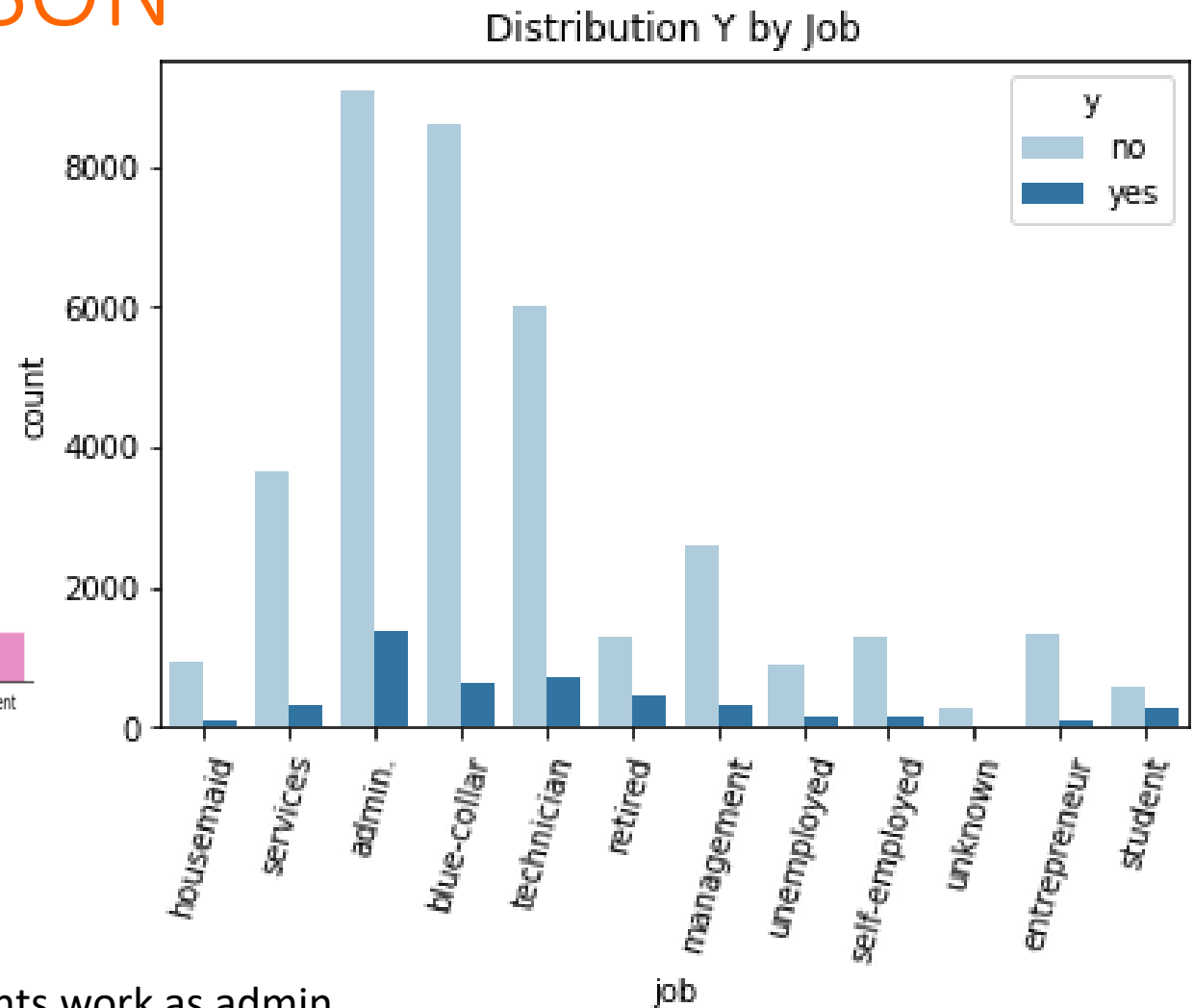
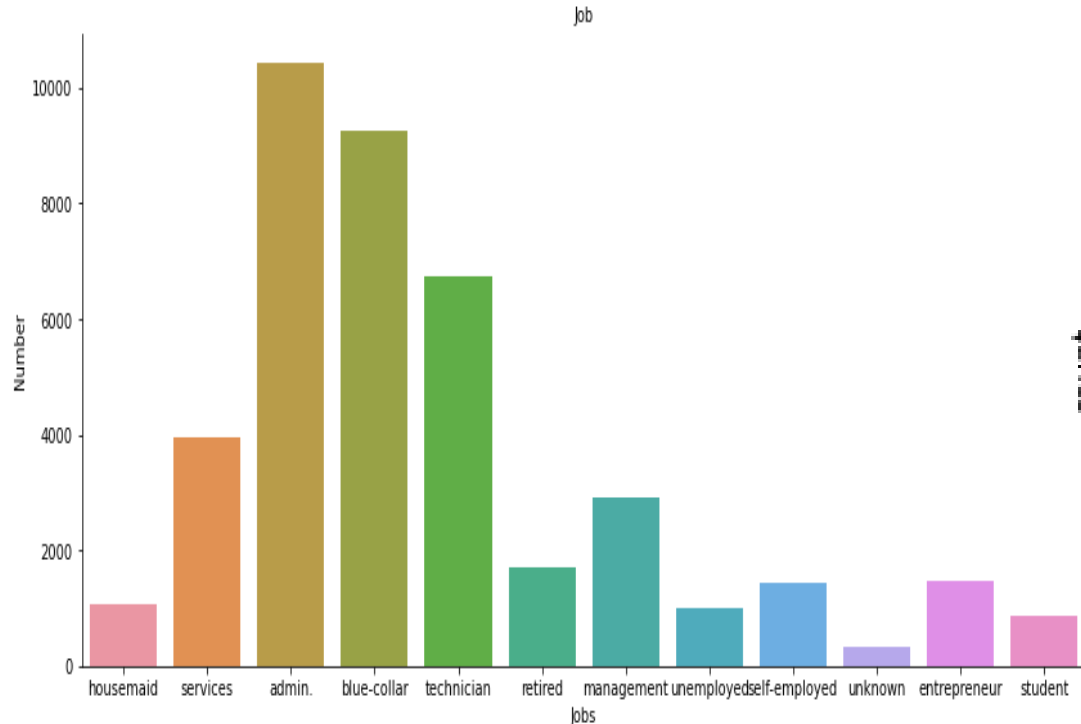
EDA Summary

Recommendations

PROBLEM DESCRIPTION

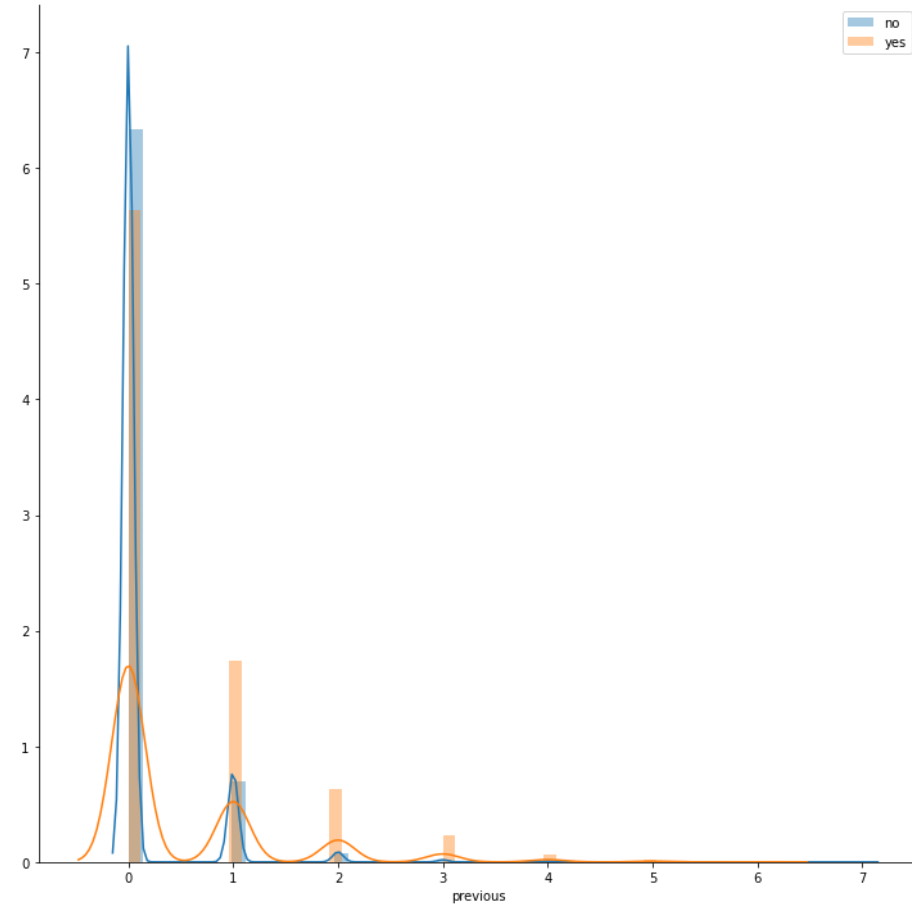
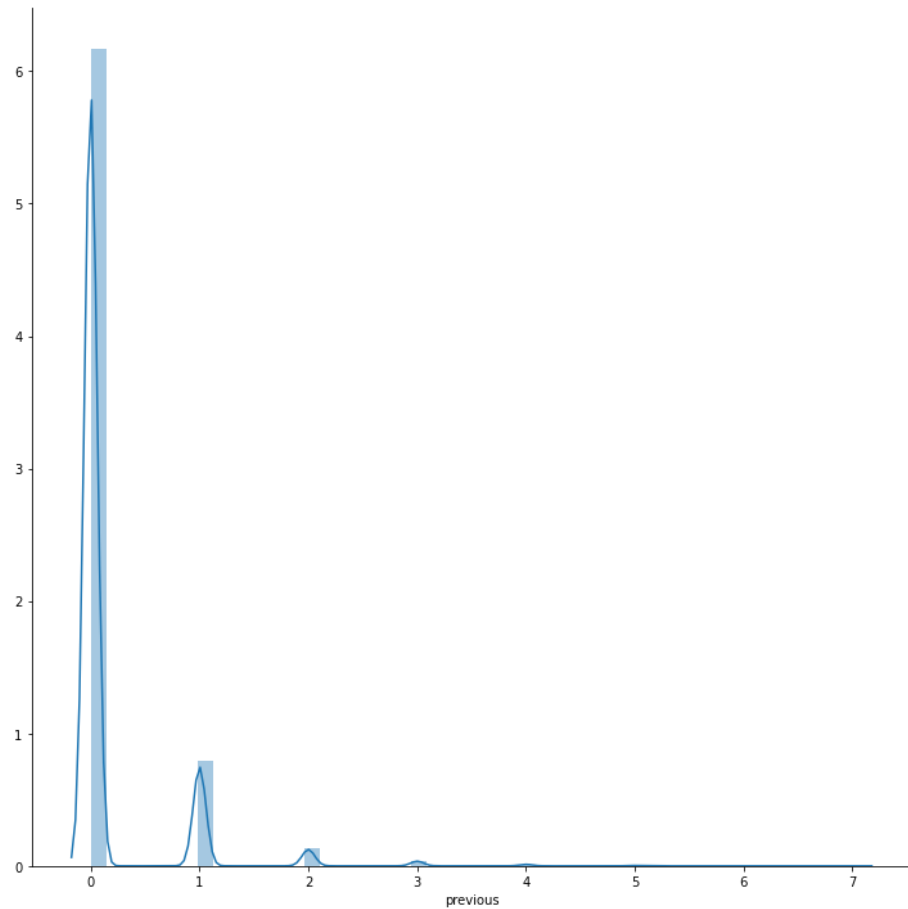
- We are given data related to direct marketing campaigns i.e. phone calls of a bank in Portugal.
- The classification goal is to predict whether a client will subscribe or not(yes/no) to a term deposit (variable y).

