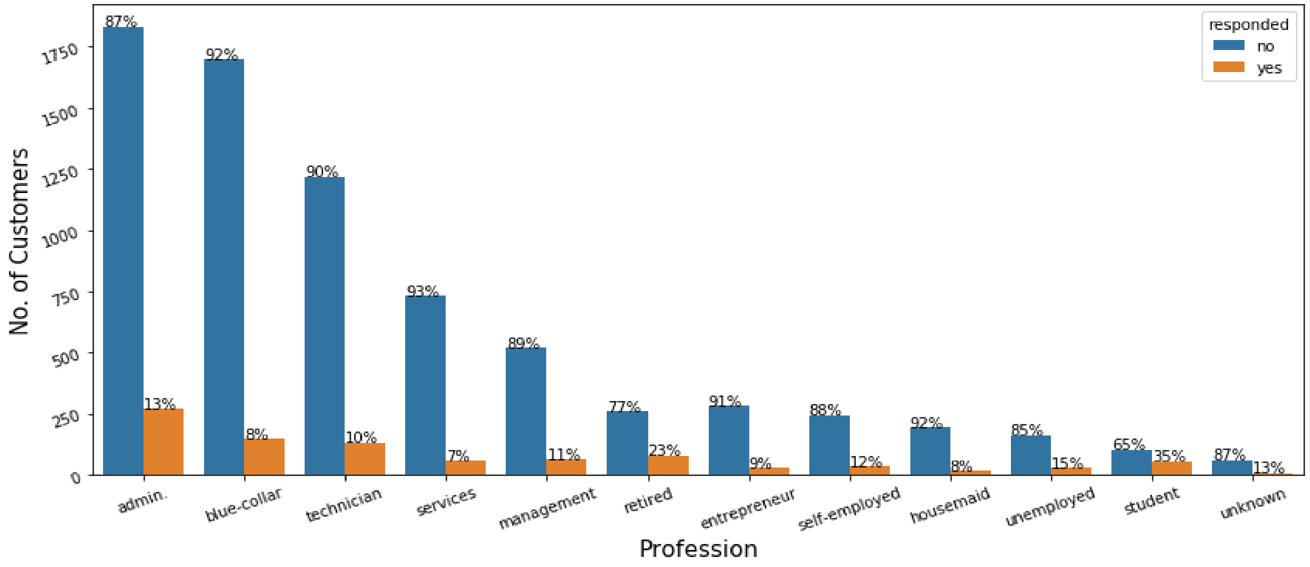
Customer Quality Prediction

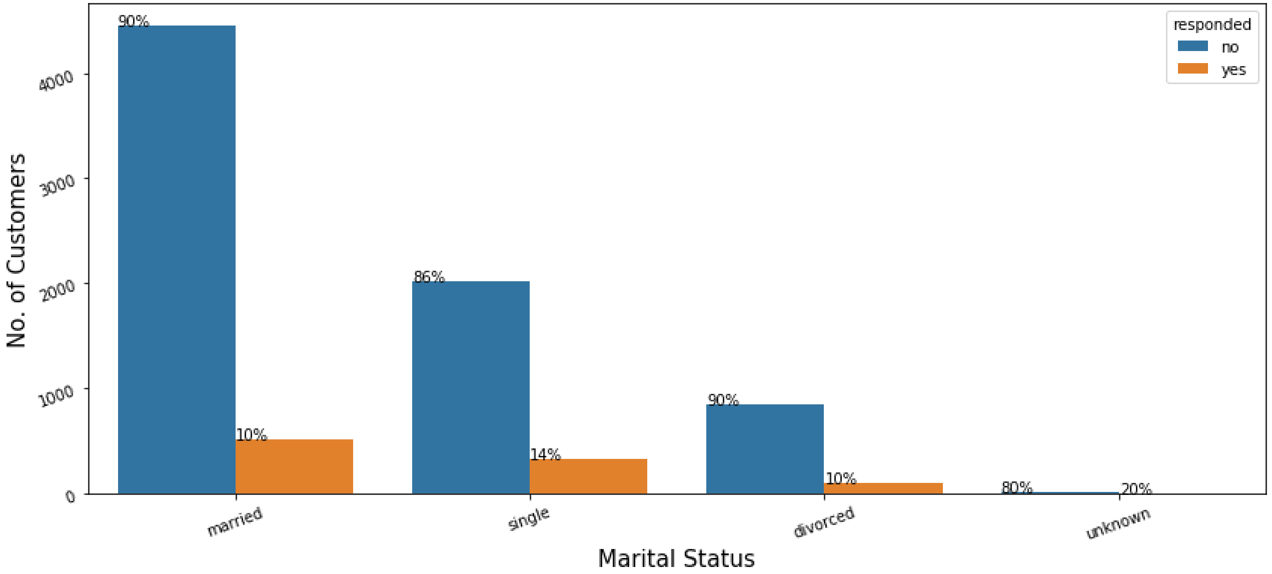
Client	Insurance Company running Marketing campaigns to sell Insurance policies to customers
Objective	Maximize the profits by running a Targeted Marketing Campaigning
Project Objective	Analyze the available data to develop insights into the customer quality Build a Machine Learning model to predict whether or not a customer should be targeted in order to maximize the profits
Historical Dataset: Features used for training the predictive model	• Customers profile: Age, Marital status, Professional & Educational background, Credit history, Preferred contact type • Economy indicators: Employment Variation Rate, CPI (Inflation indicator), Consumer Confidence Index, Euribor rate (Bank Interest rates), etc. • Previous Campaign details: Days since previous campaign, Outcome, etc.
Predictive Model Output	• Whether or not a customer responded to the present campaign ('Yes' / 'No') • Profit made by company if customer responded and purchased a policy
Approach	Data Analysis: Studying the trends to check if any strong correlations exist between the features highlighted above Build a model to predict whether or not a customer should be contacted for pitching the policy

