Bank Marketing Prediction for future Campaigns

Project Presentation:
Arya Dhorajiya
Rijul Saini
Sumit Hawal
Spandan Maaheshwari

Introduction

- The data is related with direct marketing campaigns of a Portuguese banking institution.
- The marketing campaigns were based on phone calls.
- Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be (‘yes’) or not (‘no’) subscribed.”

Goal: Predict if the client will subscribe (yes/no) a term deposit

Business Understanding

- The Bank wants to market one of its products - Term deposit. Term deposits are an extremely safe investment and are therefore very appealing to conservative, low-risk investors

- Instead of mass marketing, the bank has opted to be more proactive in identifying potential buyers and contact the customer directly

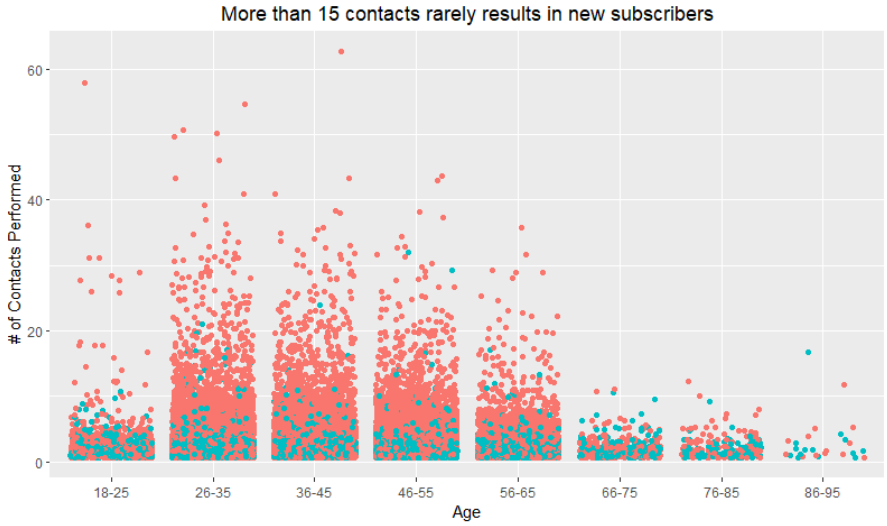
- The goal of this project is to perform post-campaign analytics to identify the potential subscribers of the term deposit product for future campaigns

Data Description

45212 instances / 16 inputs

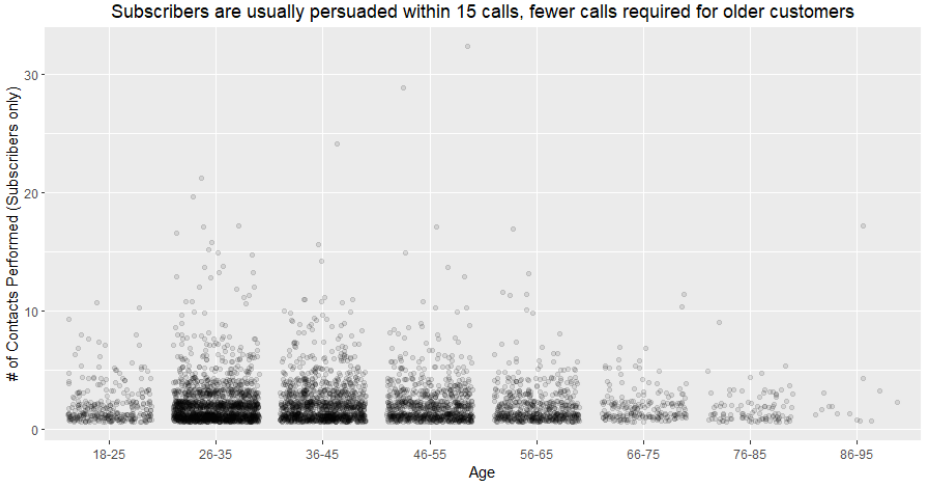
Attribute	Data Type	Description
Age	Numeric	Age of client
Job	Categorical	Type of job ("admin.","unknown","unemployed","management",etc.)
Marital	Categorical	Marital status ("married","divorced","single")
Education	Categorical	Level of education ("unknown","secondary","primary","tertiary")
Default	Binary	Has credit in default? ("yes","no")
Balance	Numeric	Average yearly balance (in Euros)
Housing	Binary	Has housing loan? ("yes","no")
Loan	Binary	Has personal loan? ("yes","no")
Contact	Categorical	Contact communication type ("unknown","telephone","cellular")
Day	Numeric	Last contact day of the month
Month	Categorical	Last contact month of year
Duration	Numeric	Last contact duration (in seconds)
Campaign	Numeric	Number of contacts performed during this campaign and for this client
Pdays	Numeric	Number of days since the client was last contacted from a previous campaign
Previous	Numeric	Number of contacts performed before this campaign and for this client
Poutcome	Categorical	Outcome of the previous marketing campaign ("unknown","other","failure","success")
Y	Binary	Has the client subscribed a term deposit? ("yes","no")