FEATURE COMPARISON



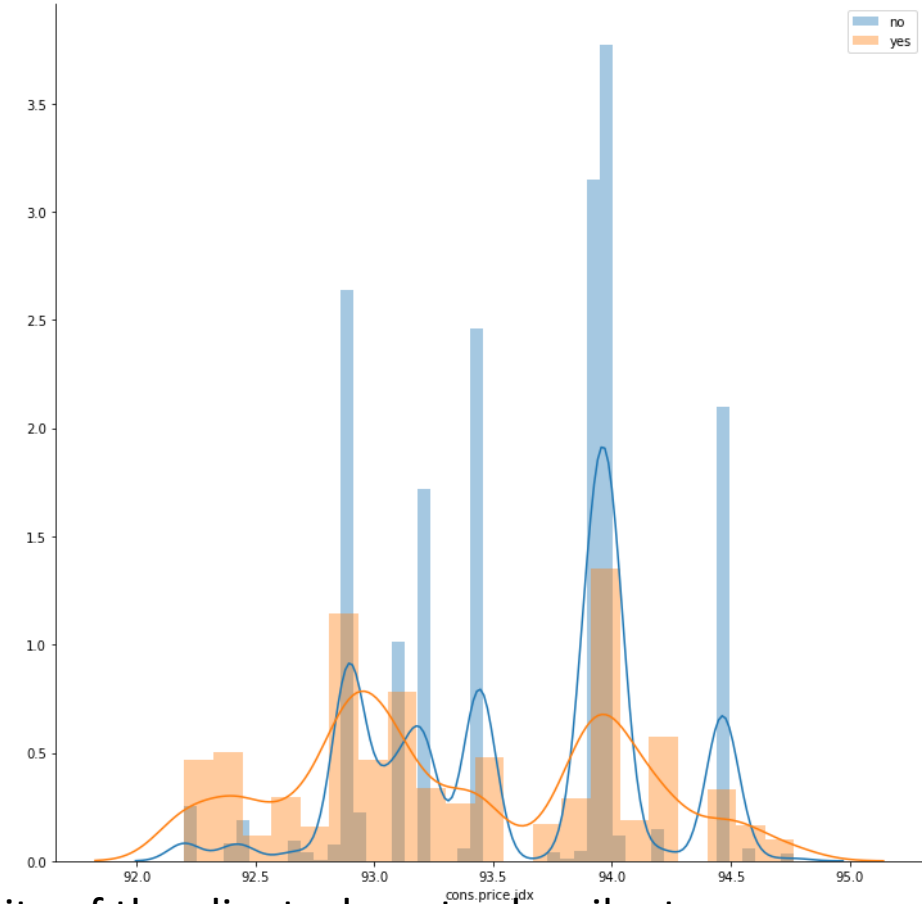
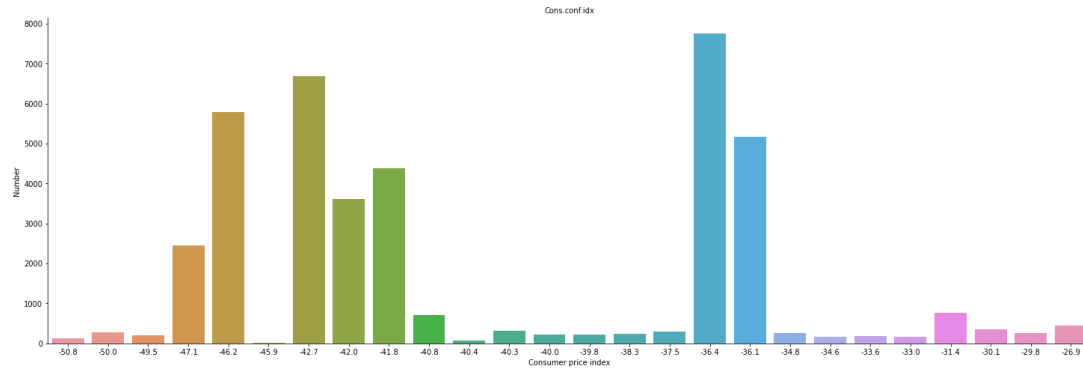
From the above we can see that majority of the clients work as admin. Clients with “admin”, “blue-collar” and “technician” jobs subscribed to the term deposit.

PREVIOUS



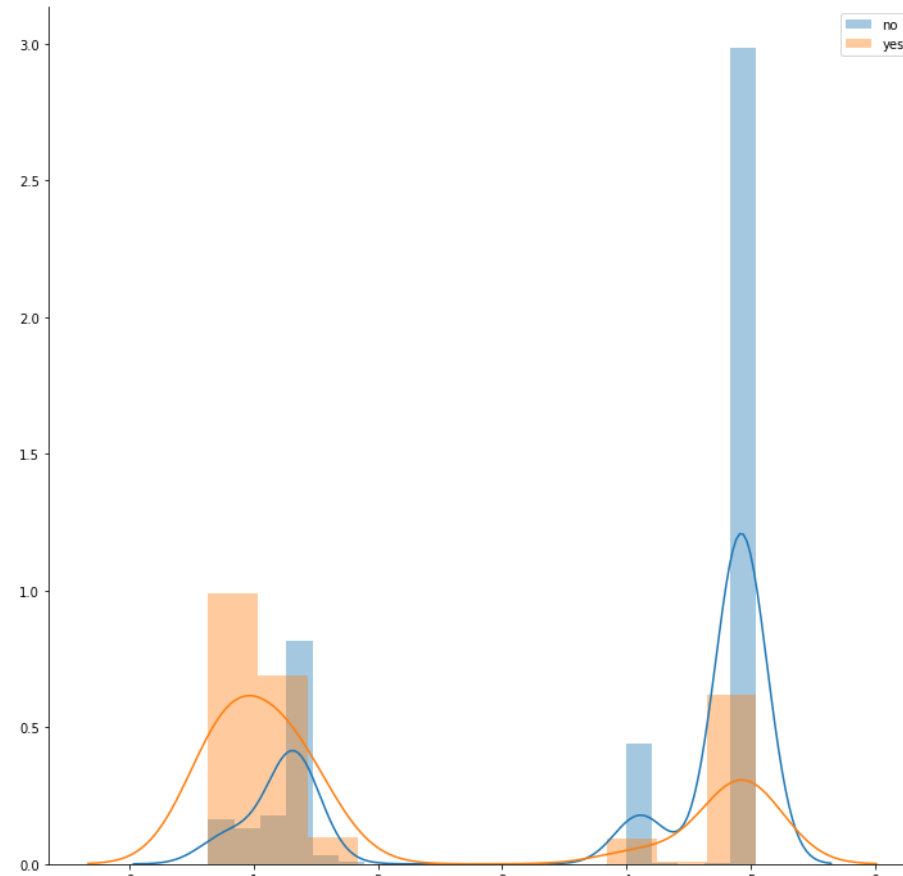
Majority of the clients the bank contacted in this campaign had not been contacted in the last campaign however those who had been contacted previously subscribed to the term deposit at a much higher rate.

