Outline

INTRODUCTION

DATA INFORMATION

EXPLORATORY DATA ANALYSIS

MODEL BUILDING RECOMMENDATION

INTRODUCTION

Problem Statement

- ABC Bank wants to sell its term deposit product to customers, but before doing so, they want to create a model that will help them understand whether a specific customer will buy their product or not (based on the customer's previous interactions with the bank or another financial institution)

- <u>Objective:</u> Provide insights to assist ABC bank in shortlisting customers with a higher likelihood of purchasing the product so that their marketing channel can focus solely on those customers with a greater chance of purchasing the product

- The analysis was divided into three parts:
  - ➢ Data Understanding
  - ➢ Visualization
  - ➢ Recommendations for Model Building

DATA INFORMATION

Data Background

Investigate and analyze two datasets to achieve insights within the banking industry:

- bank-additional-full.csv : All examples with 41188 observations and 20 features, dating from May 2008 to November 2010

- bank-additional.csv : 10% of the examples (4119) randomly selected from bank-additional-full.csv, and 20 features

Since bank-additional-full.csv is the original dataset and bank-additional.csv is a subset of the dataset we will remove the examples from the bank-additional-full.csv that are present in bank-additional.csv

Data Background

Assumptions:

- ❖ Clients were <u>not</u> previously contacted by the bank from a previous campaign

- ❖ Number of contacts performed before this campaign for the clients is <u>0</u>

- ❖ Take into consideration the assumptions that are followed with model building
  - ❖ May affect the results we've obtained and may need further investigation, however, can be used as a baseline for understanding the Banking Industry
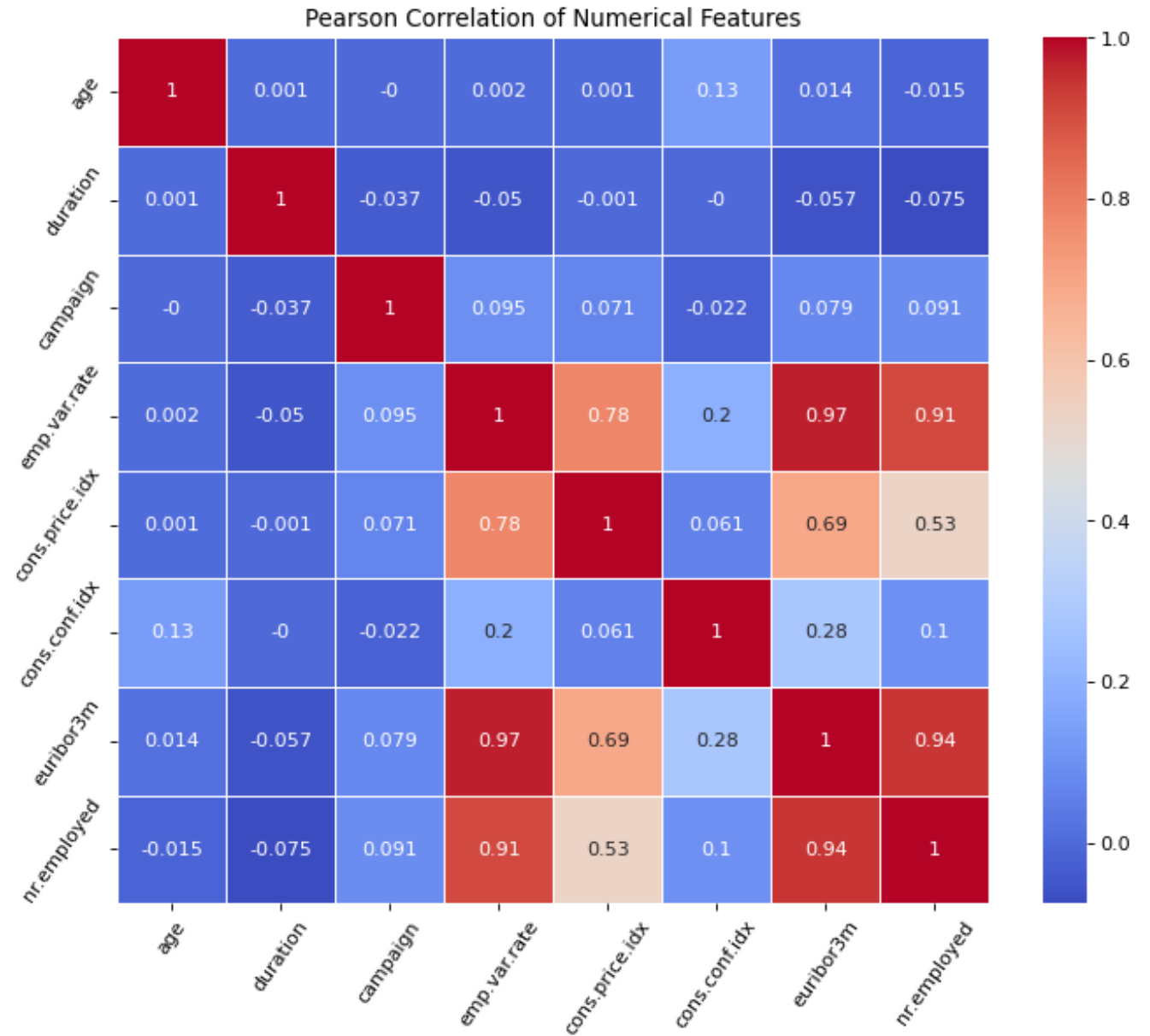
# EXPLORATORY DATA ANALYSIS