EXPLORATORY DATA ANALYSIS

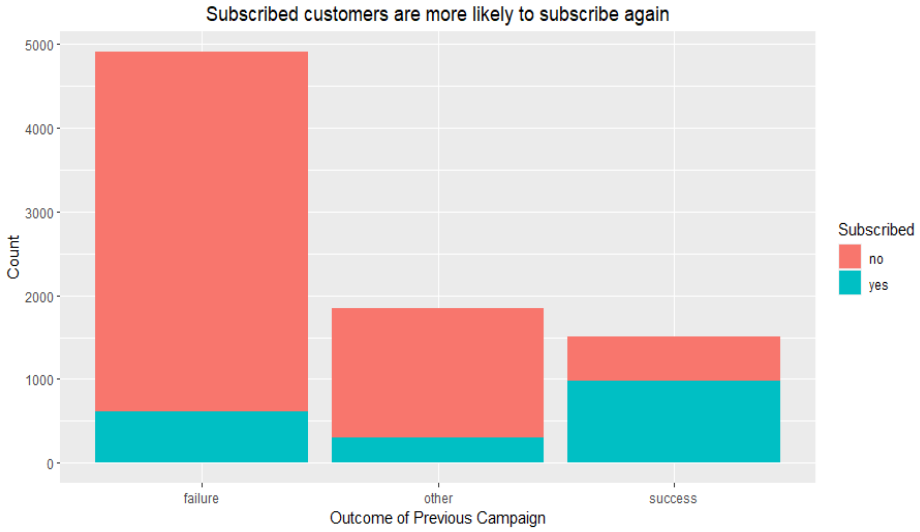


- There is clearly scope for improvement
- Majority of contacts are wasted

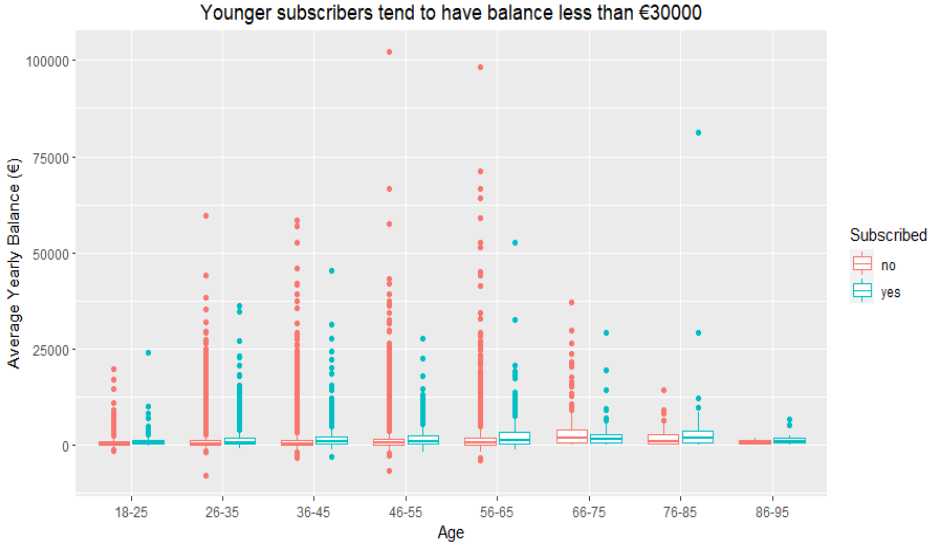
We can learn and improve from trends observed in previous campaigns



