Data Glacier

Your Deep Learning Partner

Banking Marketing Campaign Insights

Data Science Virtual Internship

Submitted by: Elissa Kuon

Date: March 30, 2023

Outline

INTRODUCTION

DATA INFORMATION

EXPLORATORY DATA ANALYSIS

MODEL BUILDING & ASSESSMENT

FINAL STATEMENT

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)
- Objective: 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

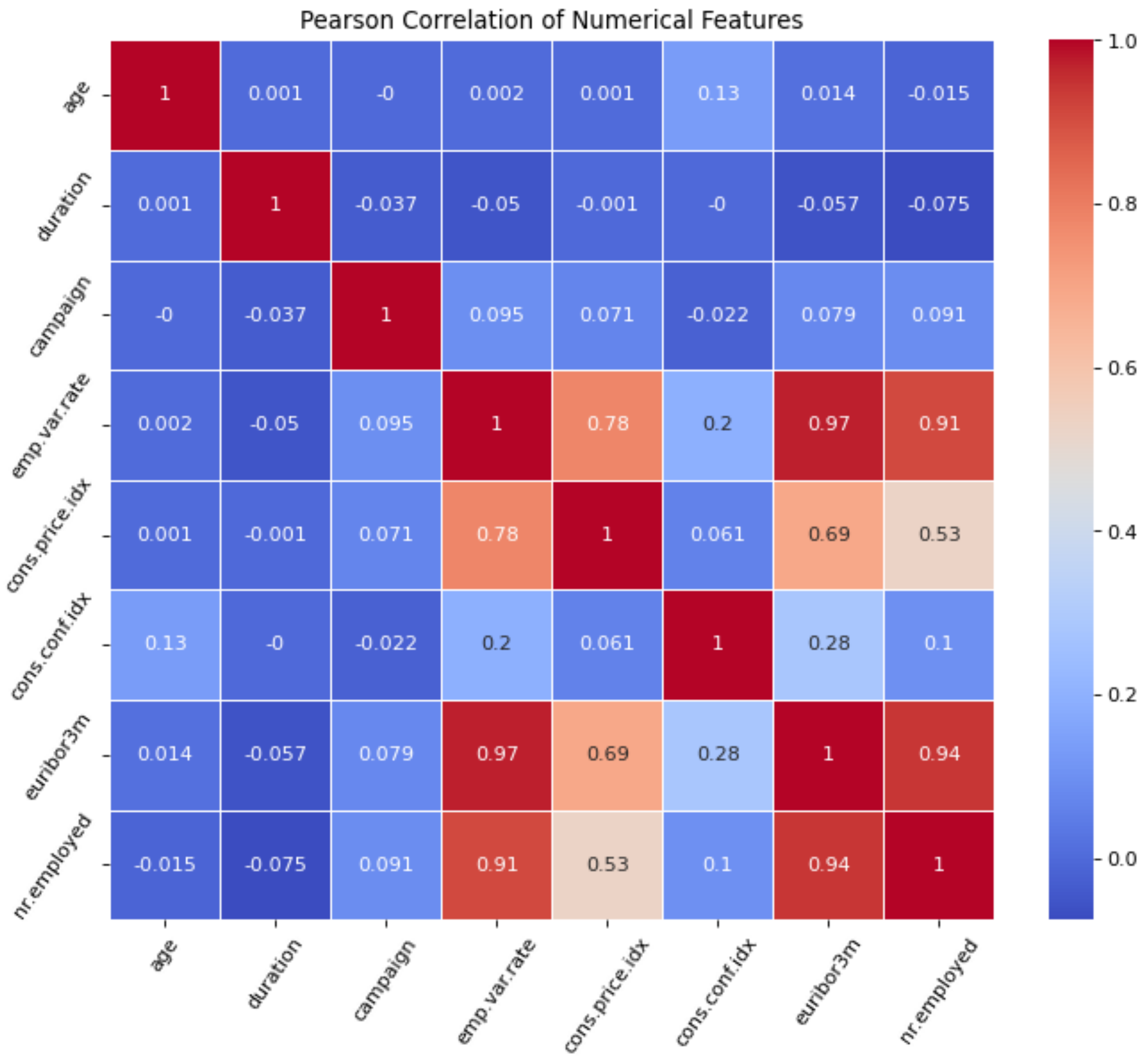
Assumptions:

- 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

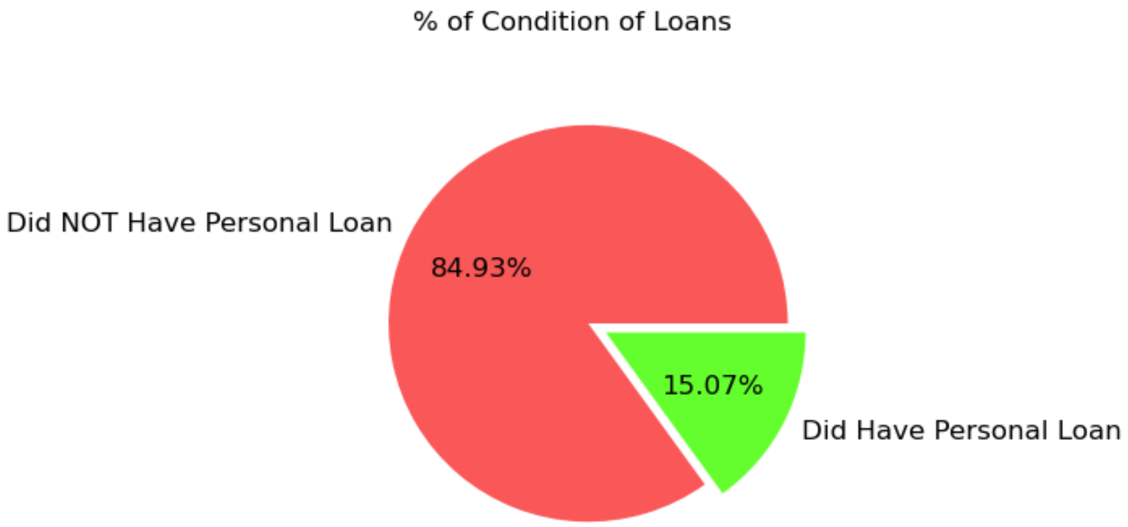
Correlation Between Variables

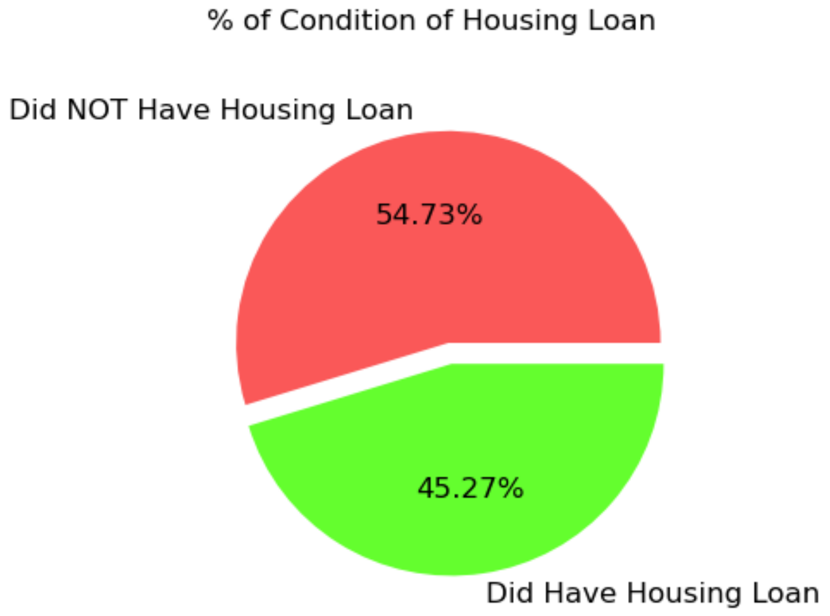
- Social and economic traits are all strongly correlated with one another, apart from the consumer confidence index



