# Bank Direct Marketing

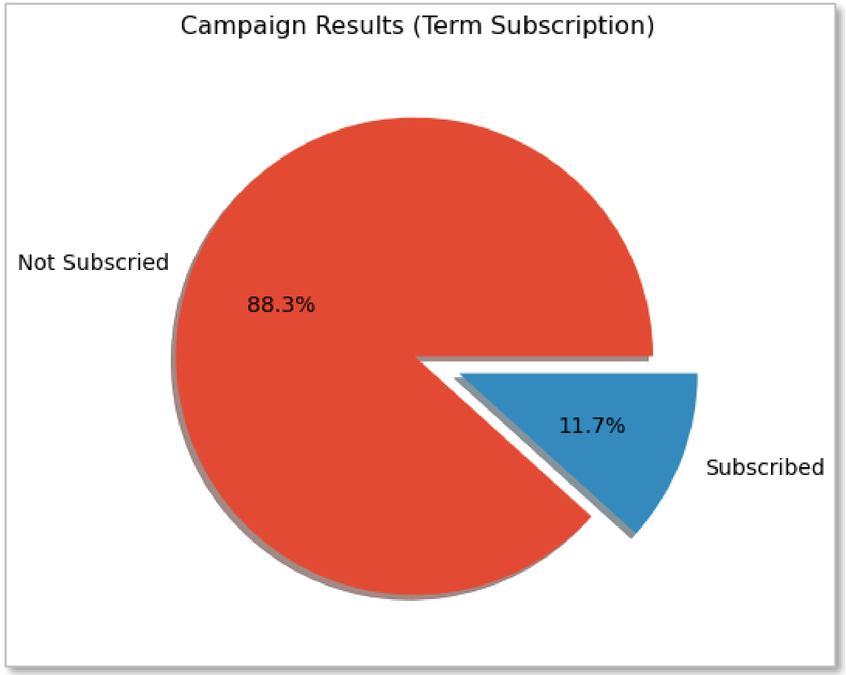
- PERFORMANCE EVALUATION
- ANALYTICS RECOMMENDATION

# Agenda

- Executive Summary
- Campaign Performance and Analysis
- ML Approaches
- Final Thoughts and Next Step

# Executive Summary

- The analysis is conducted using the direct marketing campaign data collected by a Portuguese retail bank from period of May 2008 to November 2010
- Data consist of 17 variables including demographic, financial status and campaign metrics
- Objective : Provide analytical solution to improve campaign targeting using ML approaches




- Overall, 11.7% of the target population responded and opened a term deposit account with the bank
- The analysis suggests there are opportunities to apply ML models to create a rule-based engine for target selection or to use a propensity model to help the callers to prioritize the call
- Propensity model approach could be expanded to a recommendation system for multiple other banking products/services

# Profile Analysis - Who Responded?

- **Responder vs. Non-Responder**

| | | Non-Responder | Responder |
|---|---|---|---|
| | **Volume** | 39,922 | 5,289 |
| | **Avg. Age** | 40.8 | 41.7 |
| **Marital Status** | **divorced** | 11.48% | 11.76% |
| | **married** | 61.27% | 52.09% |
| | **single** | 27.25% | 36.15% |
| **Education** | **primary** | 15.68% | 11.17% |
| | **secondary** | 51.98% | 46.32% |
| | **tertiary** | 28.32% | 37.74% |
| **Occupation** | **admin.** | 11.37% | 11.93% |
| | **blue-collar** | 22.6% | 13.39% |
| | **entrepreneur** | 3.42% | 2.33% |
| | **housemaid** | 2.83% | 2.06% |
| | **management** | 20.43% | 24.6% |
| | **retired** | 4.38% | 9.76% |
| | **self-employed** | 3.49% | 3.54% |
| | **services** | 9.48% | 6.98% |
| | **student** | 1.68% | 5.09% |
| | **technician** | 16.93% | 15.88% |
| | **unemployed** | 2.76% | 3.82% |
| **Loans and Balances** | **housing** | 58.1% | 36.59% |
| | **default** | 1.91% | 0.98% |
| | **loan** | 16.93% | 9.15% |
| | **Median Balance ($)** | $417.0 | $733.0 |