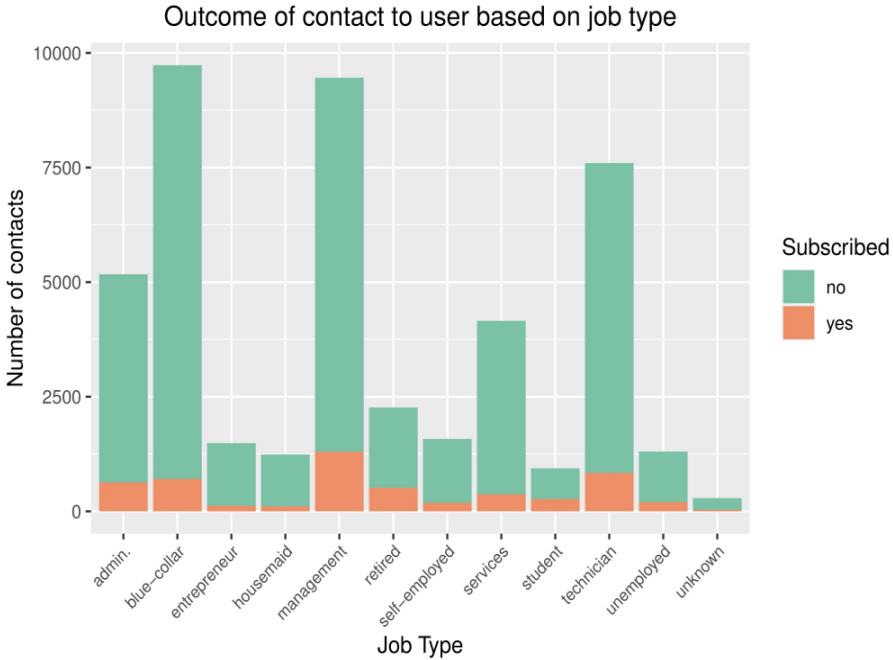
Outcome of previous campaign is important in predicting future outcome



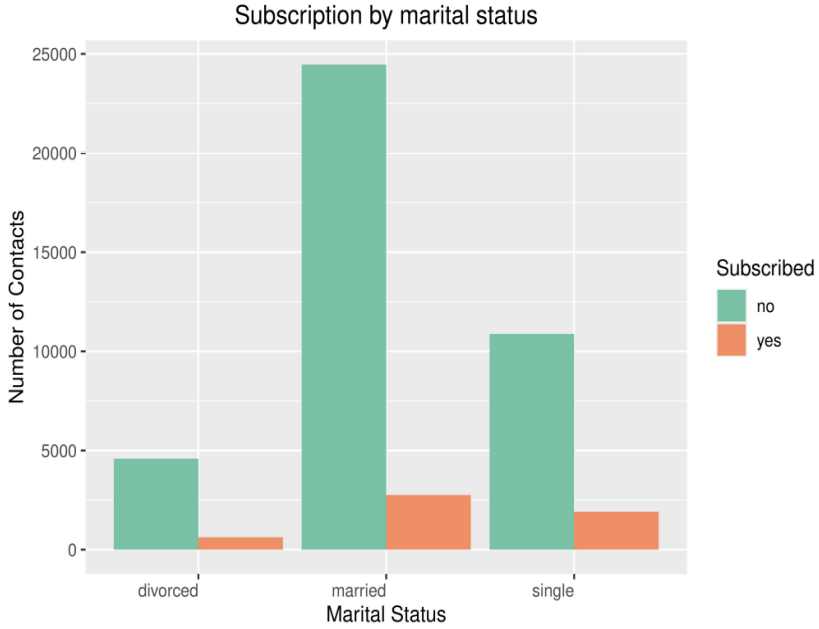
Trends in Yearly Balance change with age of client