% of Loaning Distribution

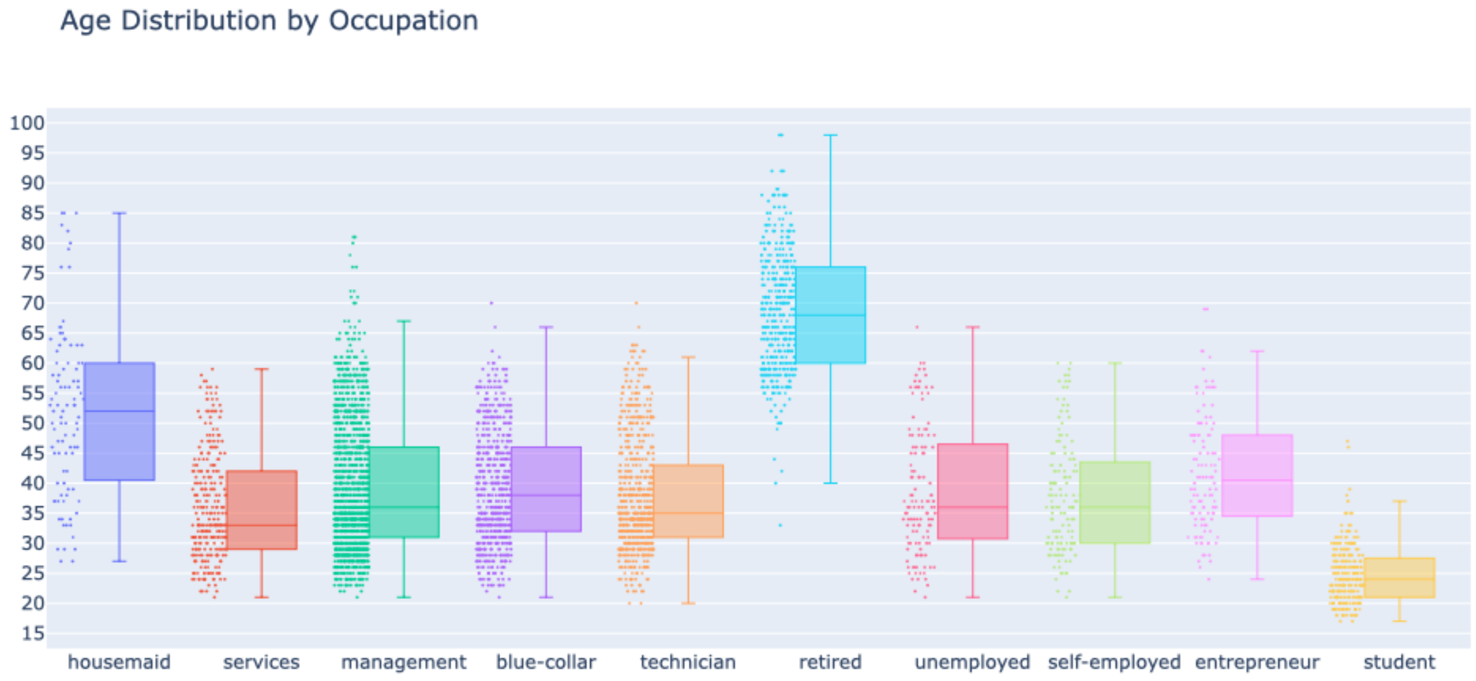
- Overall, most clients have no personal or housing loans
- Pretty even housing loans distribution with just 54.73% without housing loans