Correlation Between Variables

- Social and economic traits are all strongly correlated with one another, apart from the consumer confidence index

Pearson Correlation of Numerical Features

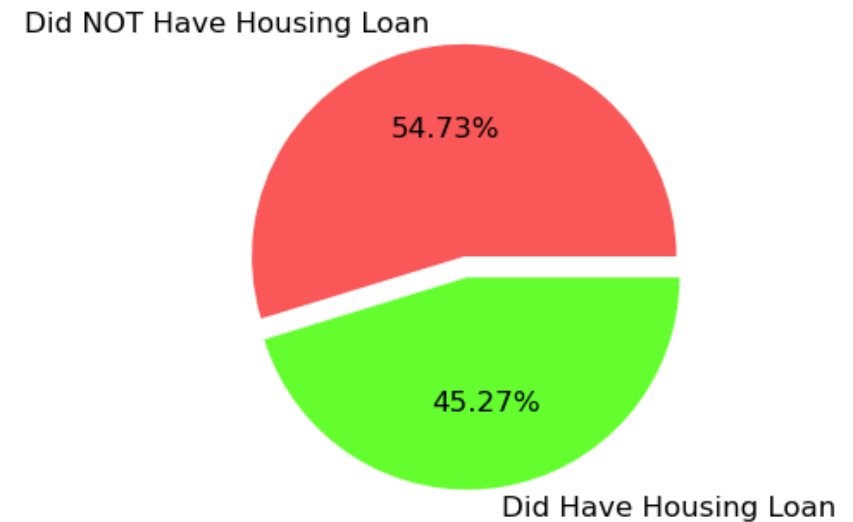# % of Loaning Distribution

- Overall, most clients have no personal or housing loans

- Pretty even housing loans distribution with just 54.73% without housing loans

% of Condition of Loans

Did NOT Have Personal Loan

84.93%

15.07%

Did Have Personal Loan
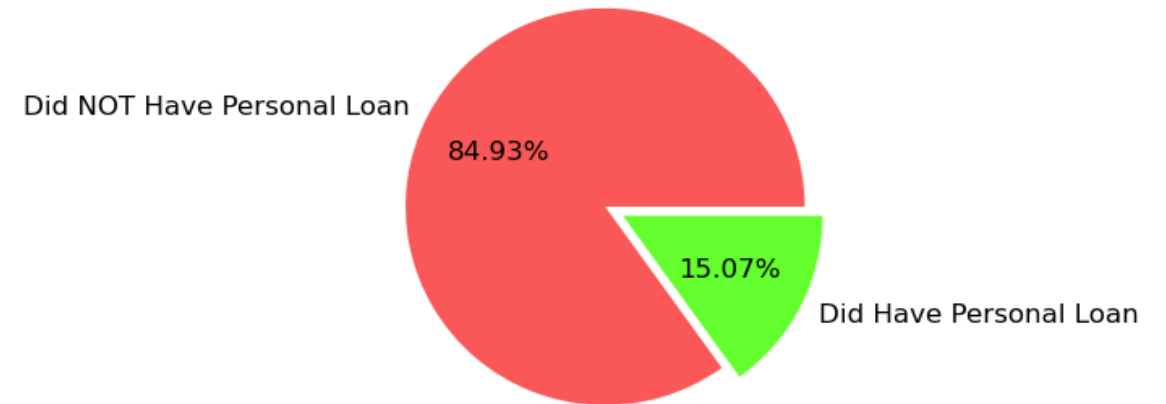