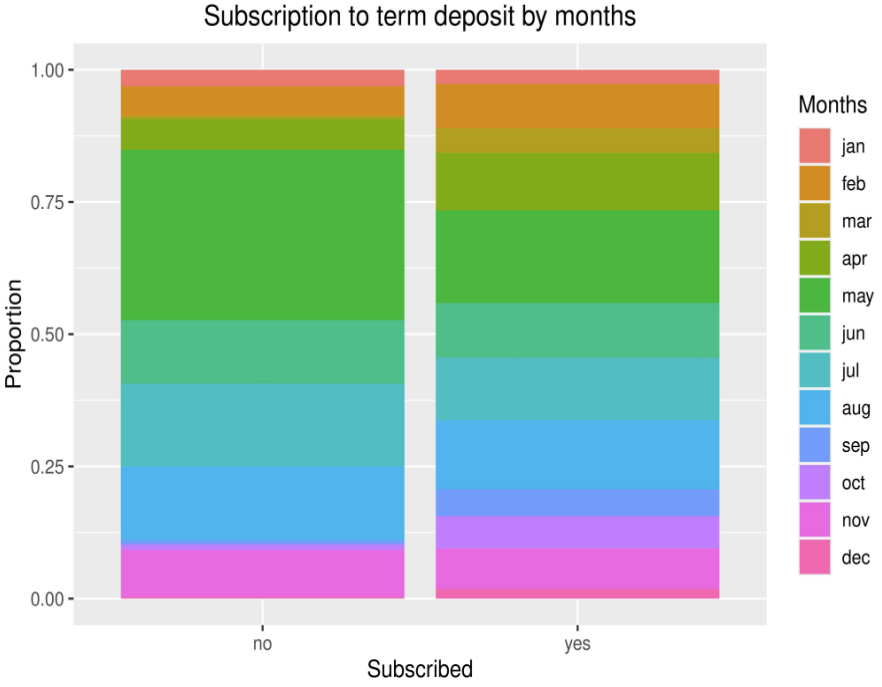
Management, student has the highest subscription rate to term deposit



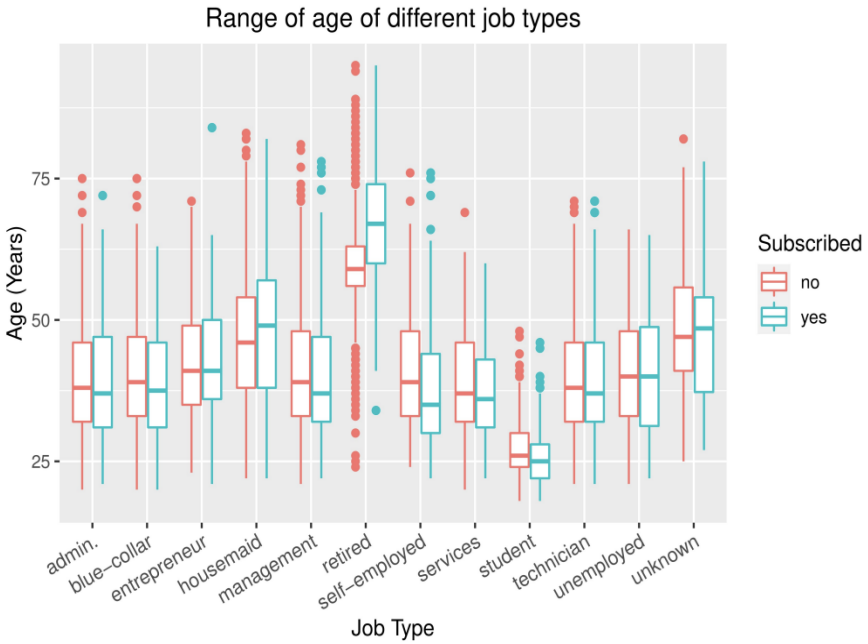
Married and Divorced are less likely to subscribe to term Deposit