CONS.PRICE.IDX



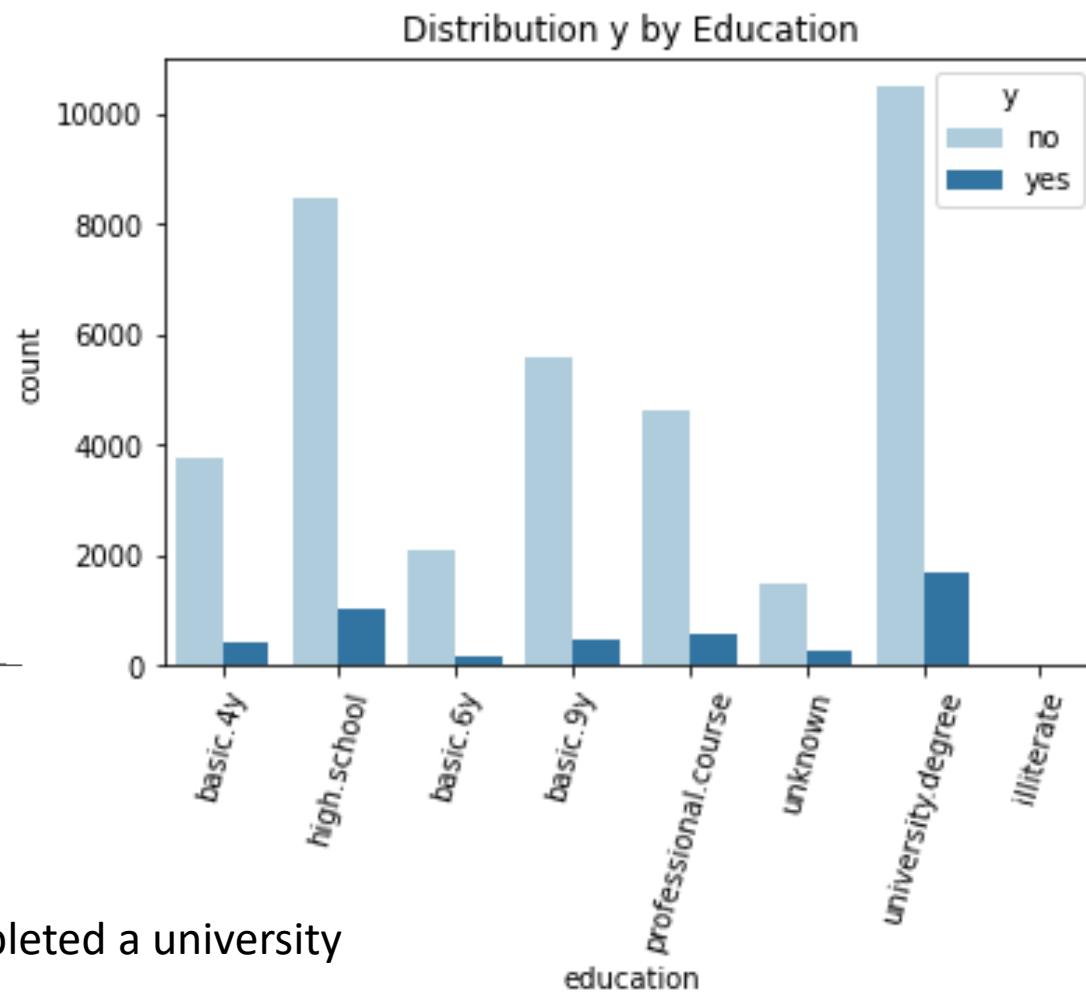
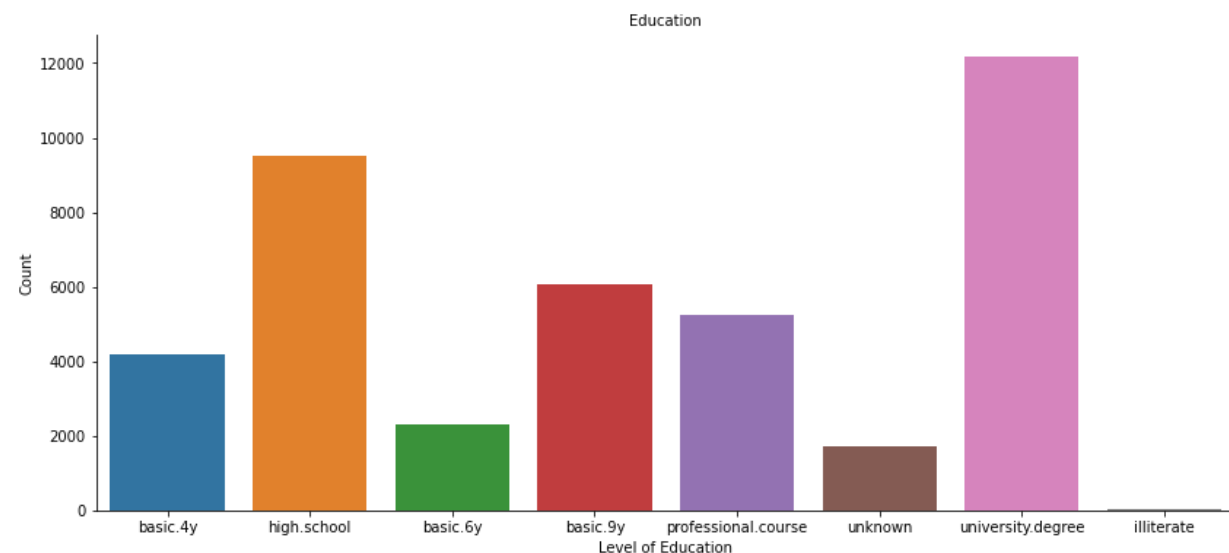
When the cons_price_idx(consumer price index) is high majority of the clients do not subscribe to the term deposit.

EURIBOR3M



The lower the euribor3m the lower the number of subscriptions to the term deposit.

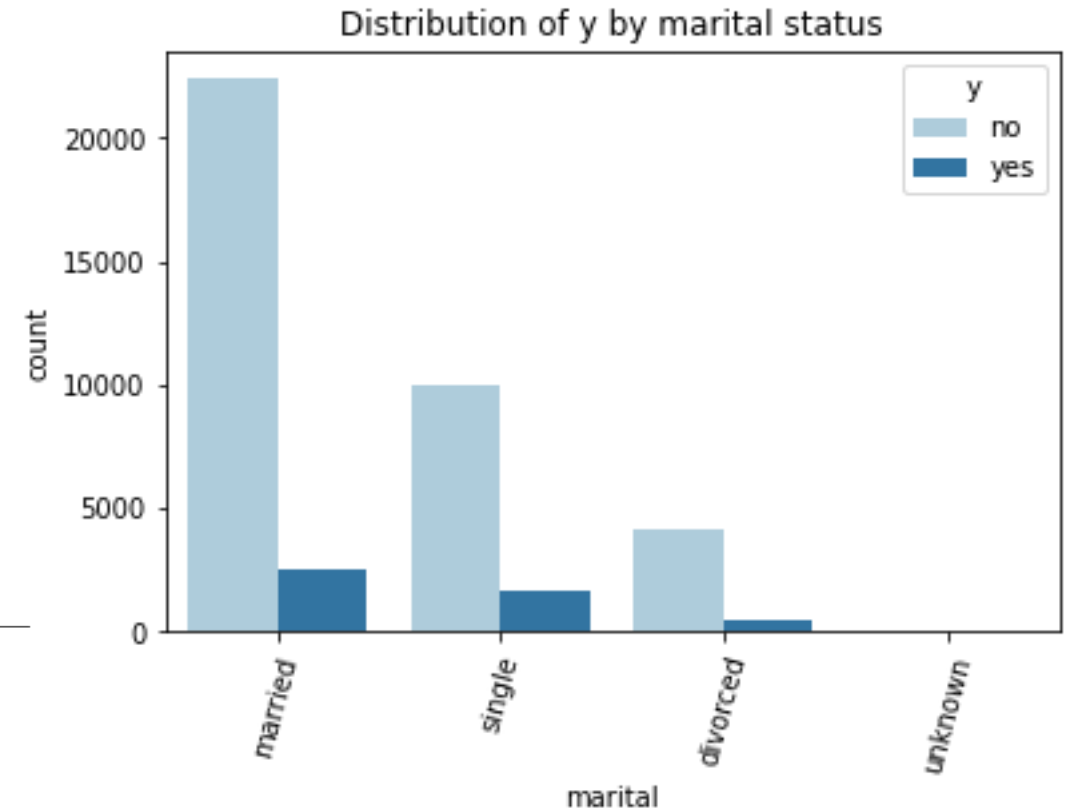
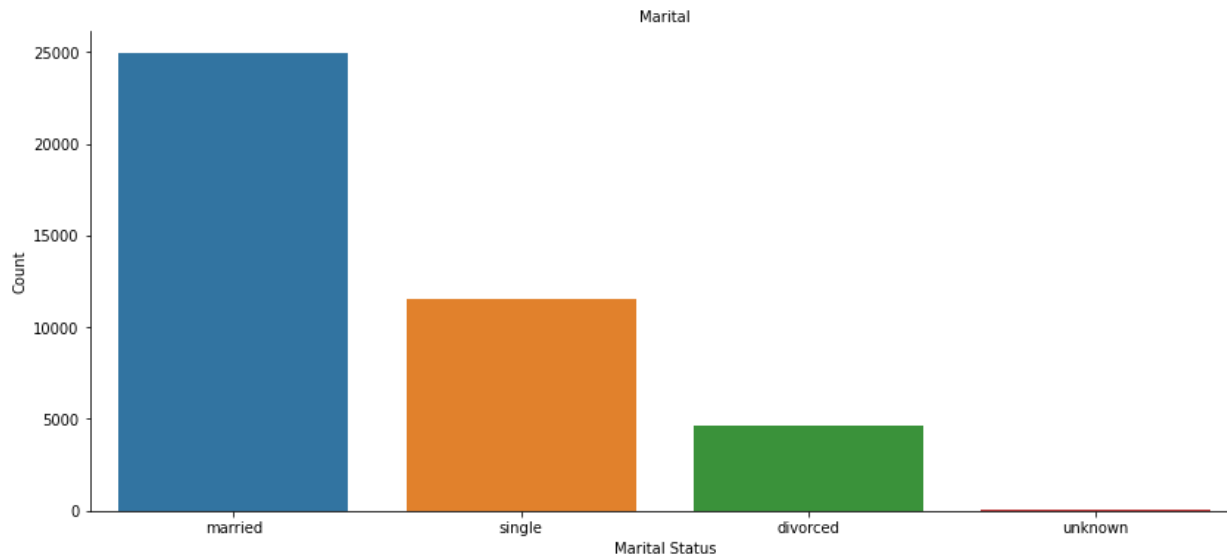
EDUCATION



Majority of the clients who were contacted had completed a university degree.