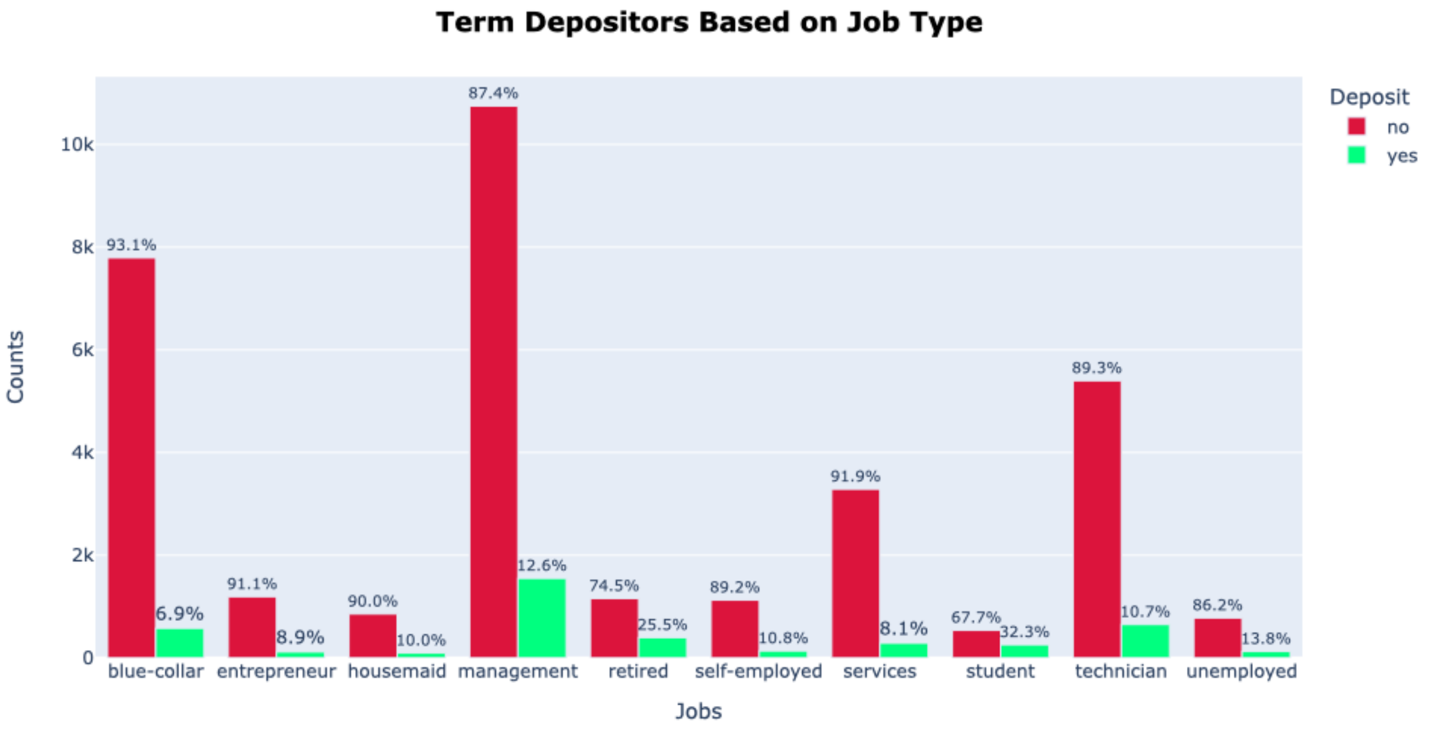
Clients' Age Distribution by Occupation

- 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



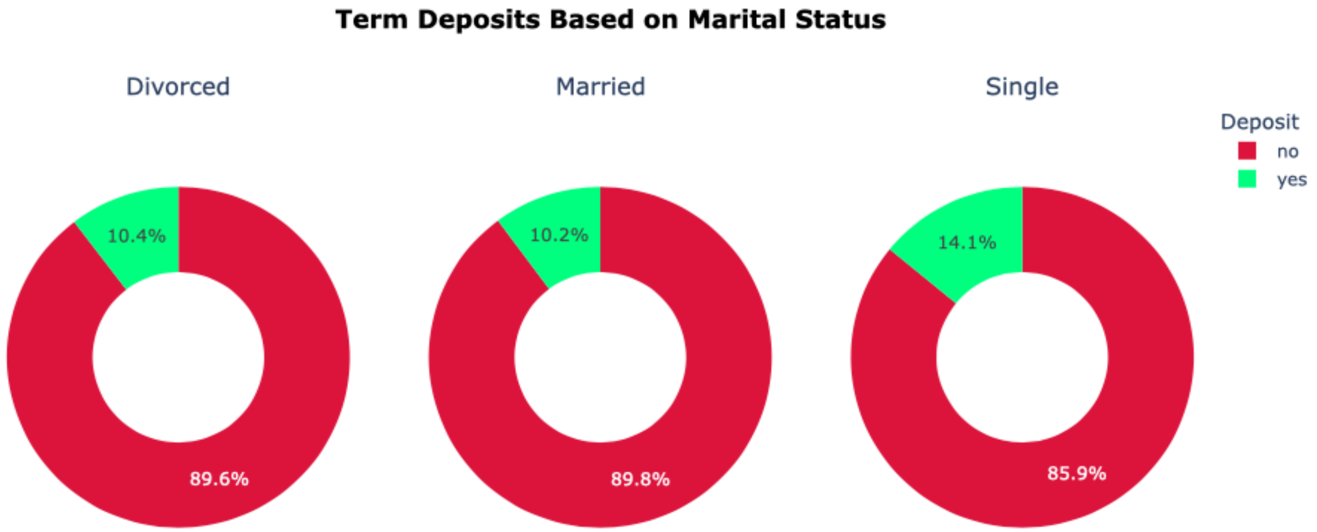
Term Depositors Distribution by Occupation