- Responders are **more likely** to :
  - Be Single – more discretionary spending
  - Have Higher Education – higher salary and/or financially savvy
  - Maintain Higher Balance
- Responders are **less likely** to have:
  - Blue collar/service job
  - Consumer Loans/Mortgages
  - Default history

# Responder Profile – Customer vs. Prospect

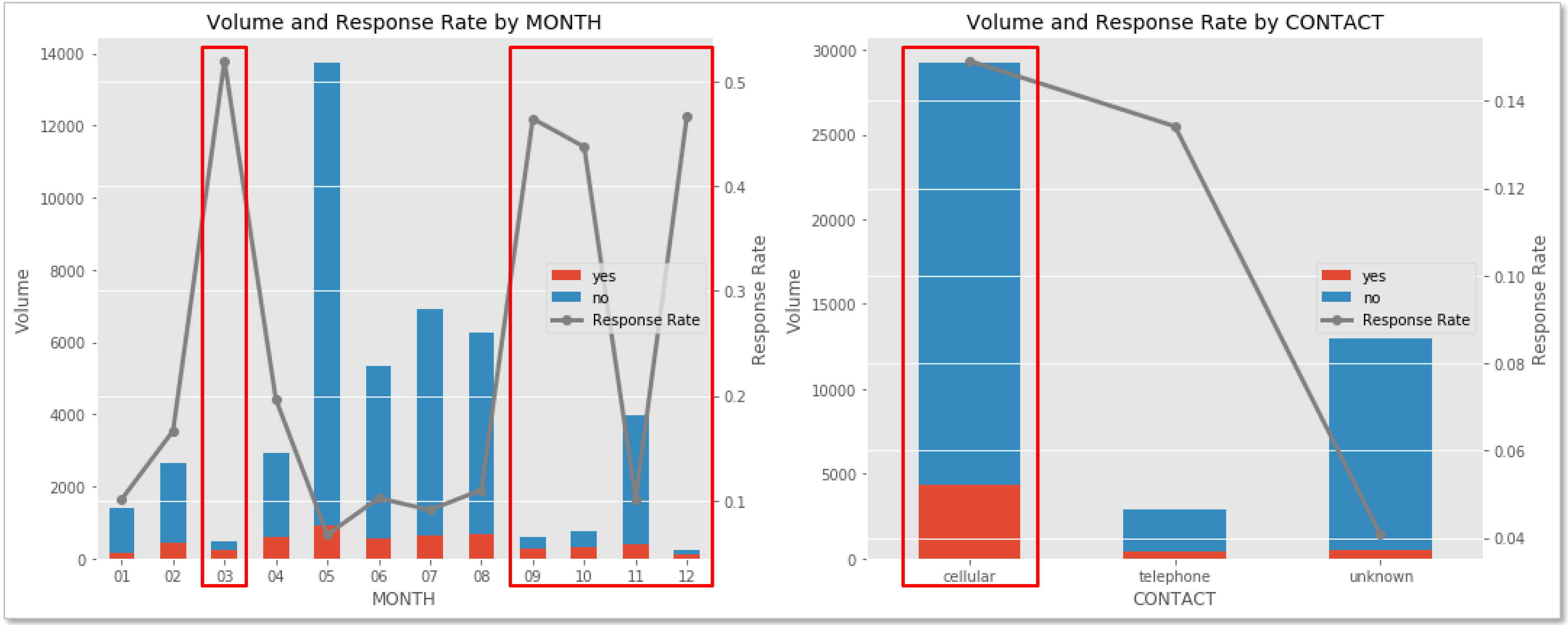
- **Customer vs. Prospect**

| | | Customer | Prospect |
|---|---|---|---|
| | Volume | 1,905 | 3,384 |
| | Avg. Age | 42.5 | 41.2 |
| Marital Status | divorced | 9.66% | 12.94% |
| | married | 55.12% | 50.38% |
| | single | 35.22% | 36.67% |
| Education | primary | 9.08% | 12.35% |
| | secondary | 44.88% | 47.13% |
| | tertiary | 41.0% | 35.9% |
| Occupation | admin. | 13.07% | 11.29% |
| | blue-collar | 9.66% | 15.48% |
| | entrepreneur | 1.42% | 2.84% |
| | housemaid | 1.78% | 2.22% |
| | management | 27.4% | 23.02% |
| | retired | 10.81% | 9.16% |
| | self-employed | 3.41% | 3.61% |
| | services | 6.25% | 7.39% |
| | student | 6.25% | 4.43% |
| | technician | 14.8% | 16.49% |
| | unemployed | 4.3% | 3.55% |
| Loans and Balances | housing | 34.17% | 37.94% |
| | default | 0.26% | 1.39% |
| | loan | 6.35% | 10.73% |
| | Median Balance ($) | $883.0 | $674.0 |