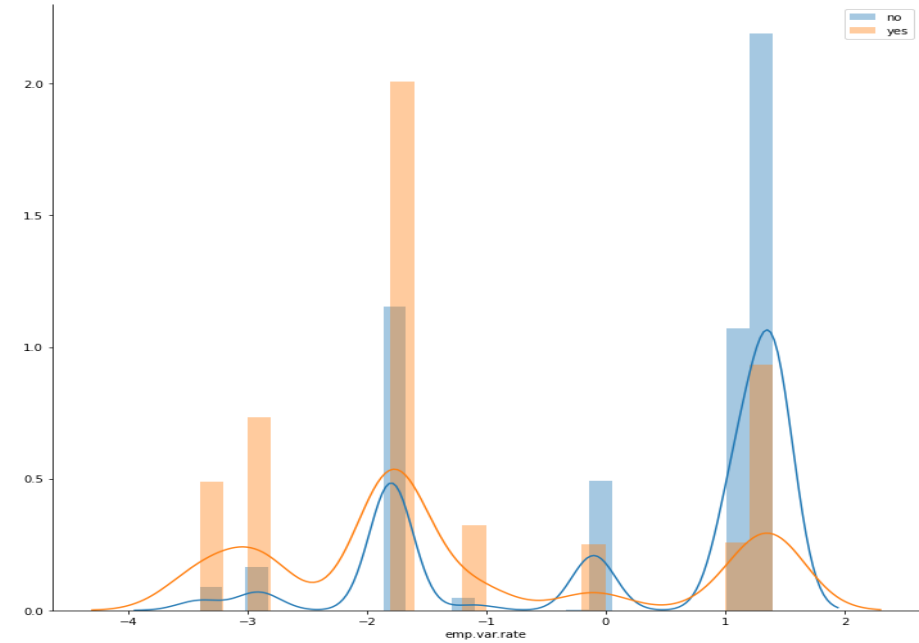
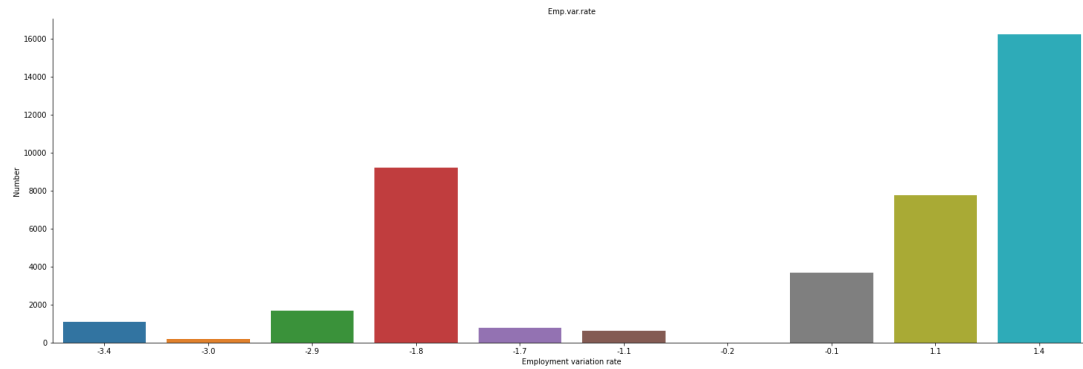
Clients with "university degree", "high school", "basic 9y" and "basic 4y" subscribed more to the term deposits.

MARITAL



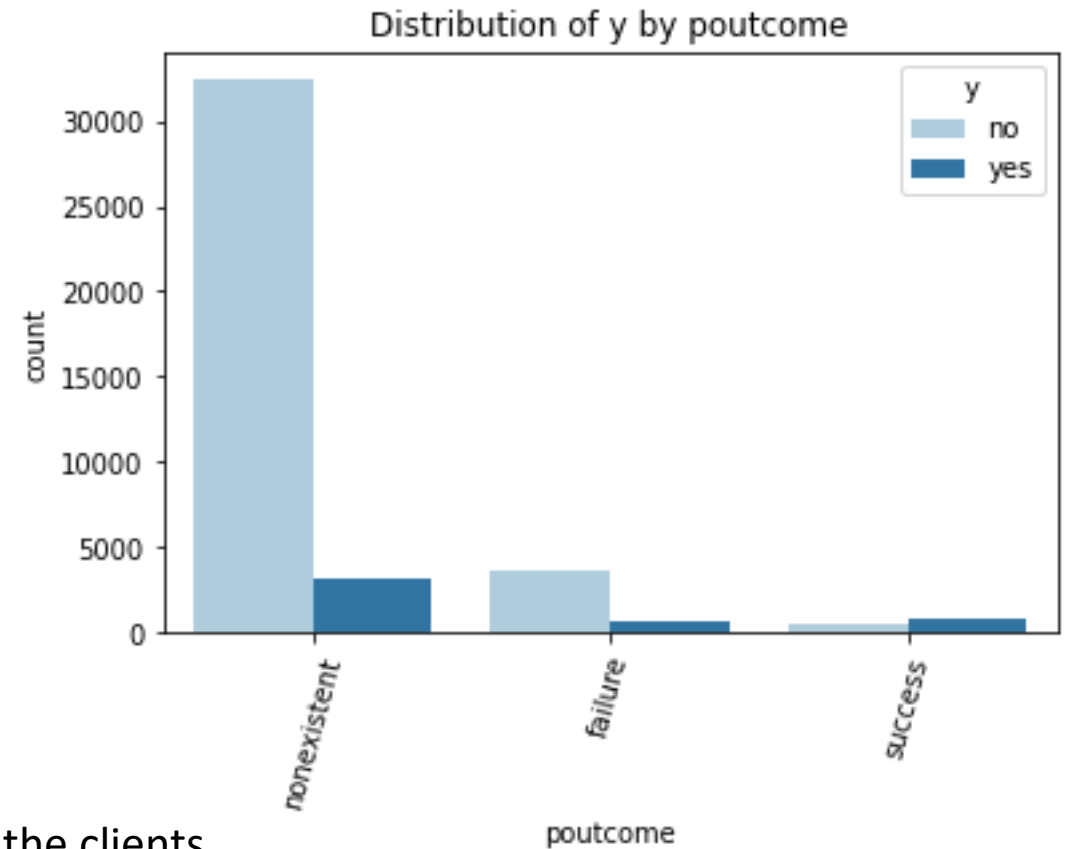
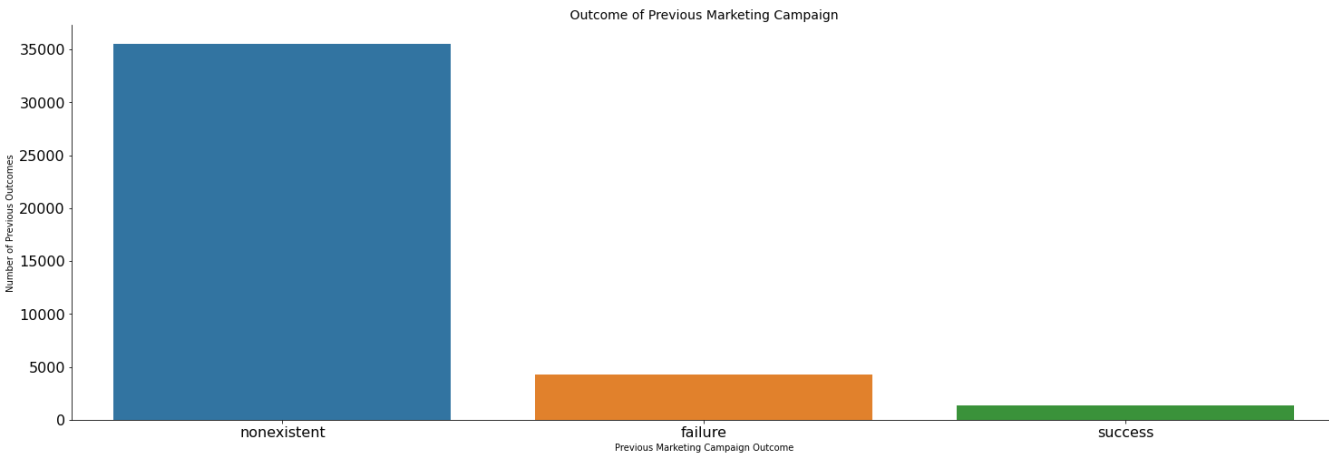
Majority of the clients were married and they subscribed slightly more than those who were single.