% of Condition of Housing Loan

Did NOT Have Housing Loan

54.73%

45.27%

Did Have Housing Loan

Clients' Age Distribution by Occupation

- Bulk of bank clients are between the ages of 30 to 40

- Majority of the clients come from management, followed by blue-collar workers, service providers, and technicians

- Most unlikely to use bank services is students

Age Distribution by Occupation
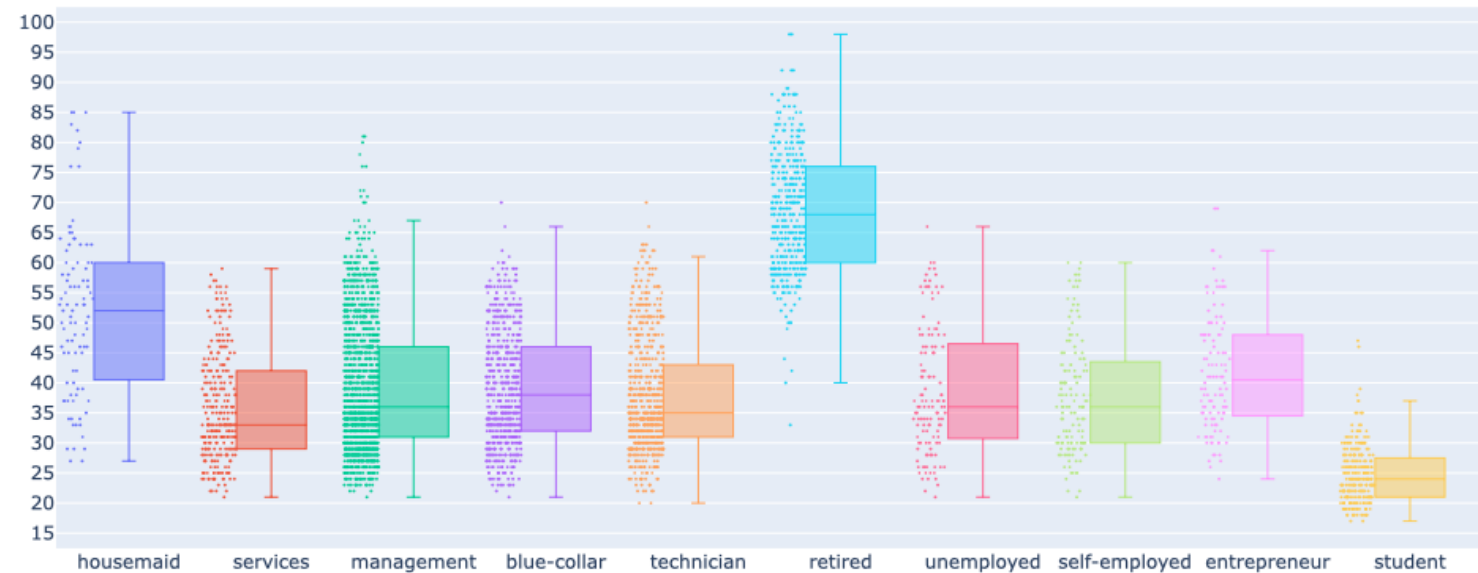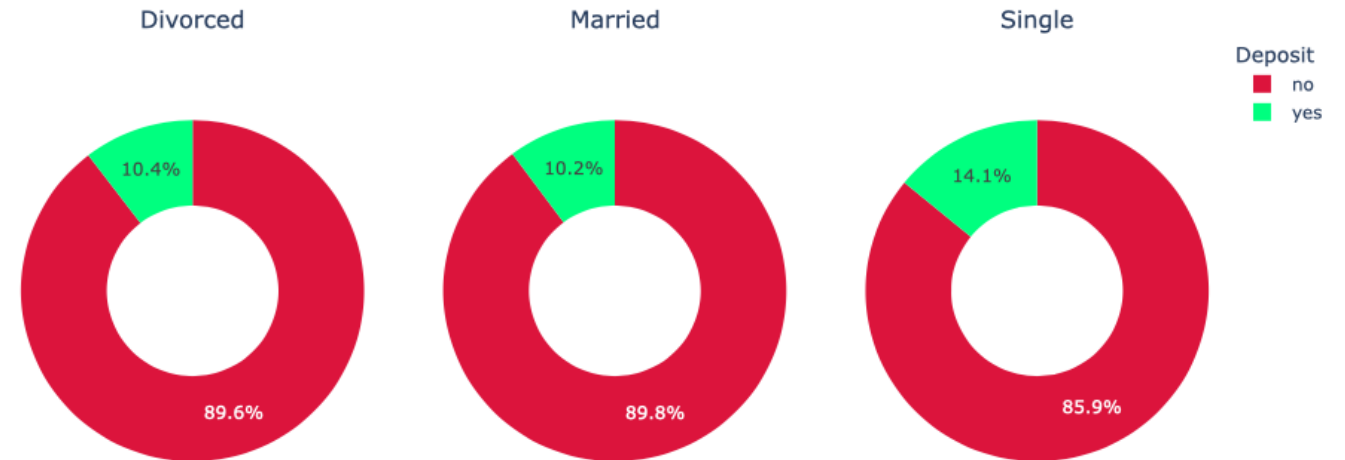
Term Depositors Distribution by Occupation