Insights from the Data Analysis



Profession:

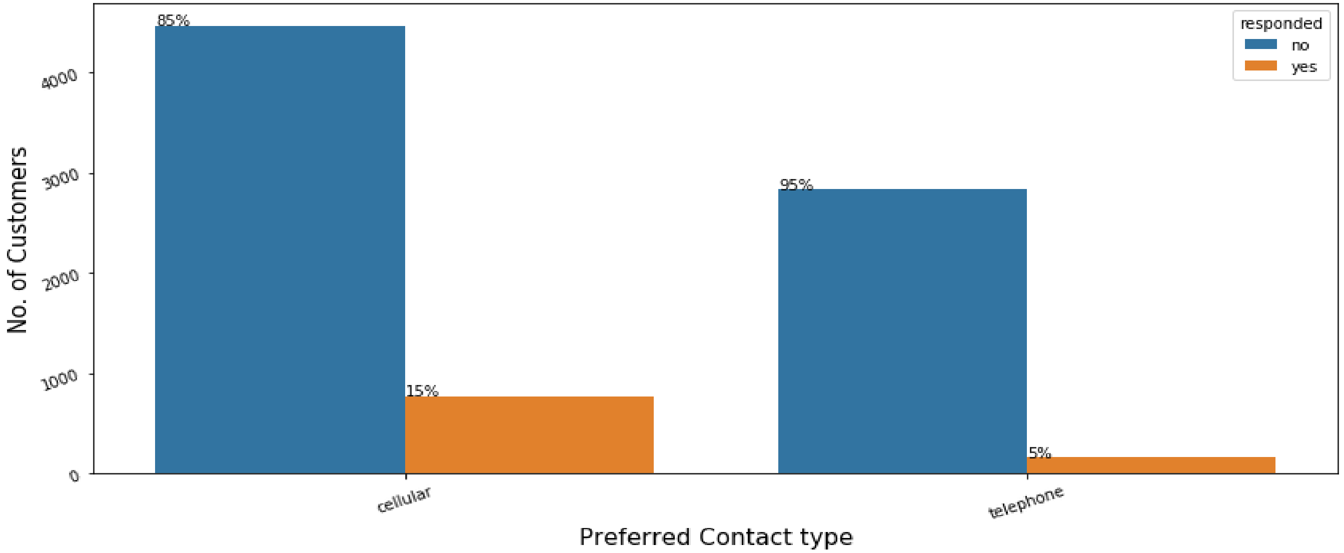
- Customers in Admin. jobs most marketed, but positive response rate not highest
- Retired customers – Highest positive response rate



Marital Status:

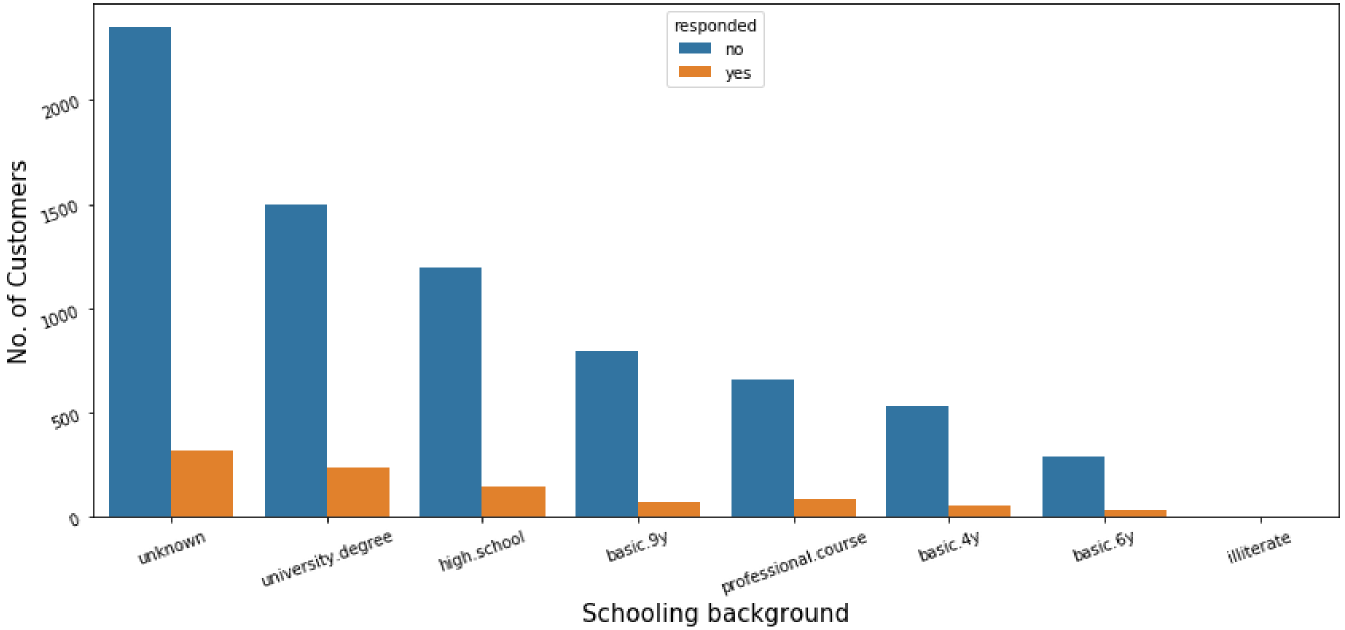
- Married customers are most marketed
- Positive response rates are close
- Single customers have highest response rate