- 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 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

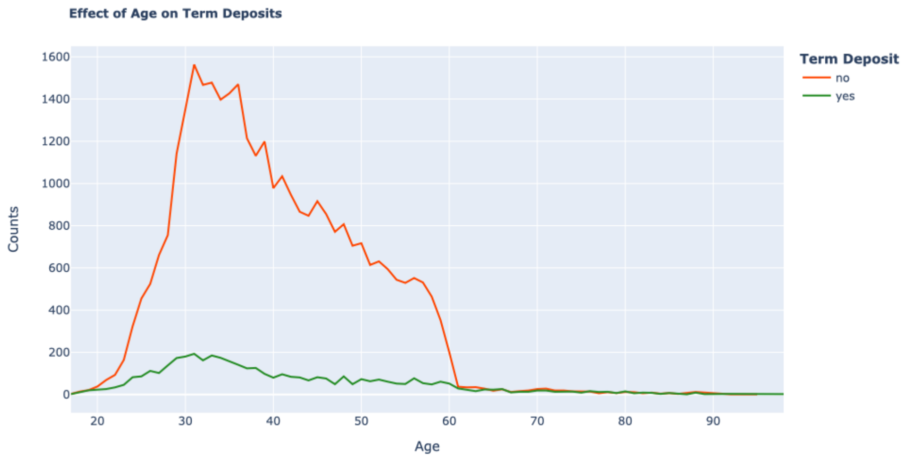


Term Depositors Distribution by Last Contact (Month)