- Clients did not commit to a term deposit across all occupations

- Difference between those who subscribe and those who didn't is overwhelmingly significant, except for retirees and students

Term Depositors Based on Job Type
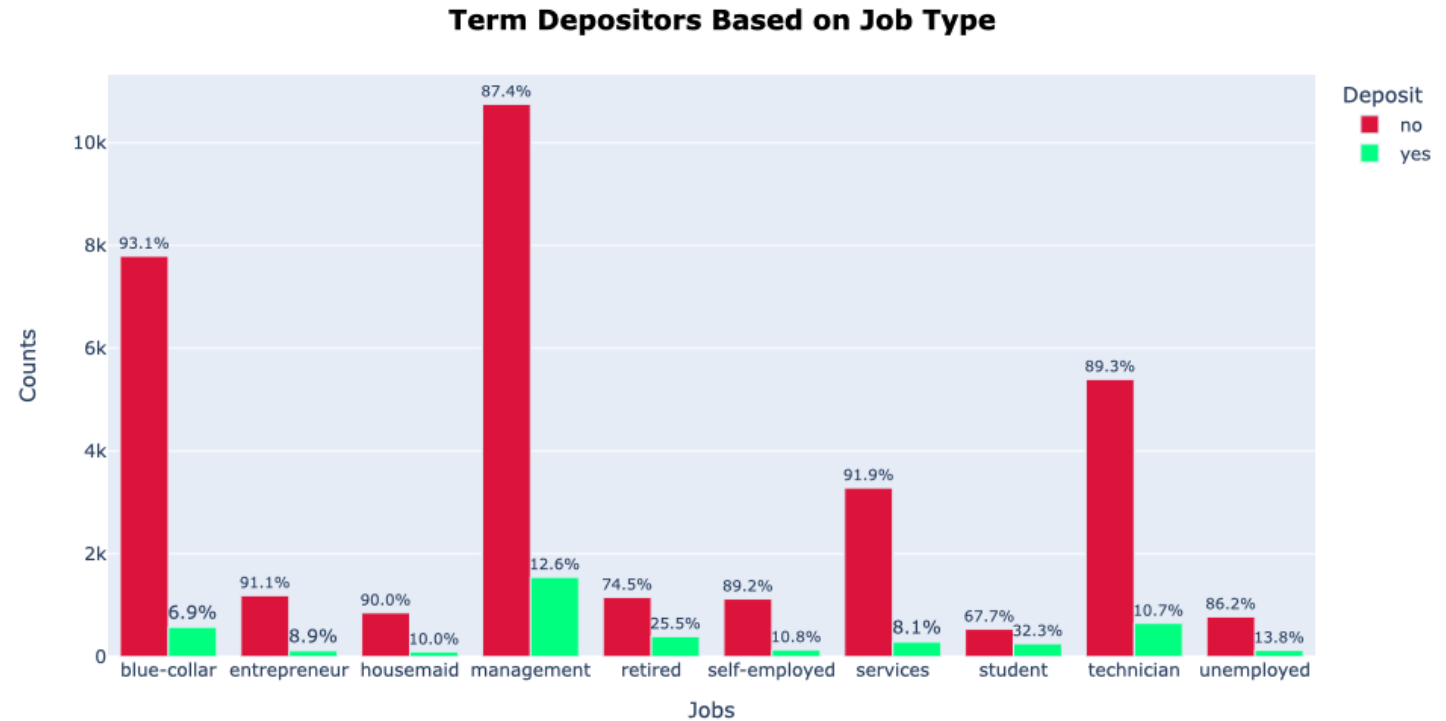
# Term Depositors Distribution by Marital Status