Data Analysis…



Preferred communication mode:

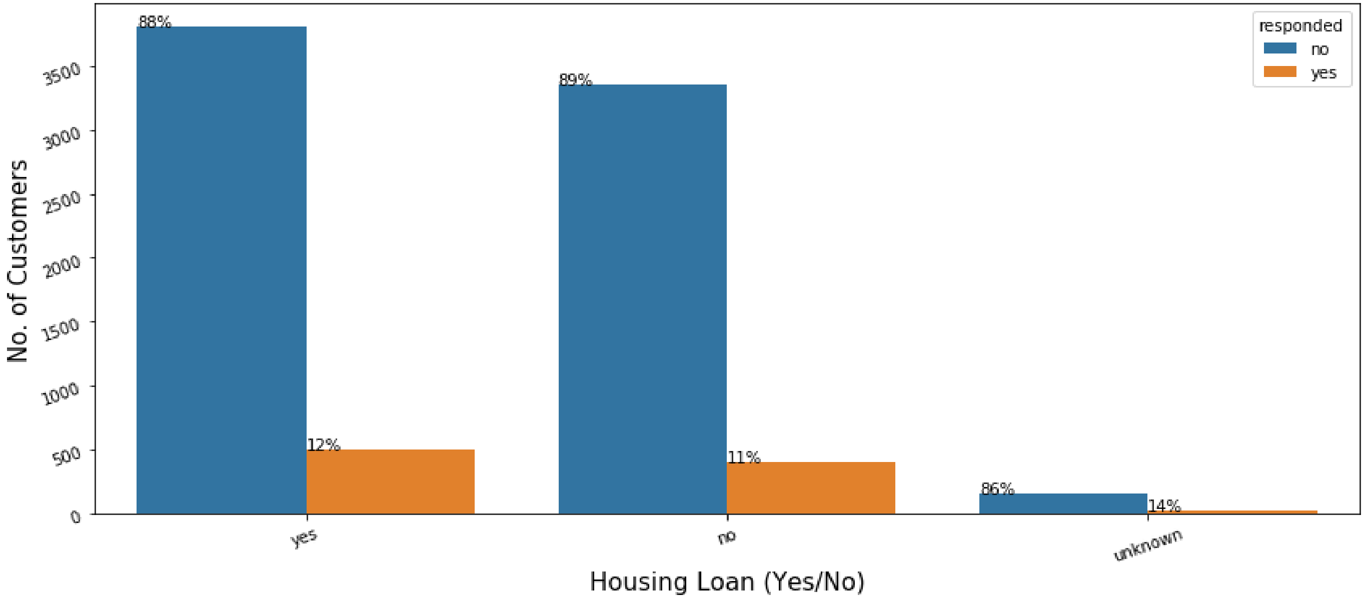
- Customers preferring communication over mobiles more marketed
- They are also more likely to respond



Schooling background:

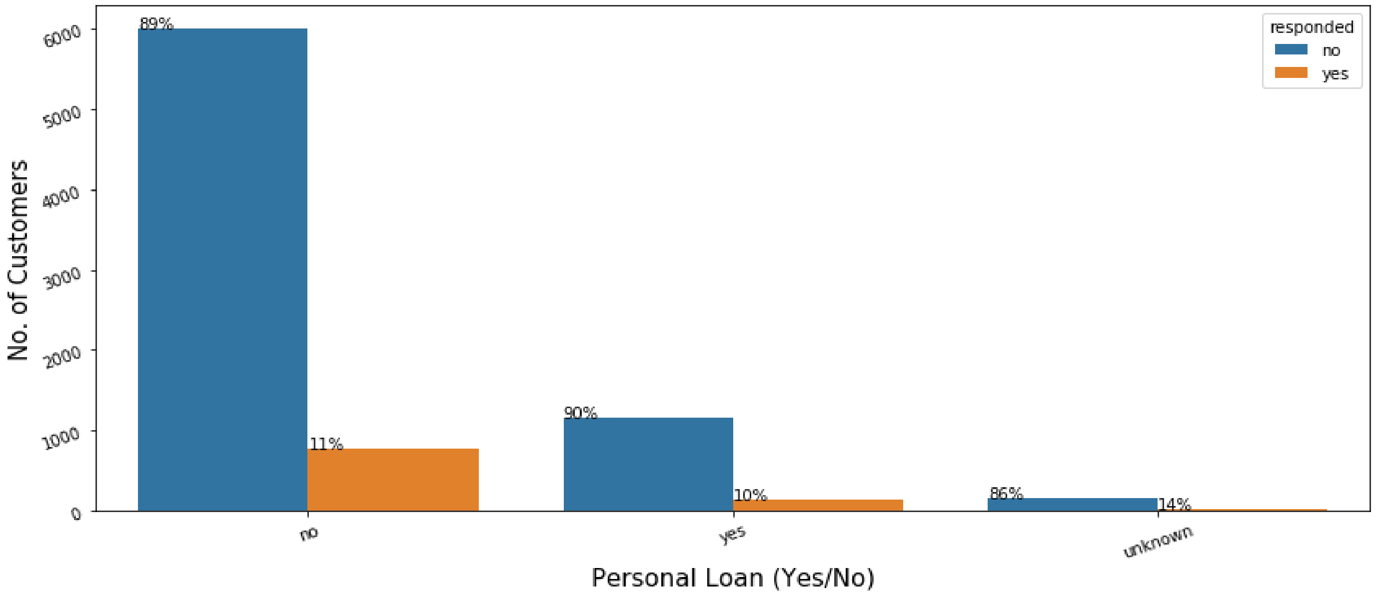
- Schooling background of a large number is unknown
- Customers with University degree and High school education are most marketed