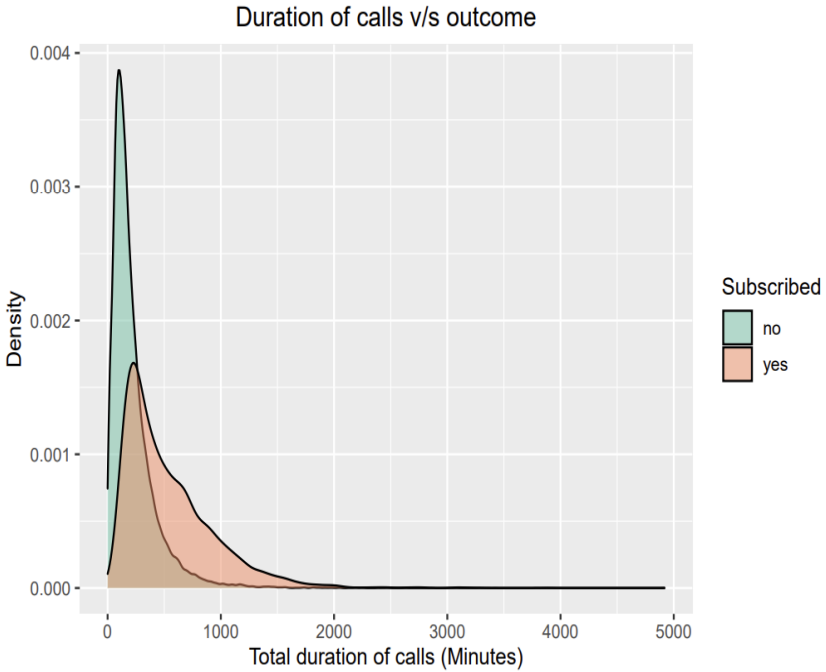
September,October,March,April had more subscription to term deposit than rejection



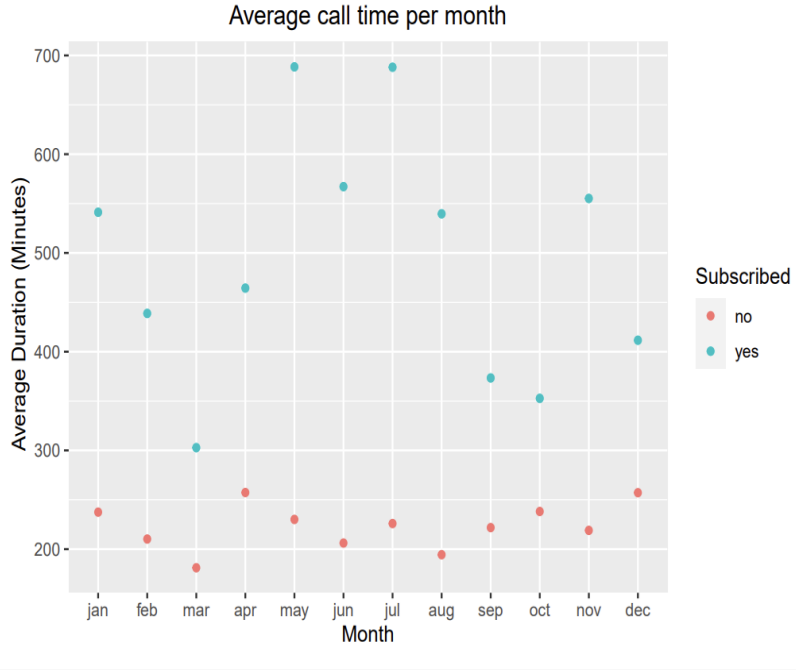
On Average Retired people tend to say ‘yes’ more to subscription then reject it.