- Most clients by a wide margin did not subscribe to a term deposits

- About less than 2500 subscribed to a deposits for each martial status

**Term Deposits Based on Marital Status**

Term Depositors Distribution by Last Contact (Month)

- Bank contacts clients mostly in May, June, July, and August

- Very few of the clients are contacted in March, September, October, and December
  - Better chance for subscribers as there is roughly half of the clients subscribe

- Significantly more people did not sign up for a term deposit than those who did, regardless of bank interaction

Deposits Based on Last Contact
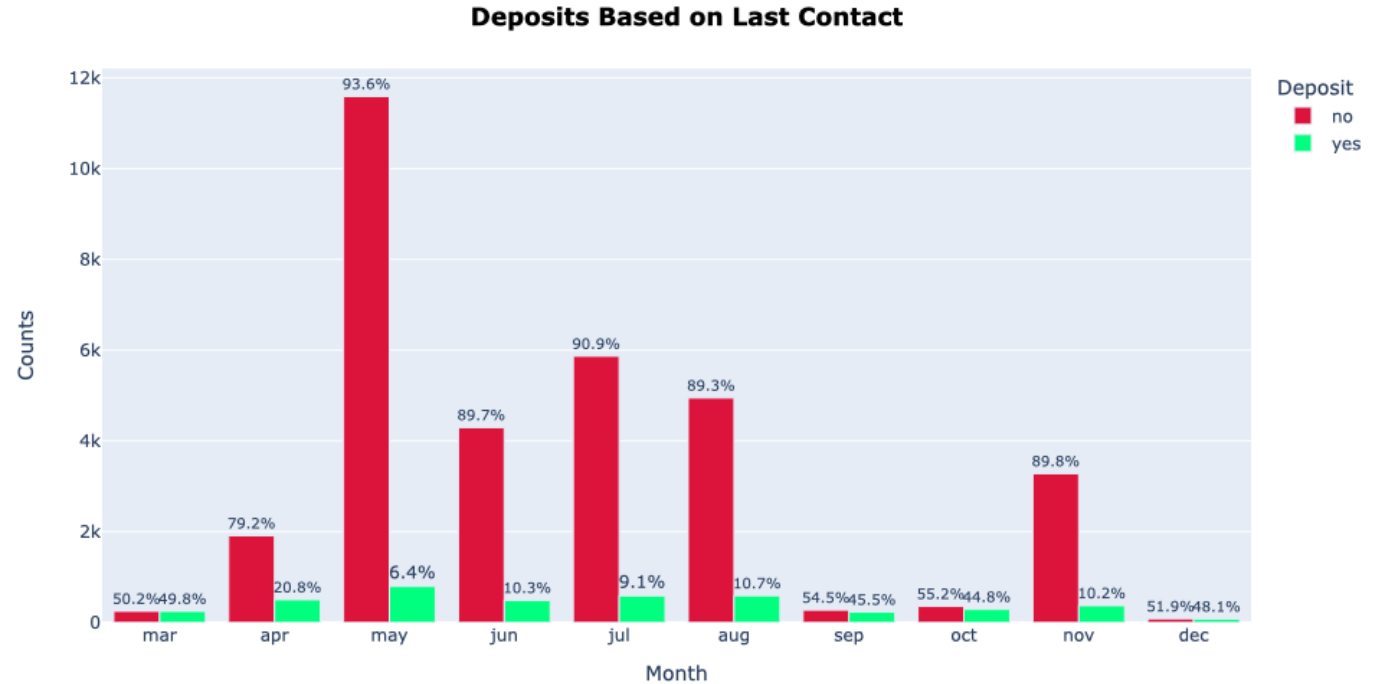
Effect of Age on Loaning History and Term Depositors