- 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

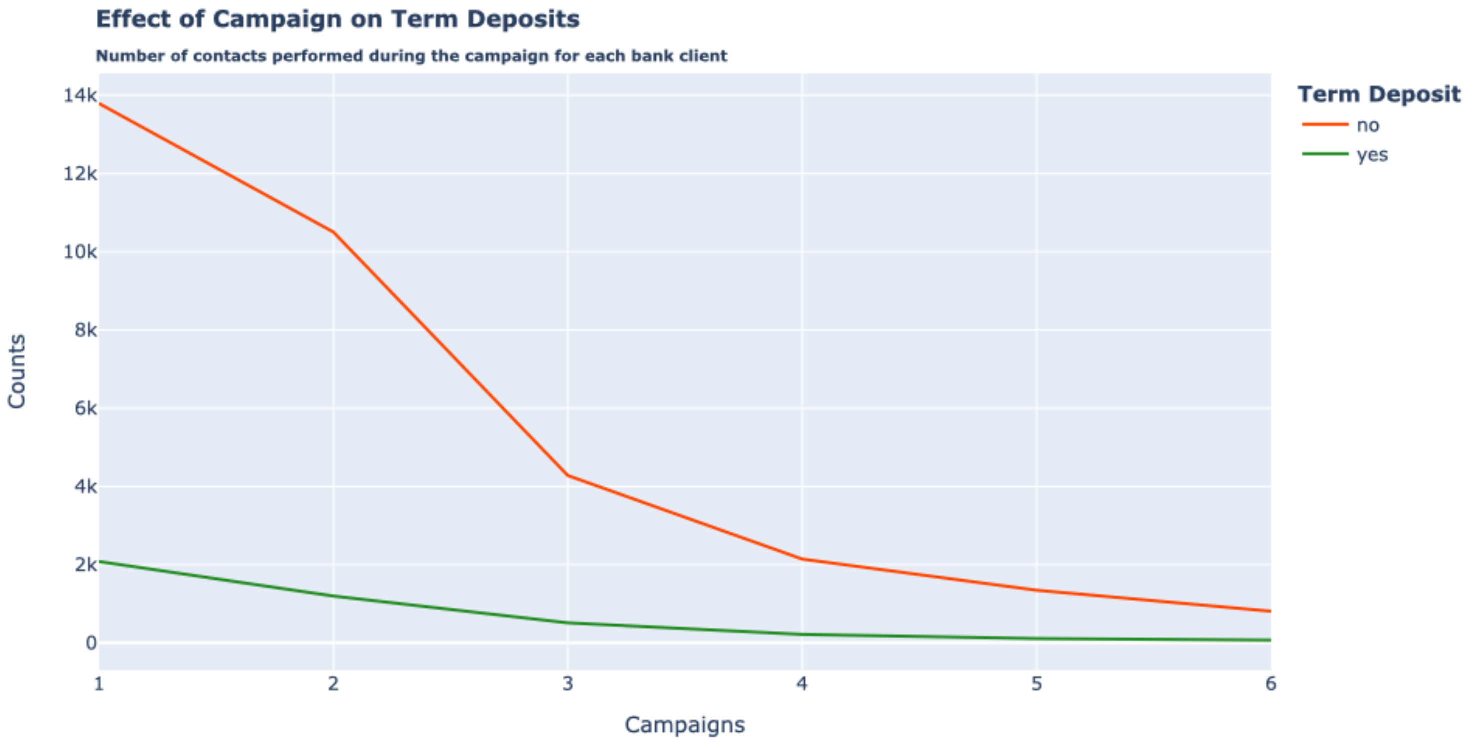
Effect of Age on Loaning History and Term Depositors