EMP.VAR.RATE



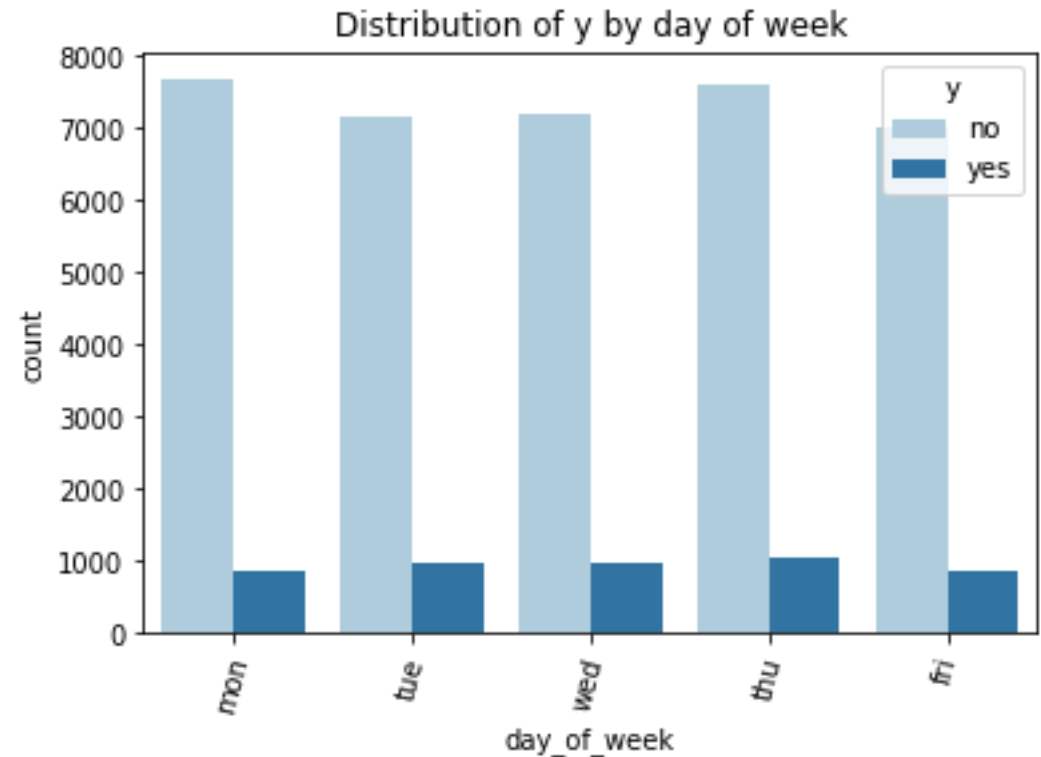
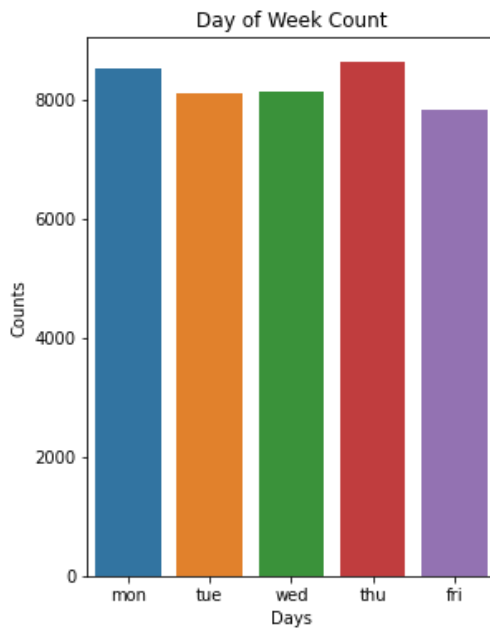
From the plots we see that when the emp.var.rate(the employment rate) is low majority of the clients subscribe to the term deposit.

POUTCOME



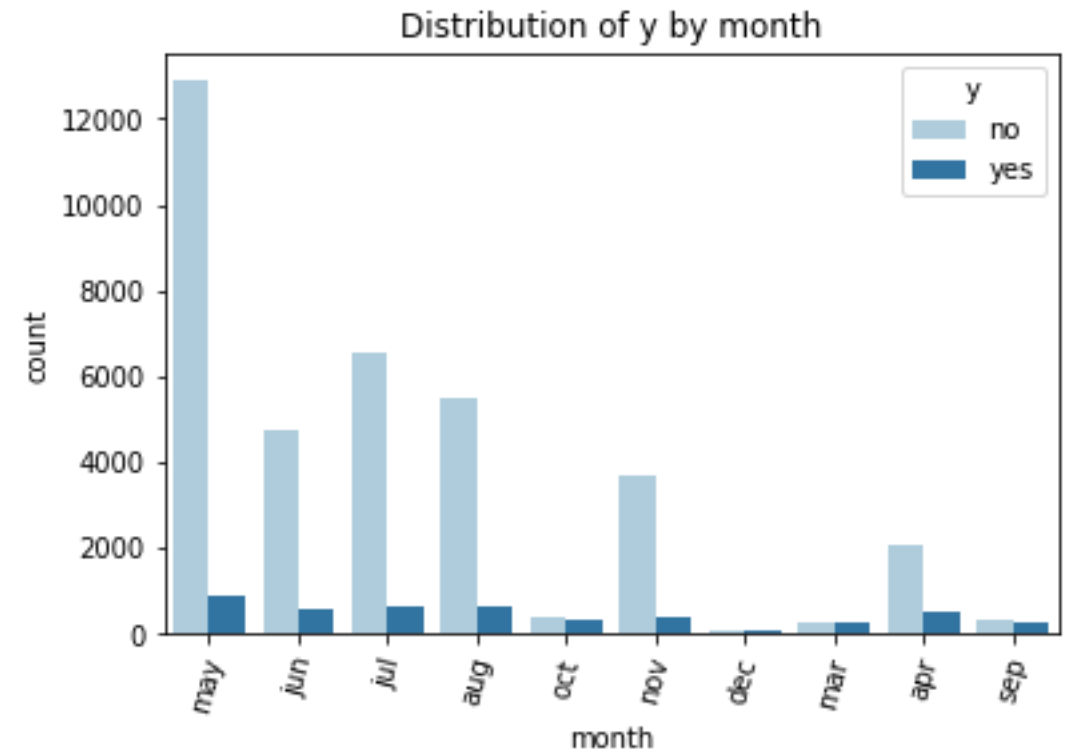
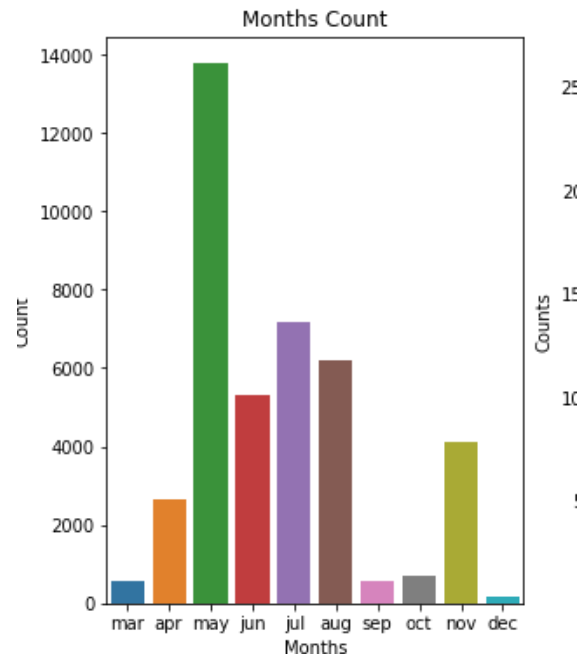
Considering the outcome of the previous campaign, majority of the clients contacted were new clients and they subscribed more compared to those who had been contacted before.

DAY OF WEEK



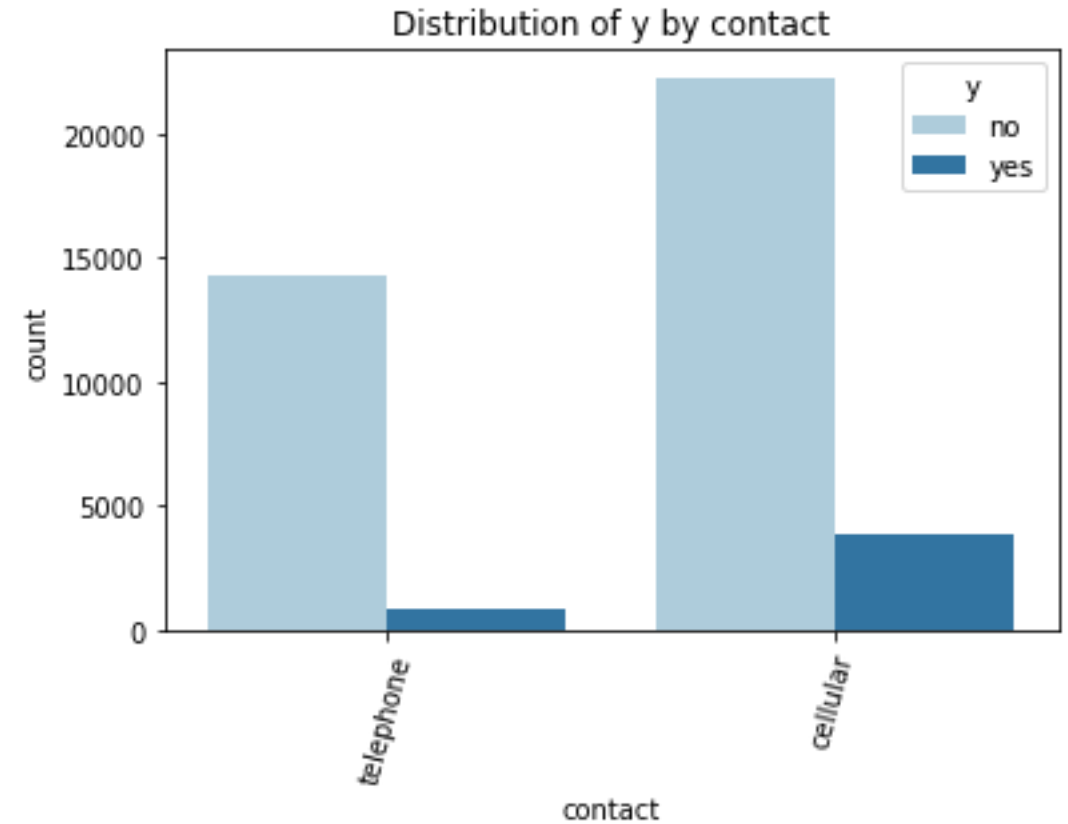
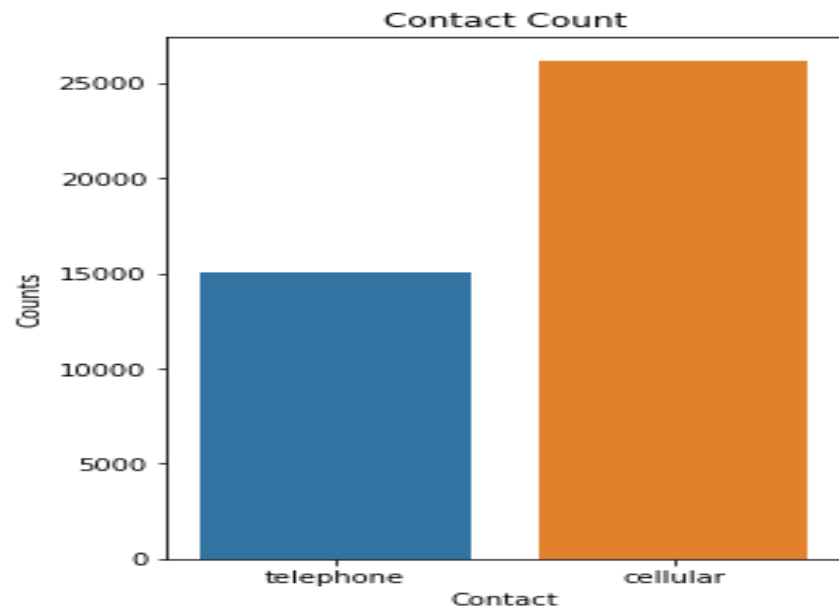
Clients were contacted on Monday, Tuesday, Wednesday, Thursday and Friday with the least number of clients being contacted on Friday. Almost an equal number of clients contacted subscribed .