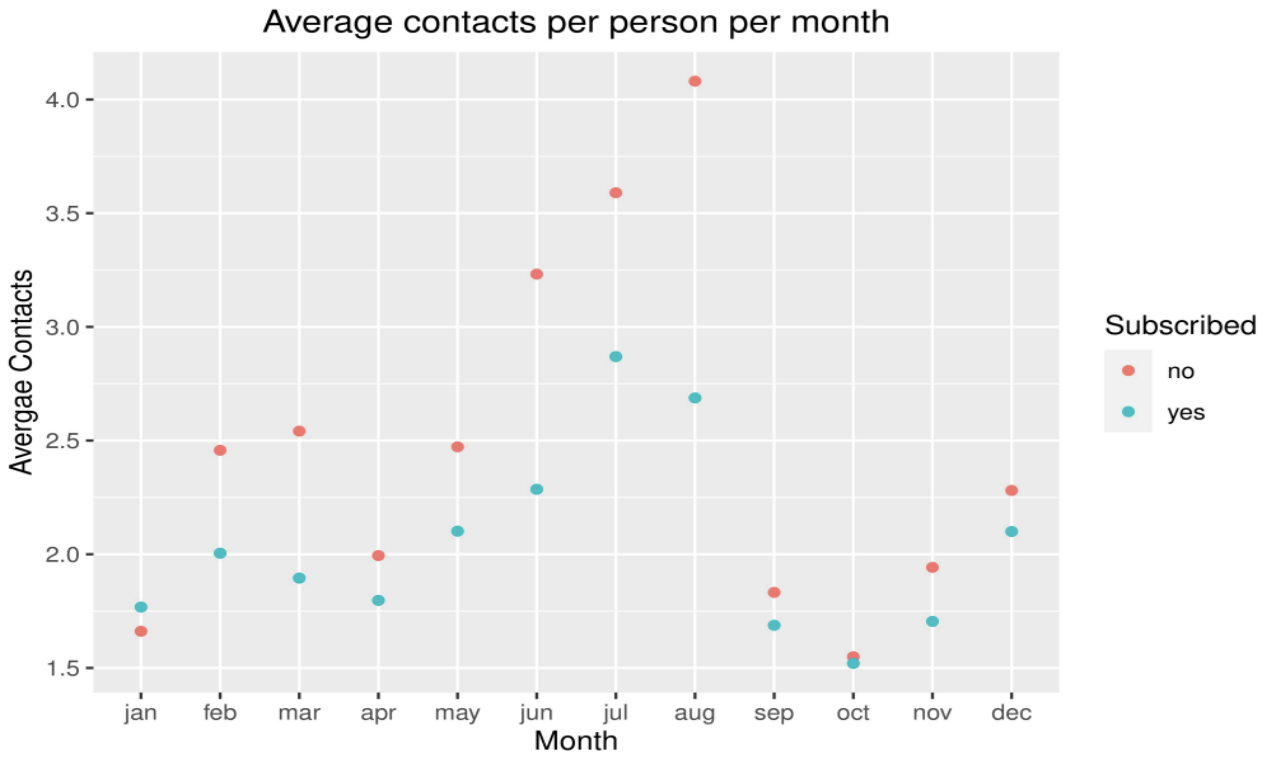
People who stay on call for longer time tend to subscribe more



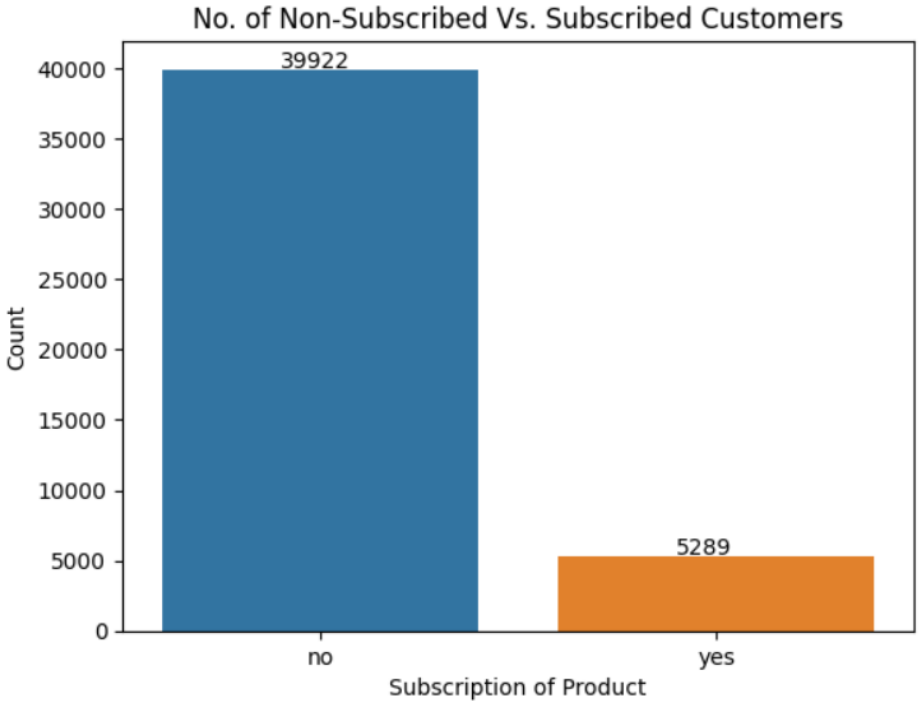
Average call time for successful subscription lasted more than 400 minutes