- None of the clients increased personal loan subscriptions across all age groups

- Majority of clients did not subscribe to term deposits across age groups

- Predominance of people in the mid-age group who subscribe to a deposits
  - No distinct dynamics after past 60 years of age like in the mid-age group
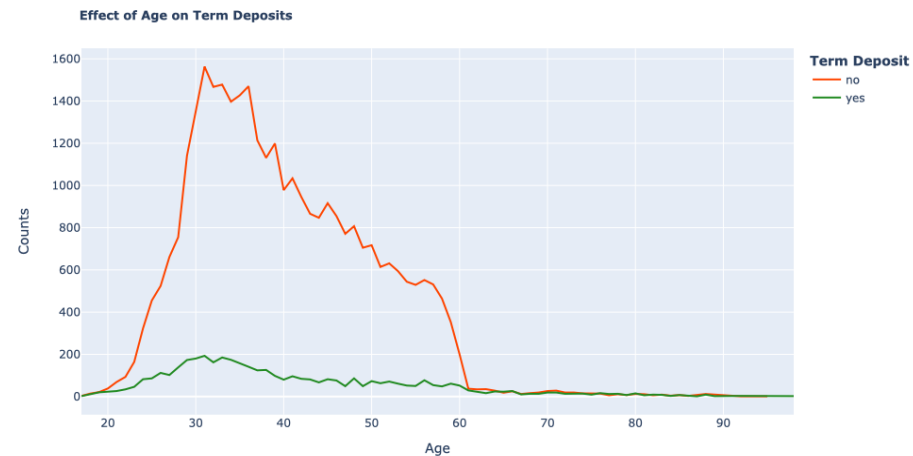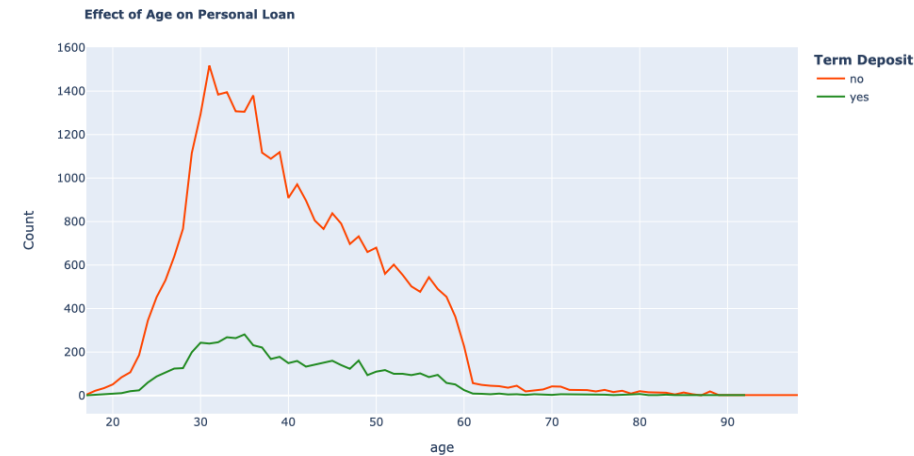
Effect of Campaign on Term Deposits