Data Analysis…



Housing Loan:

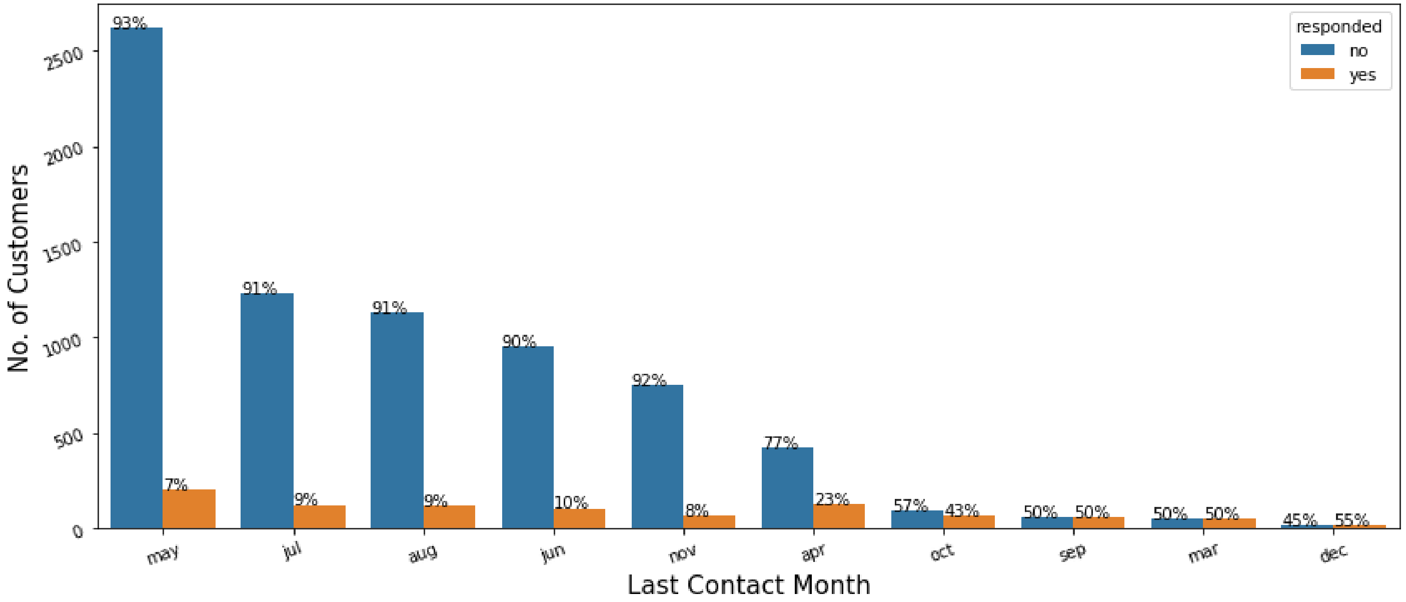
- Customers with or without Housing Loan – almost equally marketed
- Same response rates



Personal Loan:

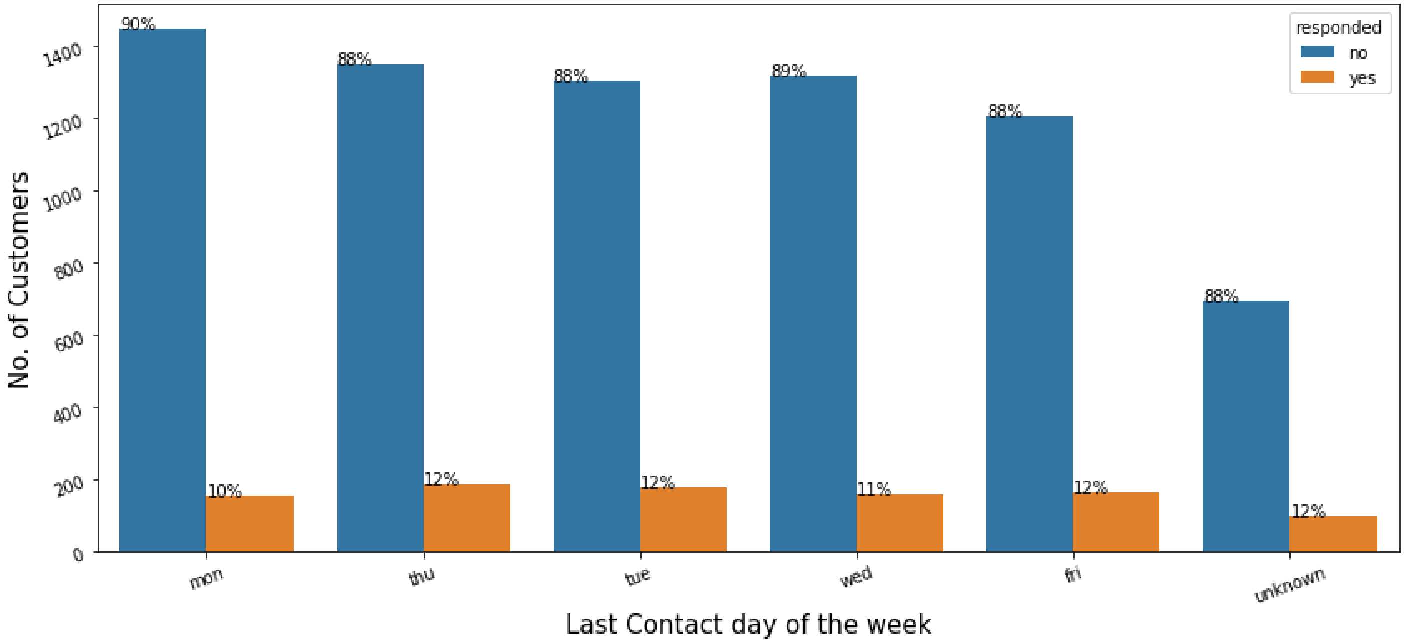
- Customers without Personal Loan are most marketed – Higher spending capacity
- However, Positive response rates are almost equal