- From the responders, there is not much noticeable difference in profile between existing customer and prospects *(* Prospect : New-to-Bank/First time Contacted)*

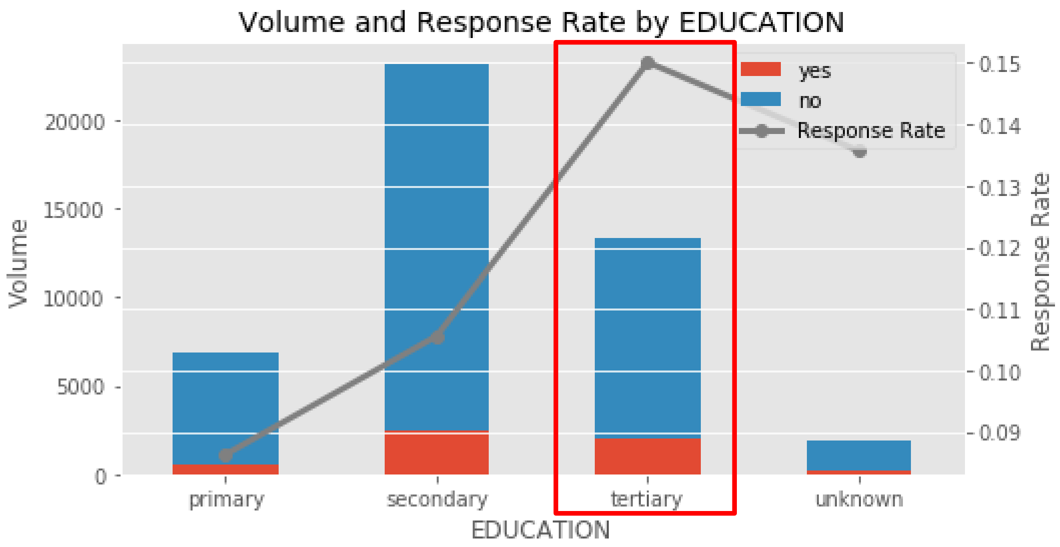
- Slightly higher median balances from existing customers

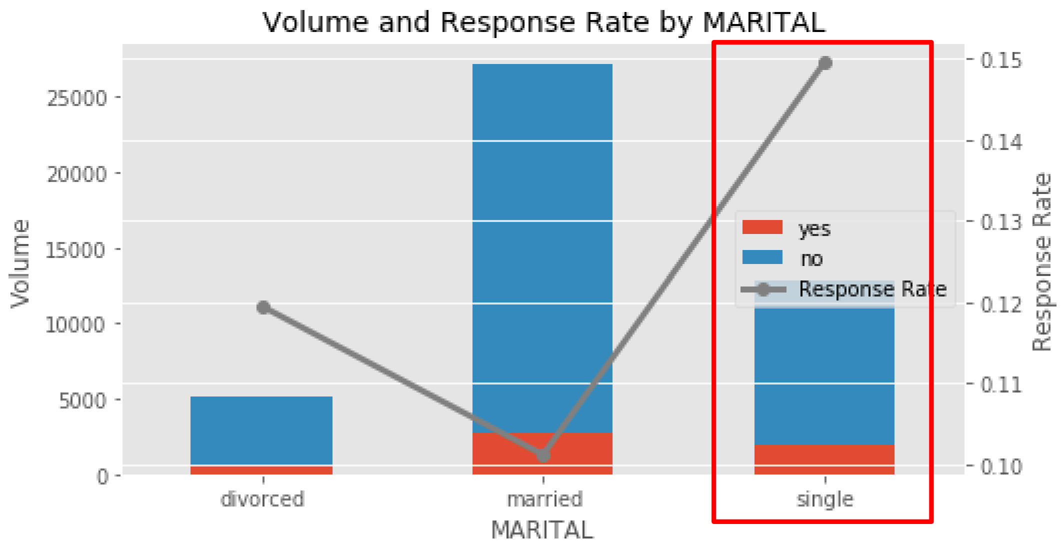
# Response rates by time of the year and preferred contact channel