MONTH



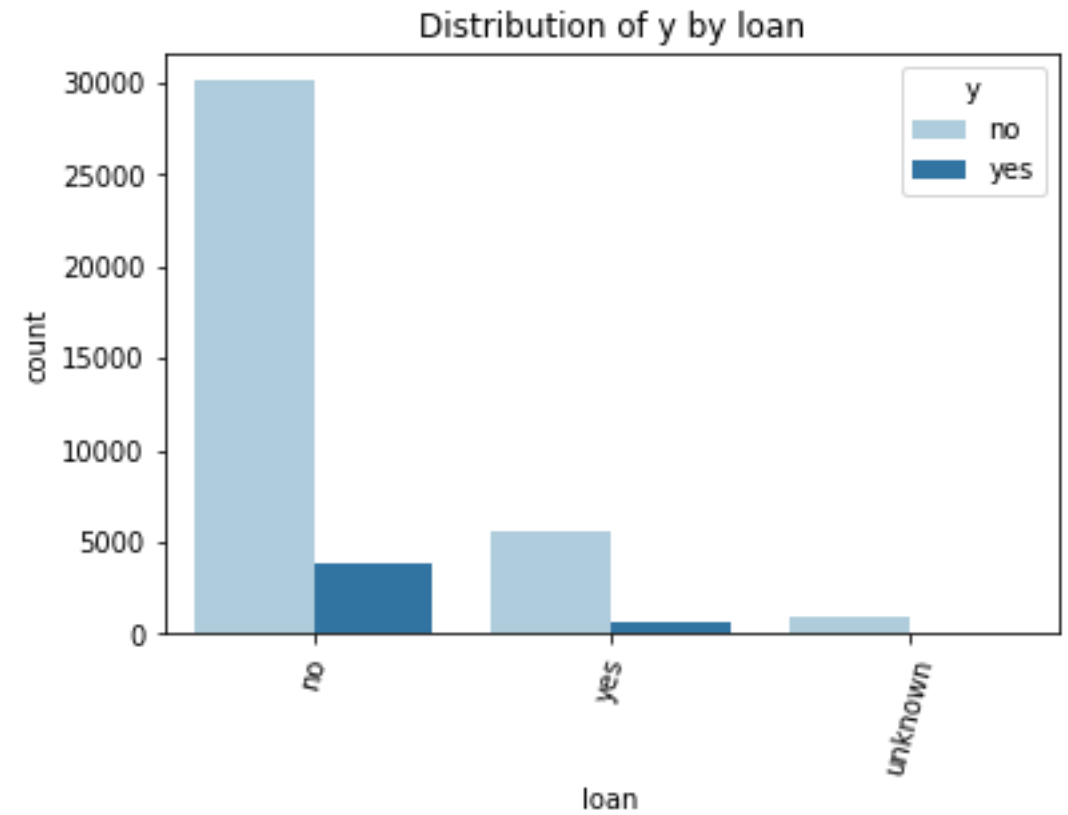
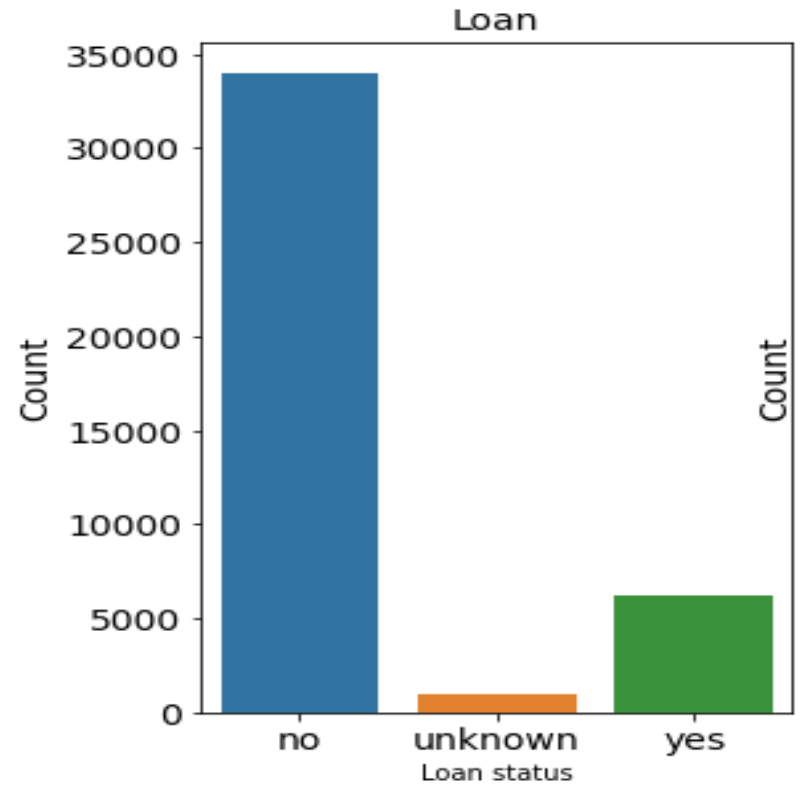
Clients were contacted most in the month of May and more clients equally subscribed in May.

CONTACT



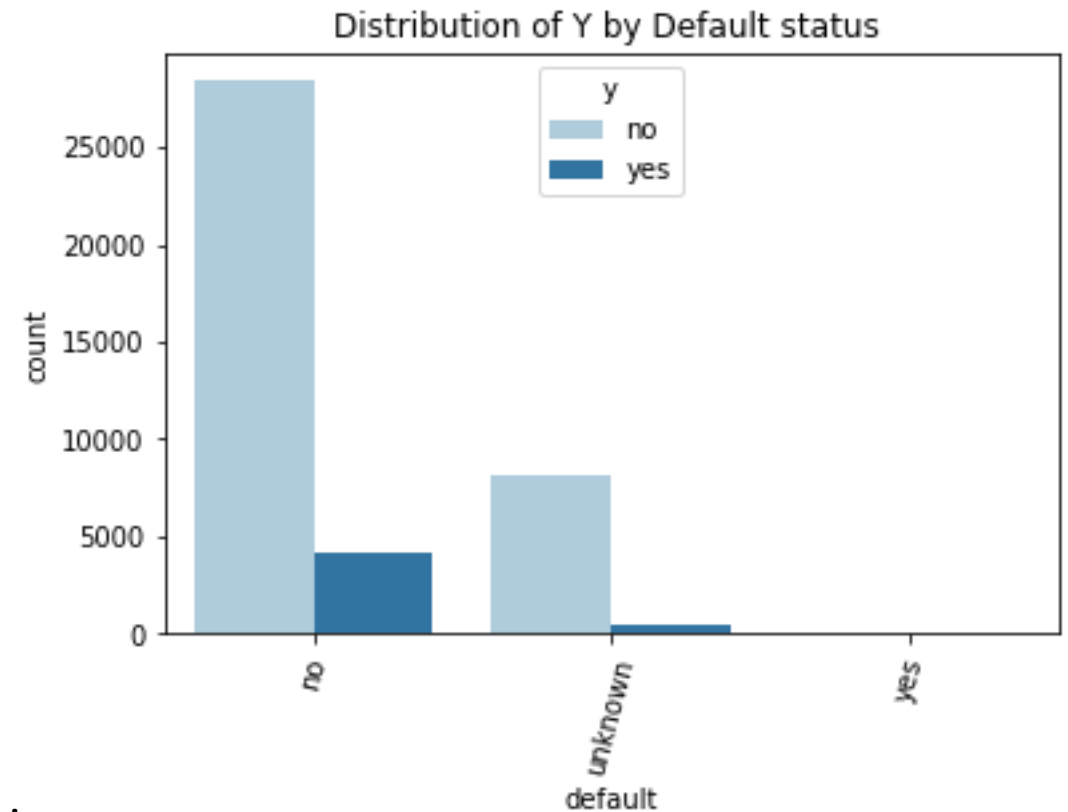
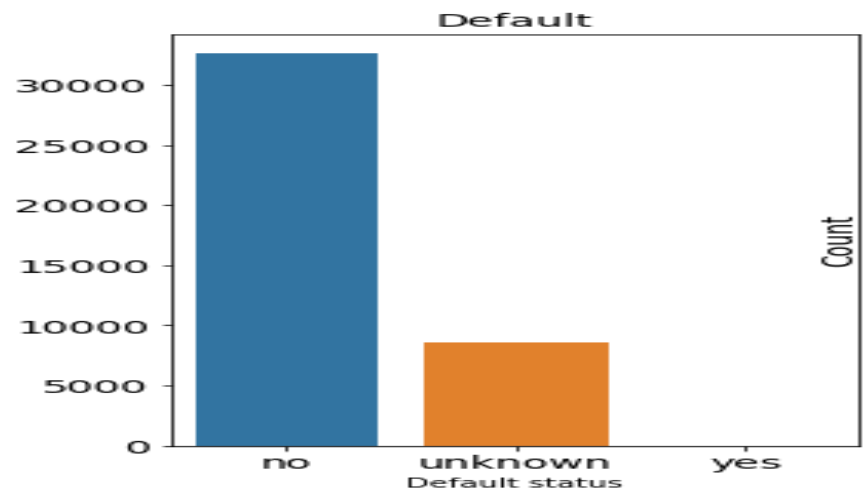
Majority of the clients were contacted by cellular and they subscribed more compared to those who used telephone.