Data Analysis…



Last Contact Month:

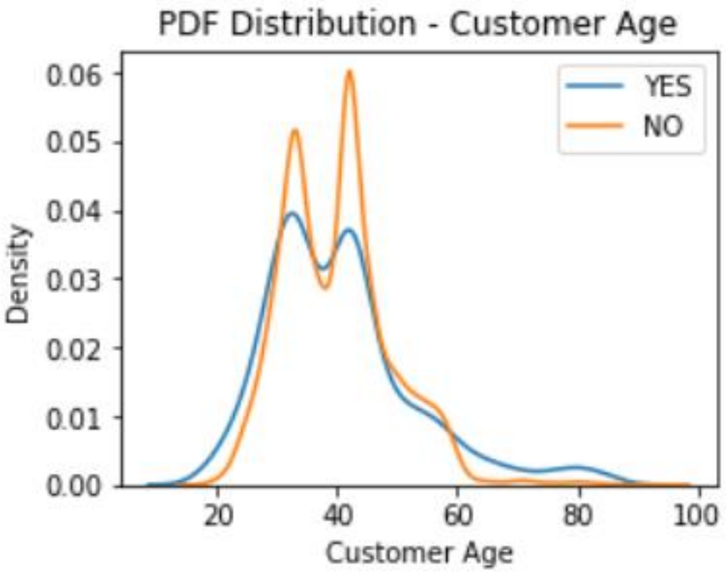
- Marketing Campaign is aggressive in the months of May-Aug
- Highest campaigning activity in May
- However, these months do not have high Positive response rates
- Months with lower campaigning activity have higher Positive response rates

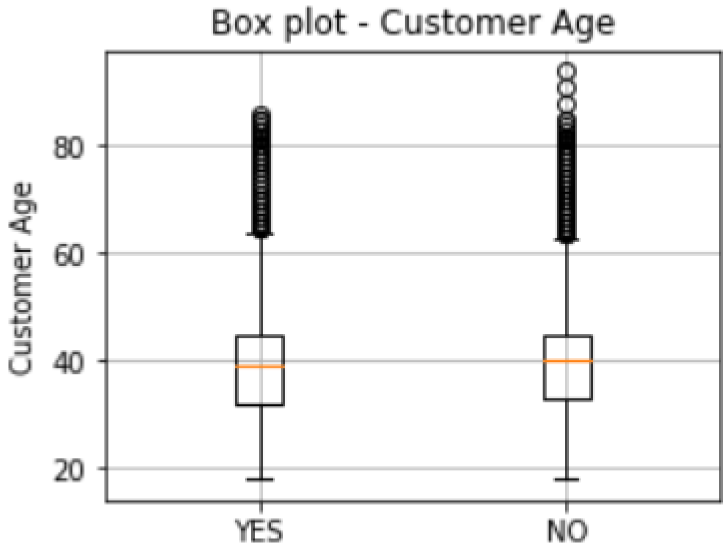


Last Contact Day of week:

- Positive response rate is also almost equal across working days of the week

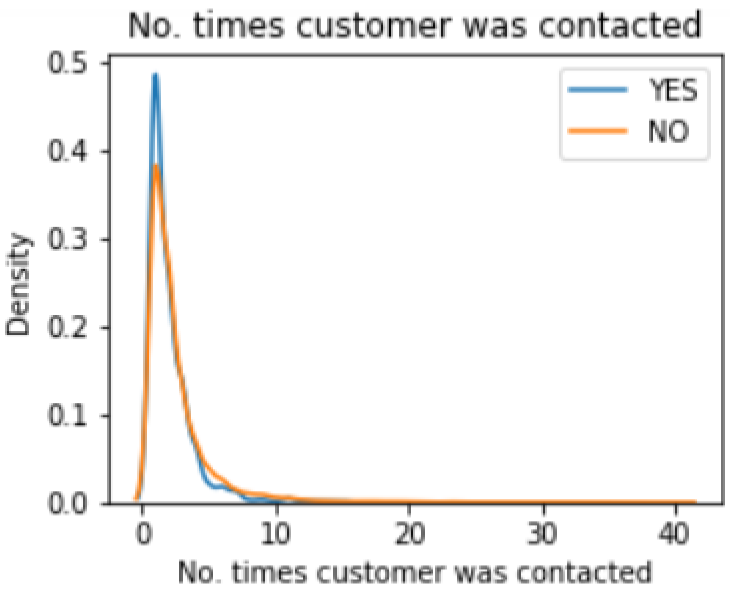
Data Analysis…





Customer Age:

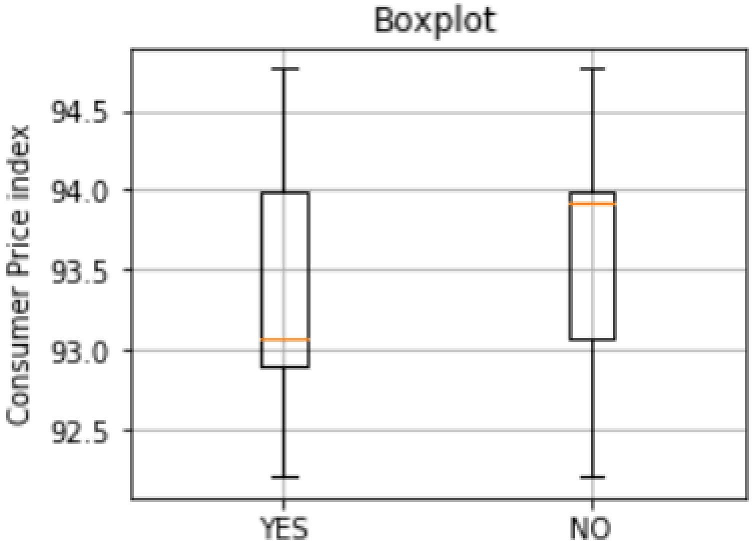
- Age distributions of customers of both classes are almost similar
- Mean = 40, Median = 38



Number of times customer was contacted in campaign:

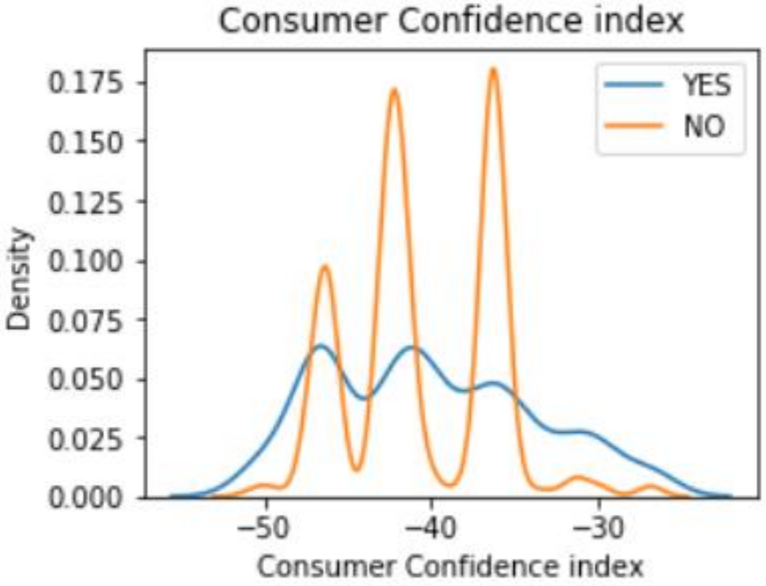
- **The number of calls does not seem to be having a large impact on the customer response**

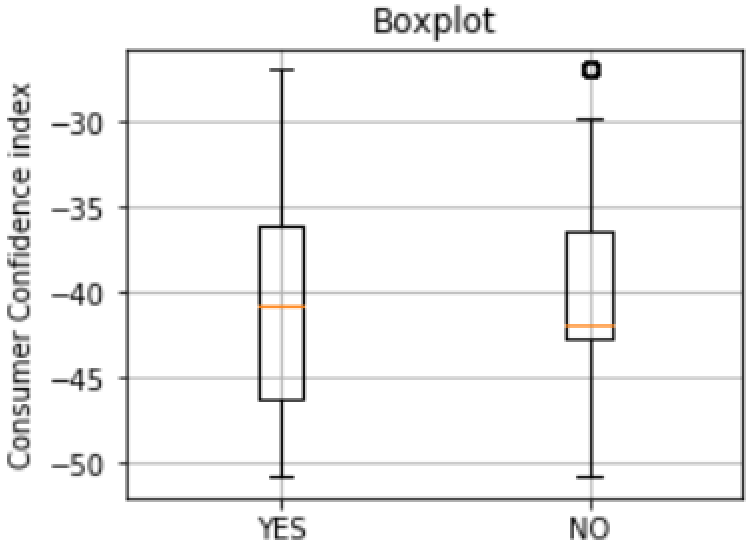
Data Analysis…



Consumer Price Index – Monthly Inflation Indicator

- Higher CPI = Higher Inflation rate and hence lower spending capacity
- **Majority of customers who have responded have done so in lower CPI periods**
- Customers are more likely to not respond during High Inflation periods
- CPI can be an important feature for our classification model