- Clients are less likely to subscribed to a term depots the more frequently the banks contact them during the campaign

Effect of Campaign on Term Deposits

Number of contacts performed during the campaign for each bank client
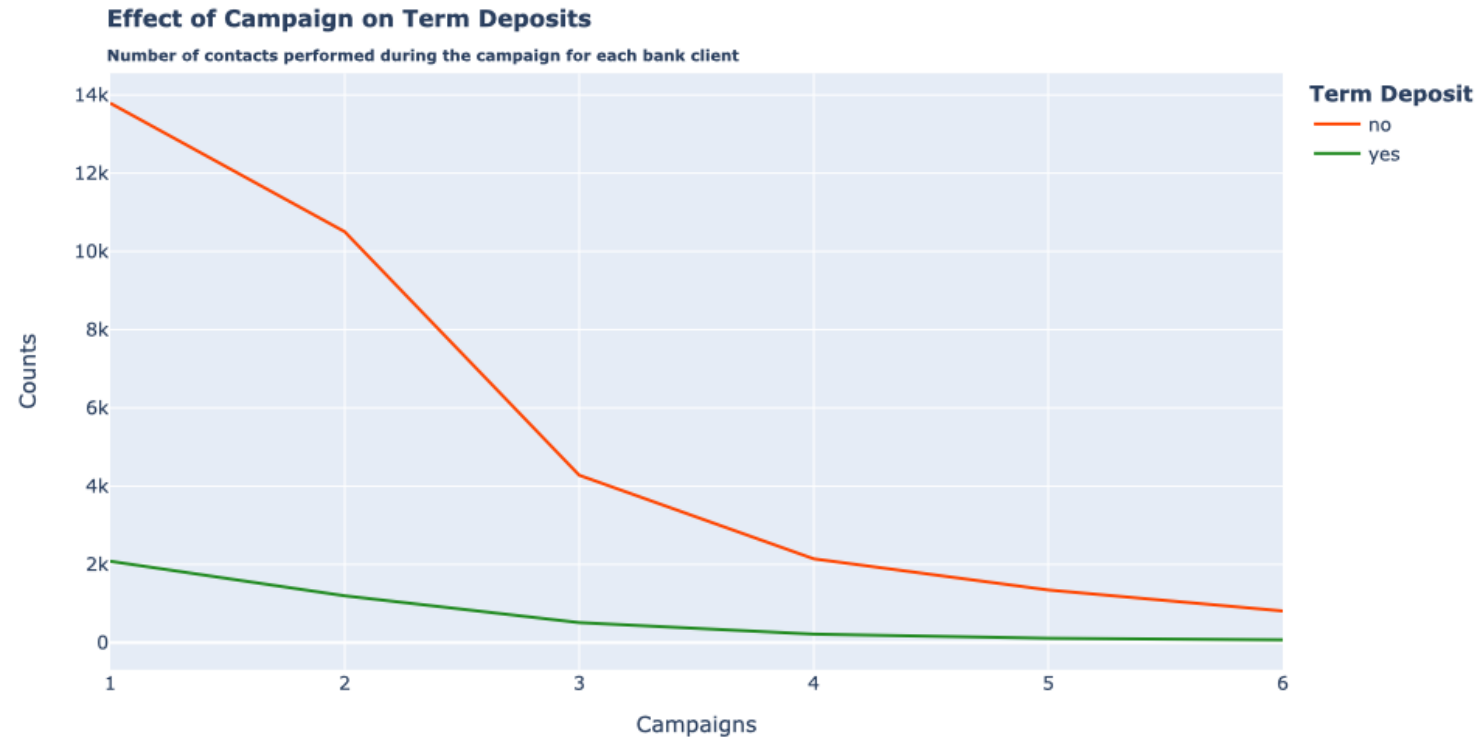
Term Deposit
— no
— yes