People who did subscribe to Term Deposit Product were contacted less than 2.5 times per month on average

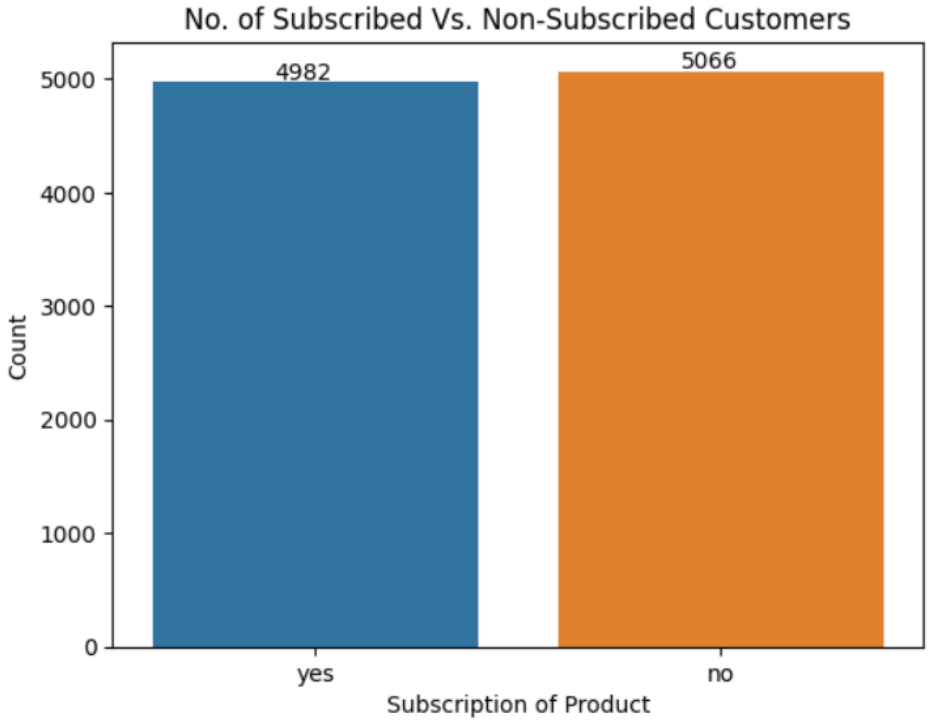


Before Data Sampling...



Class Imbalance could lead to biased predictions and decline model performance

After Data Sampling...



Downsampling the ‘no’ category leads to equal data points resulting accurate model