LOAN



The bank mostly contacted clients who did not have loans.

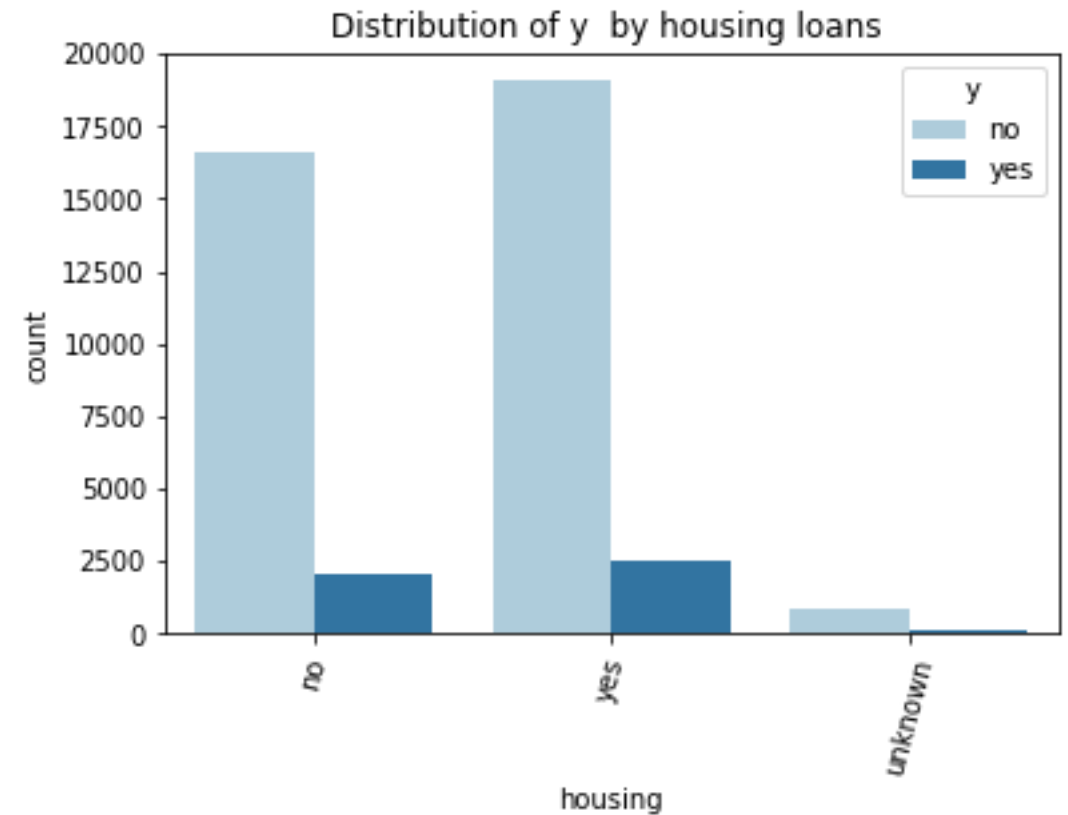
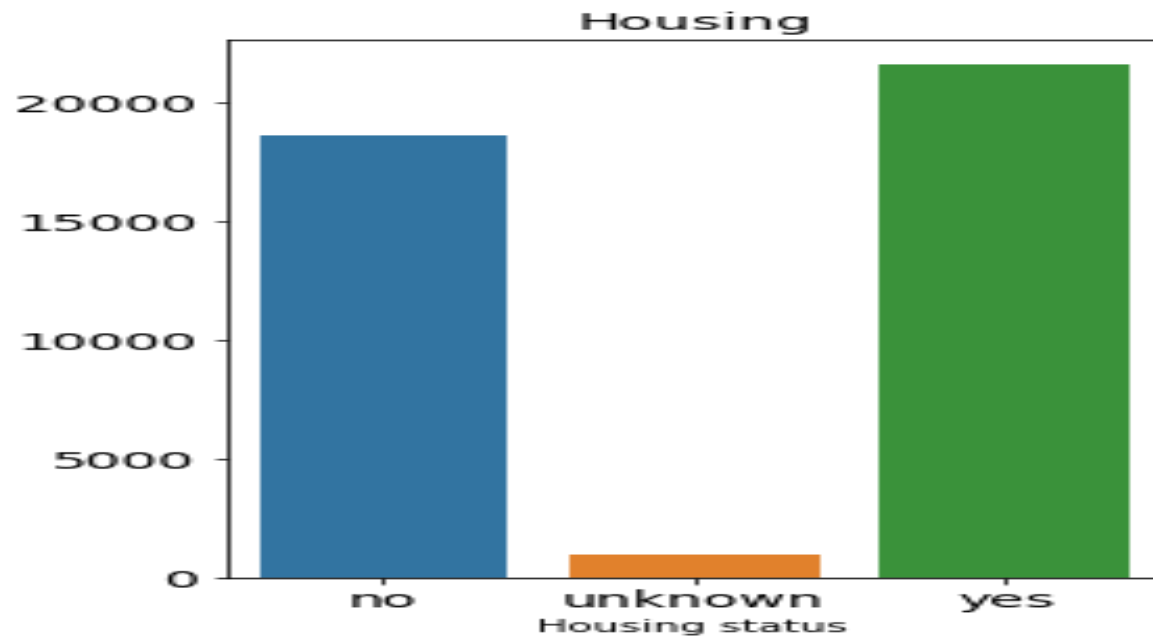
DEFAULT



The contacted mostly clients who had not defaulted .

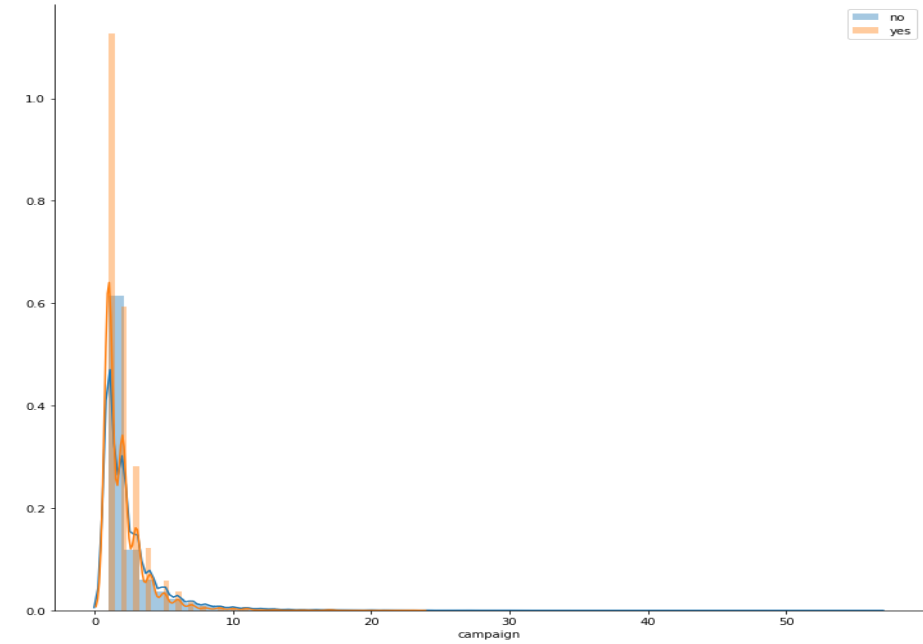
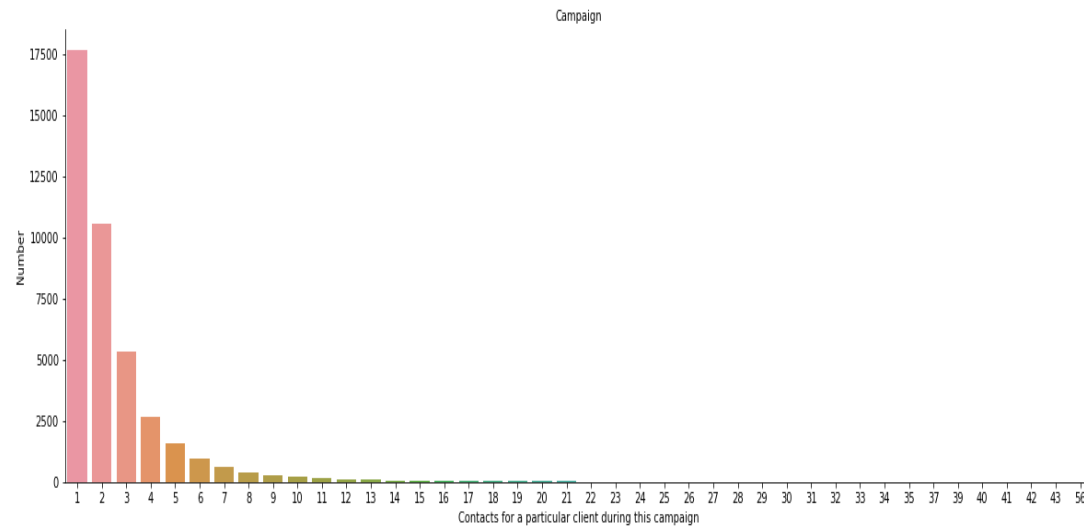
Most of the clients who had not defaulted in their credit and they subscribed more compared to those whose default status was unknown.