- Higher response rates observed in September through December as customers tend to reevaluate their finances towards the end of the year
- March and April also shows relatively higher response rates likely due to upcoming tax season during April and May
- Across contact type, Cellular showed the highest response rates

# Response rates by Education, Martial Status and Job titles

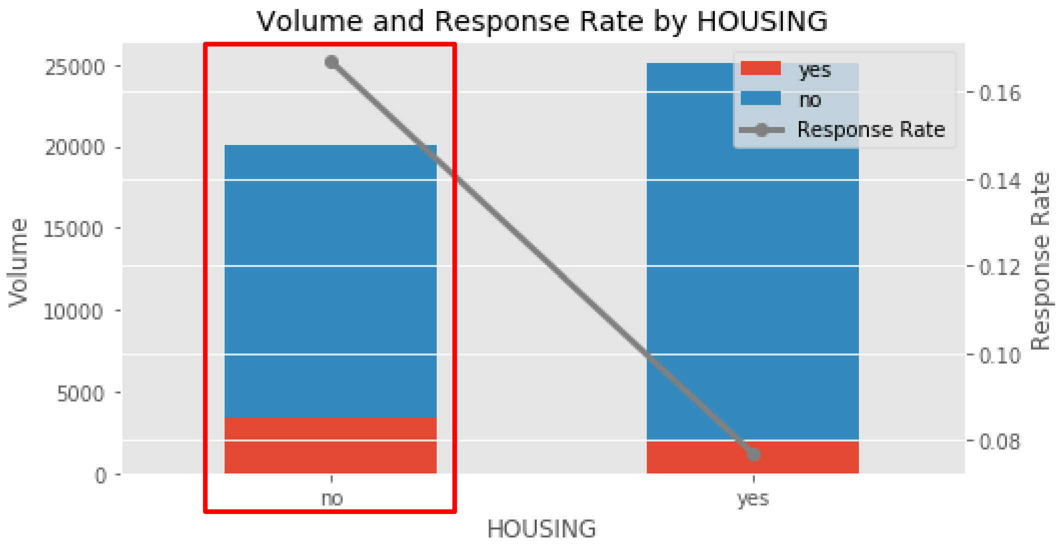


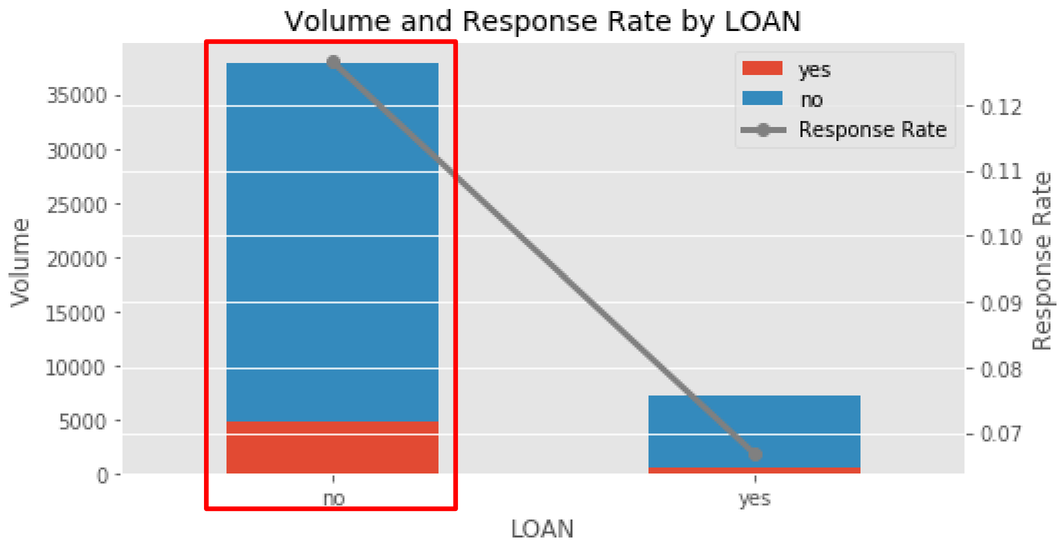