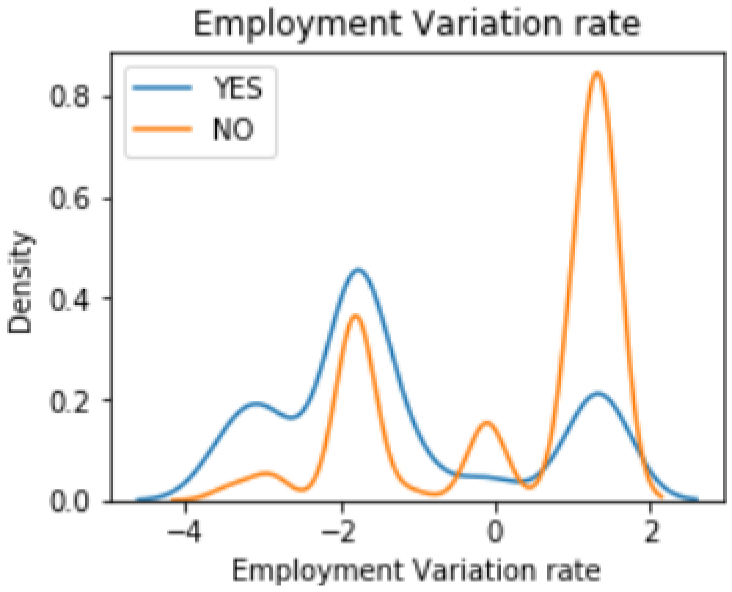


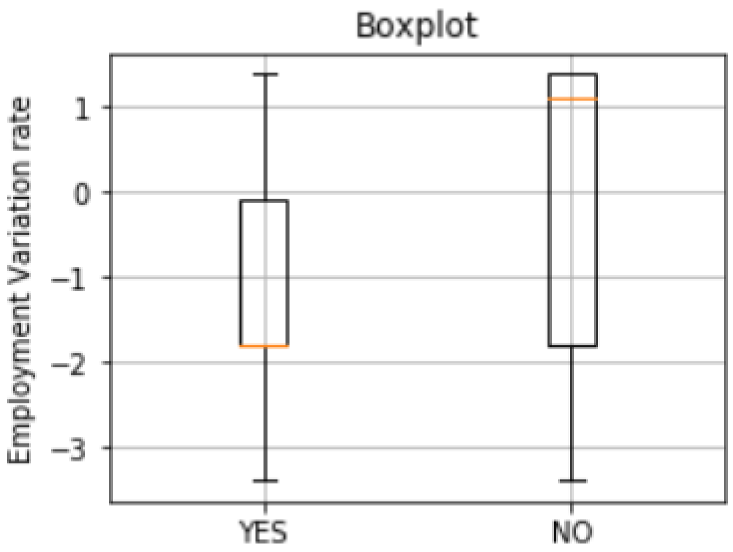


Consumer Confidence Index (CCI) – Monthly Indicator

- Indicator of optimism about financial situation
- Higher CCI = Higher spending

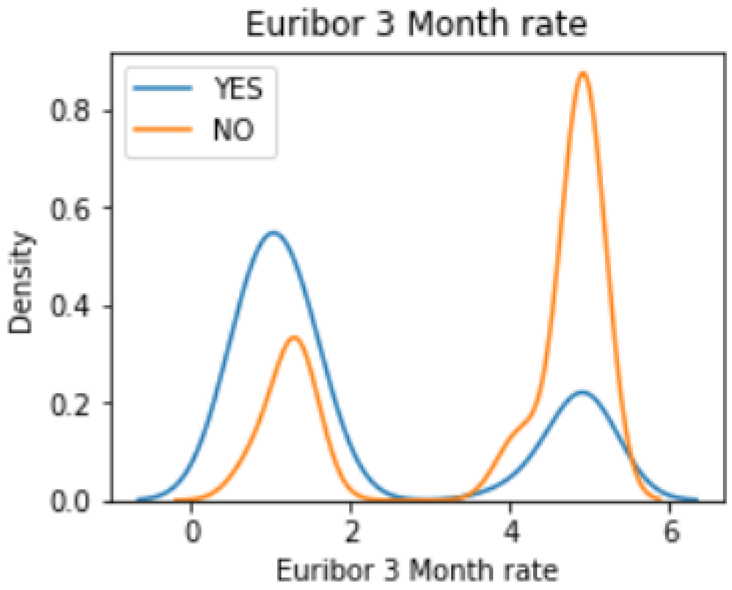
Data Analysis…

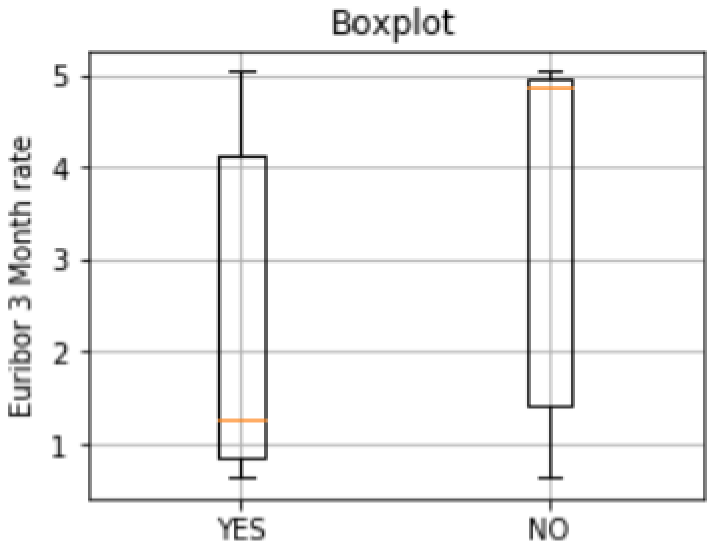




Employment Variation Rate (EVR) – Quarterly Indicator

- Majority of customers who have responded have done so in lower EVR periods
- Customers are more likely to not respond during High EVR periods
- EVR can be an important feature for our predictive model