- 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



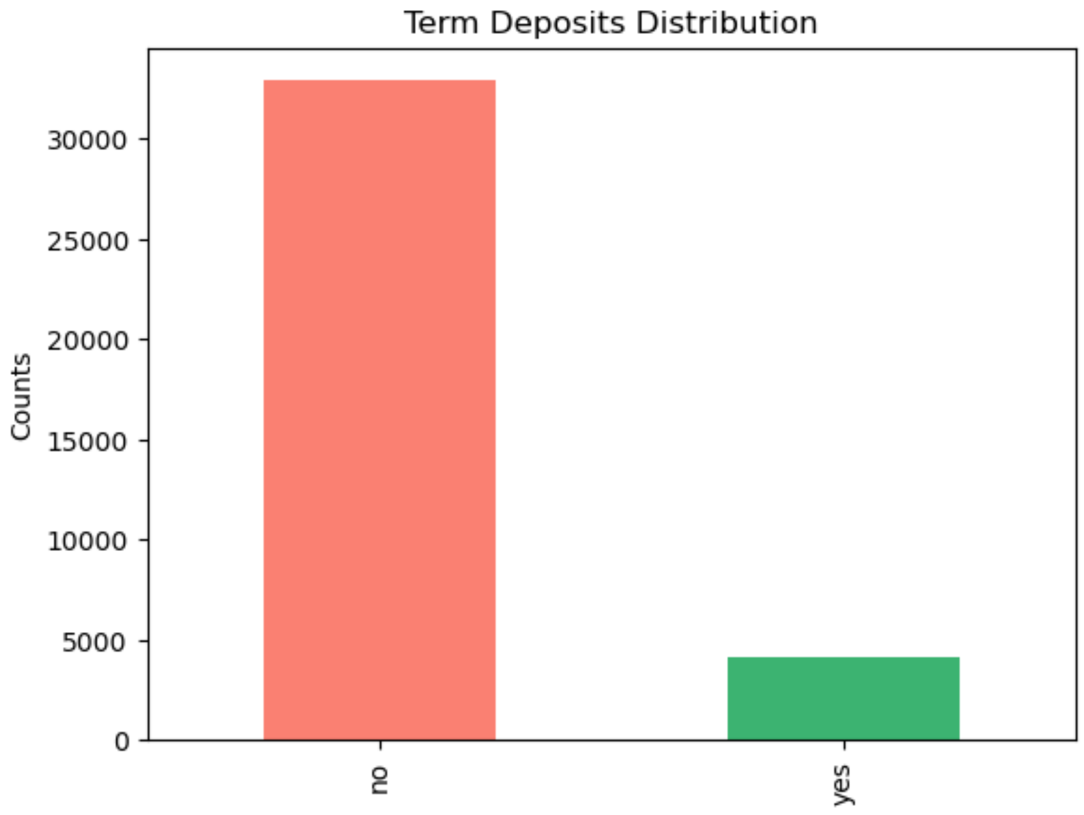
Effect of Campaign on Term Deposits

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



Term Deposits Distribution

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



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 & ASSESSMENT

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

Metrics Evaluation

Model Used:

- Logistic Regression (LR)
- Support Vector Classifier (SVC)
- Random Forest (RF)
- Gradient Boosting (GB)

With Duration	Precision Score	Recall Score	F1 Score	AUC-ROC Score	Accuracy Score
LR	0.3874	0.9157	0.5445	0.8689	0.8322
SVC	0.3660	0.9180	0.5234	0.8612	0.8169
RF	0.4113	0.9357	0.5714	0.8855	0.8463
GB	0.4149	0.9401	0.5757	0.8886	0.8483

W/O Duration	Precision Score	Recall Score	F1 Score	AUC-ROC Score	Accuracy Score
LR	0.3011	0.6208	0.4055	0.7218	0.8007
SVC	0.2665	0.6718	0.3816	0.7222	0.7616
RF	0.2724	0.6608	0.3858	0.7219	0.7696
GB	0.3500	0.6364	0.4516	0.7455	0.8308

Best Model

- Gradient Boosting is the best model because it performs best on all metrics both with and without the duration variable.
  - However, the precision score, recall score, and f1 score are low, and may consider reevaluating the model once again (by making some adjustments to the model – hyper tuning)

FINAL STATEMENT

Future Considerations

- Further investigations reveal that the social and economic characteristics have a significant impact on our best model. We may consider eliminating these traits and focusing on individual traits.
- Due to the imbalance of the classes, the judgments we made disproportionately leaned toward individuals who did not subscribe to a term deposit rather than looking at significant characteristics of those who did
  - Despite how we resolved this issue by undersampling the classes, we may consider evaluating this decision as oversampling or a mixture of both
- We may also wish to use an alternative machine learning model or further improve the gradient boosting model to provide a more thorough and accurate examination

Based on the above points, Gradient Boosting is indeed our best model from the decisions we made throughout the entire project. However, we must consider other aspects that may result in why this is our best model and consider other options.

THANK YOU

Data Glacier

Your Deep Learning Partner