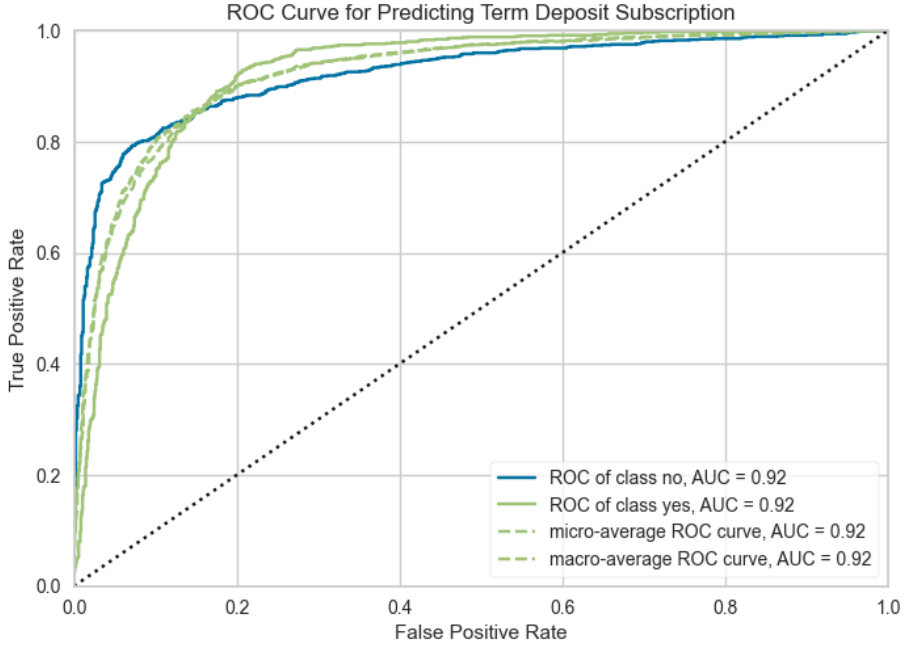
Data Pre-Processing

- Removed duplicate data points
- Downsampled data from 88-12% split to 50-50% split
- Collapse response variable into binary classes
- Dropped ‘default’ and ‘contact’ columns
- One Hot Encoded the Categorical Variables
- Scaling the Quantitative Variables using Min-Max Scalar
- Utilizing Cross-Validation on XGBoost and Random Forest models

Model Evaluation

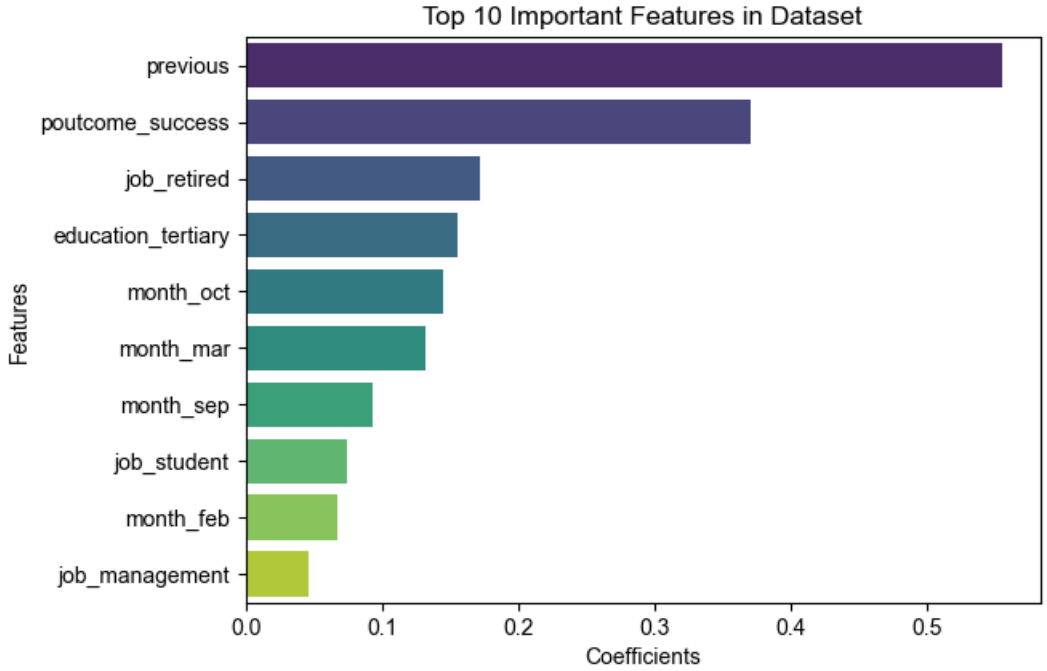
Model	Logistic Regression	Random Forest (CV)	Gradient Boosted Trees (Champion Model)
Accuracy	0.8030	0.8586	0.8670
Sensitivity	0.7788	0.8557	0.8703
Specificity	0.8266	0.8615	0.8628
Balanced	0.8027	0.8586	0.8666

Results



Area under the curve (AUC) close to 1 indicate strong classifiers

Results



The most important variables are: previous, poutcome, job, education & month

Conclusion

Will subscribe term deposit **YES**

AGE <65
PREVIOUS: less than fifteen
POUTCOME: success
JOB: retired, management, student, unemployed
EDUCATION: tertiary, unknown
MONTH: oct, sep, feb, mar, apr
MARITAL: single
HOUSING: no (has housing loan)

Will subscribe term deposit **NO**

AGE >65
PREVIOUS: zero
POUTCOME: failure, unknown
JOB: services, blue-collar, entrepreneur, technician
EDUCATION: basic.9y, basic.6y, high.school, secondary
MONTH: may, jun, jul, aug, nov
MARITAL: married
HOUSING: yes (has housing loan)

Who wants that data?
Marketing Companies / Banking Institutions

References

- Data Source : https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#

- [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

Thank You!

Questions?