Euribor 3 Month Rate

- Inter-bank Rate: Rate at which European banks lend to one another
- Majority of customers who have responded have done so in lower Euribor rate periods
- Customers are more likely to not respond during High Euribor rate periods
- Euribor can be an important feature for our predictive model

Metrics used for Maximizing Total Profit

$$Total\ Profit = P * n\ - C * N$$

P: Avg. profit per responding Customer

n: Number of customers responding

C: Marketing cost per customer ($30)

N: Number of customers marketed

Terminology used for assessing the predictive model's performance:

Number of customers our model correctly predicts as ***'Will Respond'*** = True Positives (TP)

*Number of customers our model incorrectly predicts as **'Will Not Respond'** = False Negatives (FN)*

*Number of customers our model incorrectly predicts as **'Will Respond'** = False Positives(FP)*

To maximize Total profit:

High $\frac{TP}{TP+FP}$, i.e. **Precision** (Confidence of model when it predicts as 'Will Respond') of our model should be high

High $\frac{TP}{TP+FN}$, i.e. **Recall** (Ability to predict all potential customers who 'Will Respond') should be high

Relative importance of Precision & Recall depends on Avg. Profit (P) and Marketing Cost (C)

Objective of the predictive model: Strike a balance

- Performance of this model is assessed based on the Total Profit generated from the Customers who were predicted as '***Will Respond'***

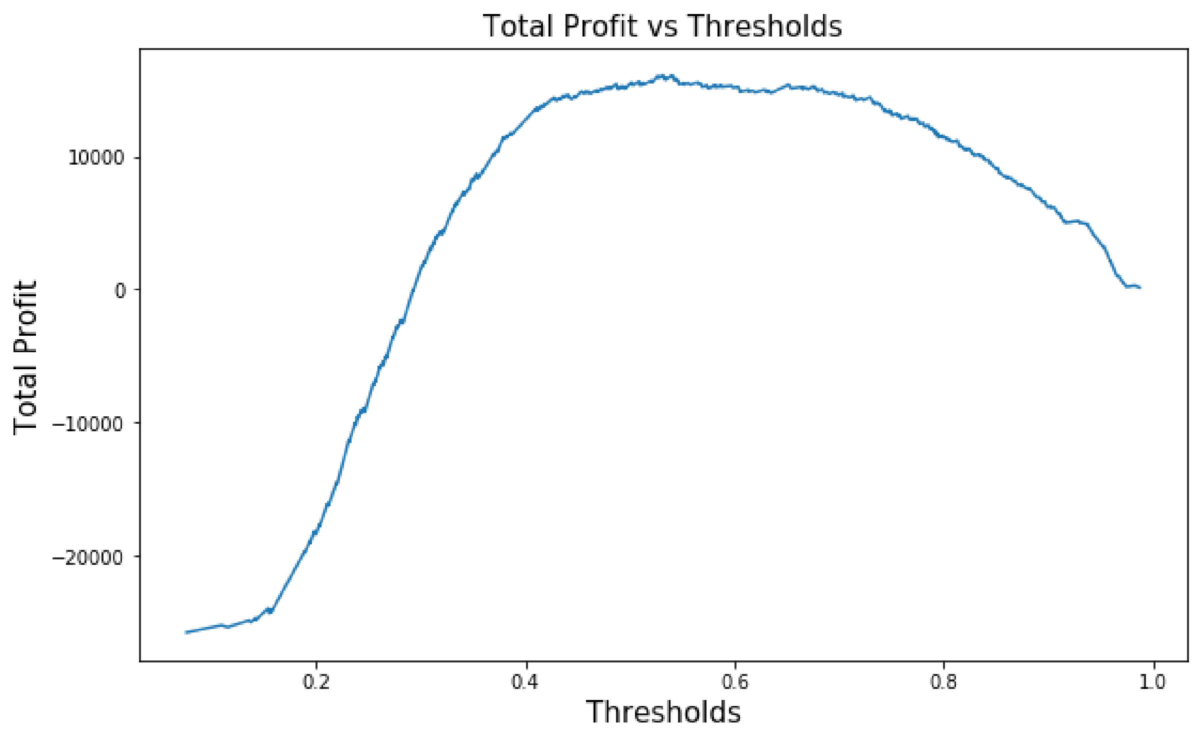
Predictive Modelling Strategy

- Training: Model learns the function f based on the relationship between Input features (X) and the desired Output Variables (Y) from the given historical dataset

$$Y = f(X_1, X_2, \ldots.)$$

- Model is trained on a subset of the historical dataset and its performance is assessed on a held-out Test set
- The model parameters are fine-tuned to maximize the performance using K-fold cross validation
- Different types of Machine Learning (ML) models are trained to pick the best model based on performance on the Test set
- Output of chosen ML Model: Probabilities of a customer belonging to Positive ('Will Respond') class
- A heuristic was used to choose the Threshold value beyond which the Model would classify a customer as 'Will Respond'

Predictive Model Performance



Metric	Value
Threshold	0.65
Precision	0.42
Recall	0.62
Total Profit	$19,684