HOUSING



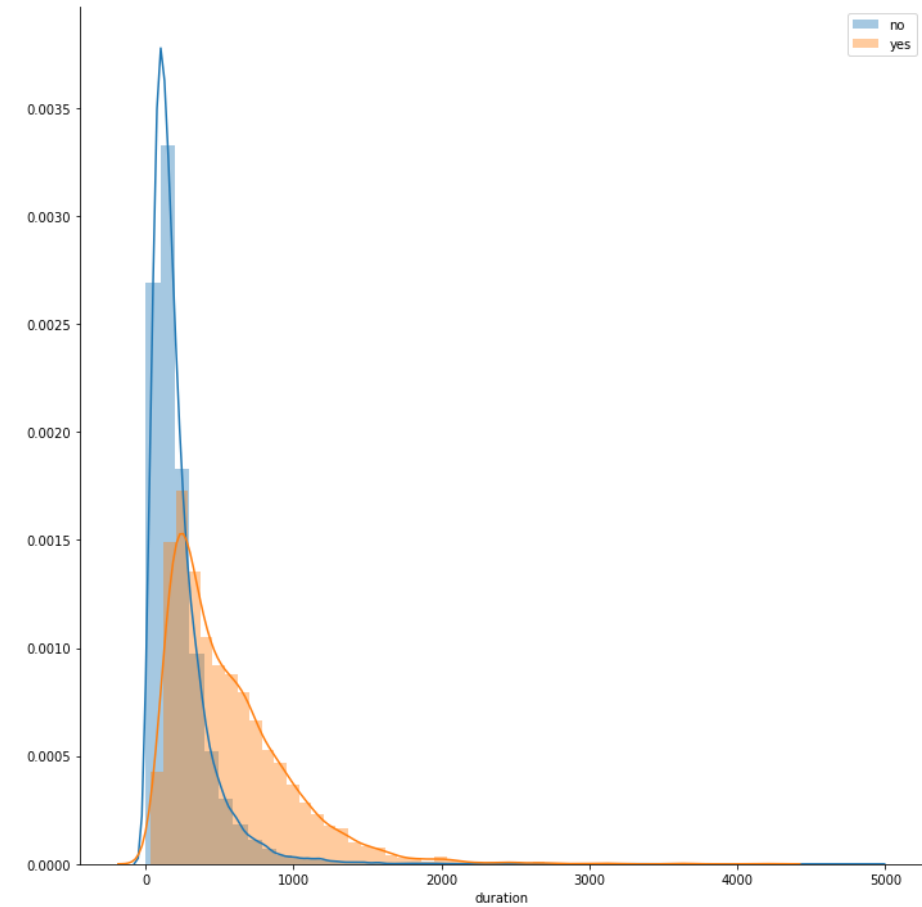
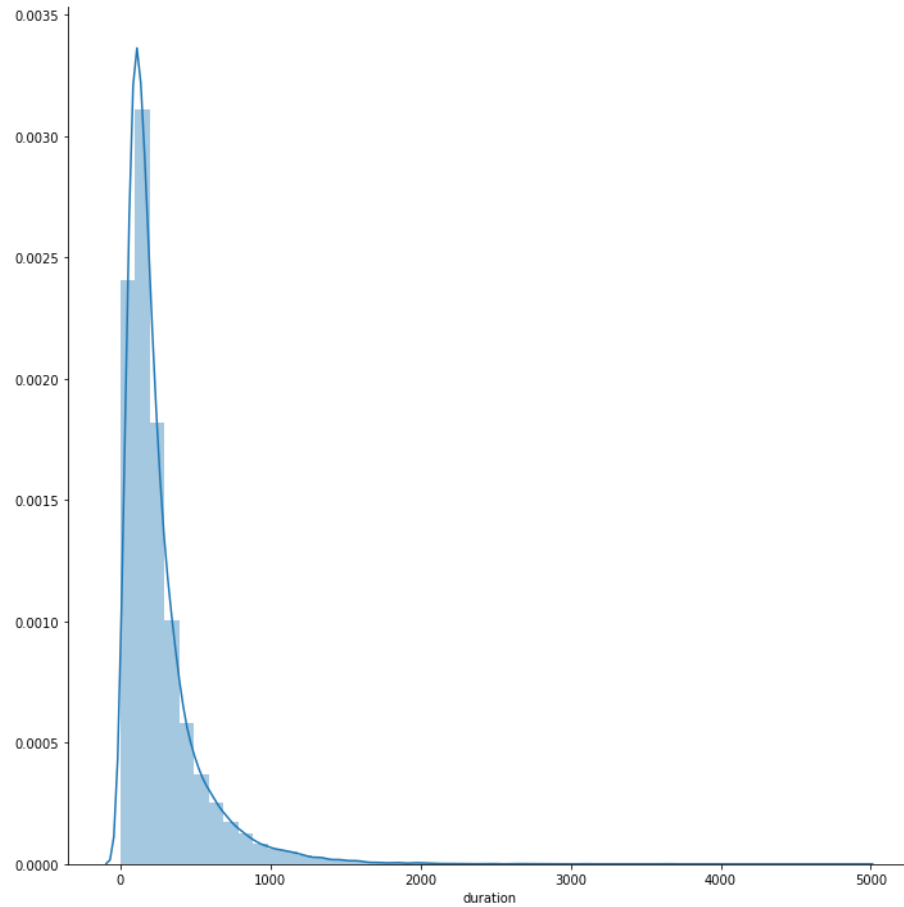
From the plots majority of clients had housing loans .
Those who had loans subscribed slightly more than those who didn't have loans.

CAMPAIGN

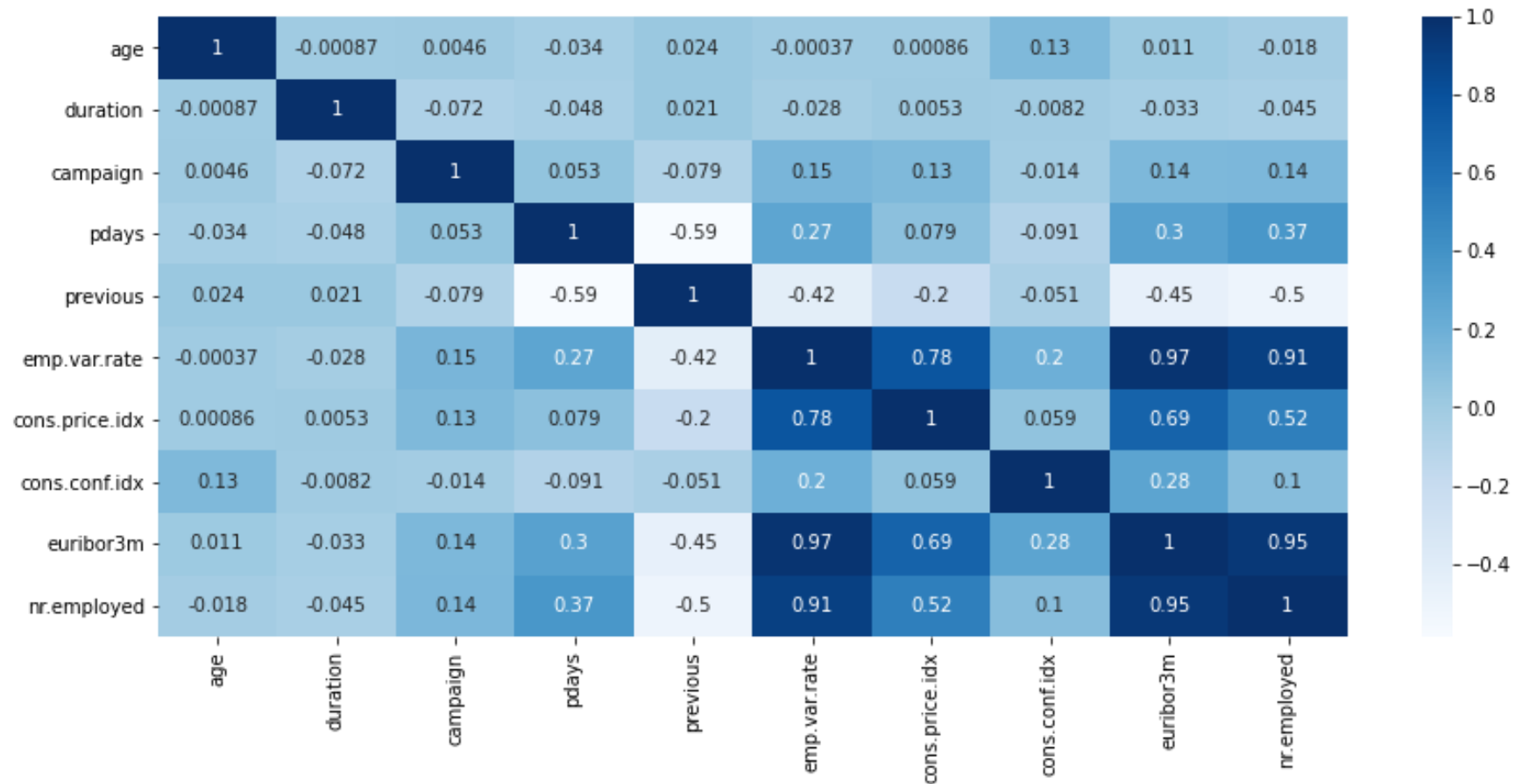


Majority of the clients were only contacted once during this campaign.

DURATION



CORRELATION MATRIX



Emp.var. Rate, euribor3m and nr.employed have very high correlation

RECOMMENDED MODELS

- Logistic Regression
- Decision Tree
- Random Forest

Thank You