Term Deposits Distribution

- Great difference between those who did and did not subscribe to a term deposit

Term Deposits Distribution
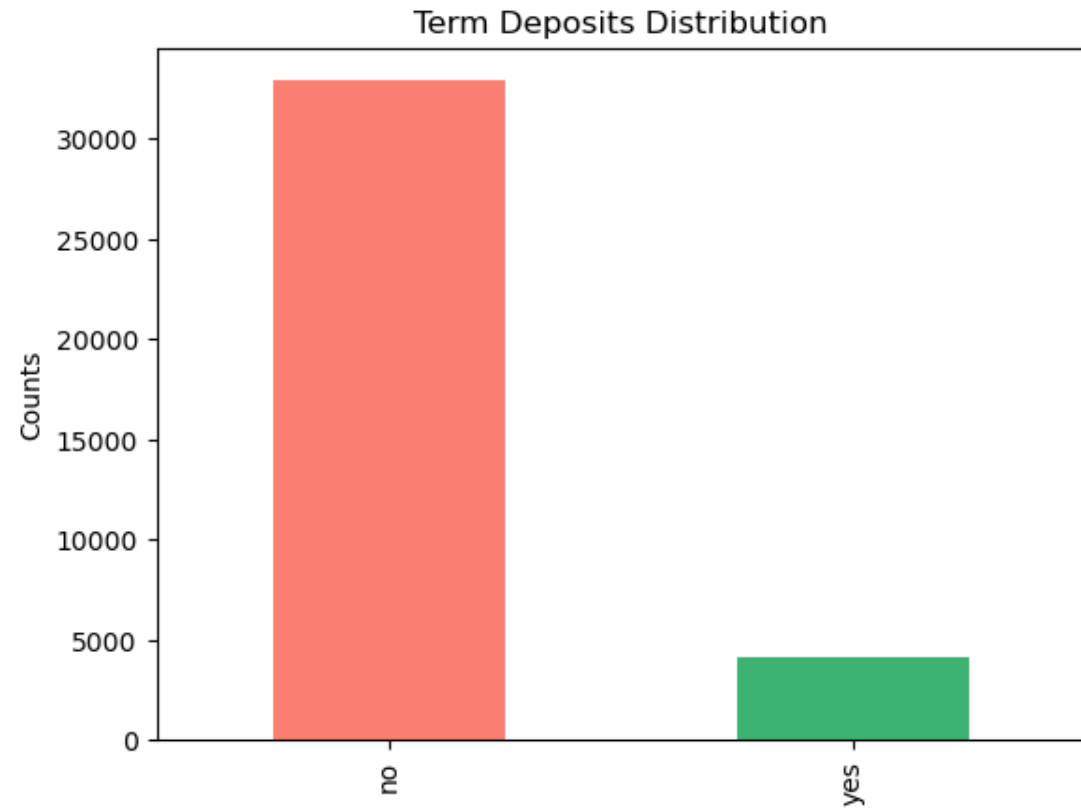
Key Observations

- Due to most clients who didn't subscribe to a term deposit, we can't really identify a distinct trait that distinguishes those of the bank's clients who did

- Because the courses were *unbalanced*, we <u>did not</u> really get a meaningful analysis because the judgments we made were heavily centered on those who did not subscribe to a term deposit instead of examining those who did

Before we start training the categorization models, we need to address this and make the necessary improvements!

We will not undertake exploratory data analysis on our original dataset after 'balancing' the classes because of the time limits for this project

MODEL BUILDING RECOMMENDATIONS

Model Building Objective

- Term Deposits will be treated as our targeted feature

- Not interested in the relationship between term deposits and the other features
  - However, it will help get a baseline of what sort of customers the banks should be keeping an eye for

- More interested in wanting to predict term deposits to shortlist customers to buying more of the bank products

Recommendations

Since our goal is to have ABC Bank to be able to shortlist their clients through machine learning models, we will concentrate on developing the following classification models:

- ➢ Logistic Regression
- ➢ Random Forest
- ➢ Boosting

We may build more classification models depending if the timing allows us to do so

THANK YOU

Data Glacier
Your Deep Learning Partner