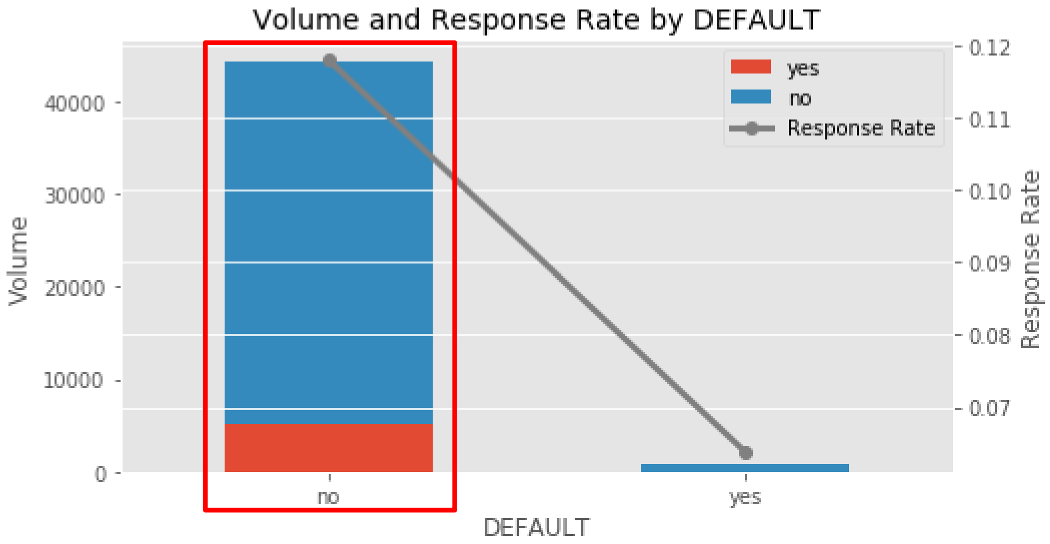


- As suggested in the profile analyses, higher conversion rates were observed from college educated group (Top Left) and single group (Top right)
- From occupation perspective (Bottom Left), students and retired group showed the highest response rates – presumably, those are the most risk adverse population

# Consumer with Less Debt are more likely to subscribe to a term deposit



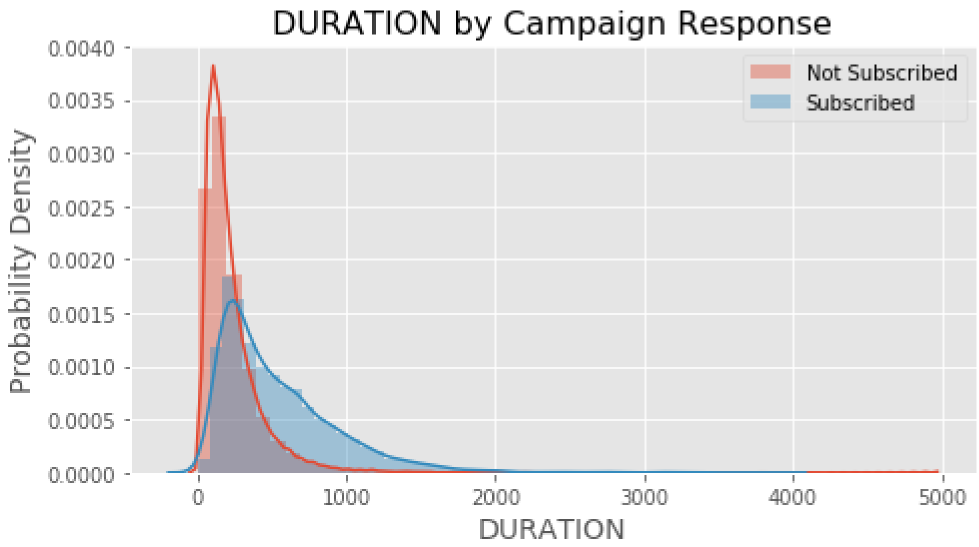



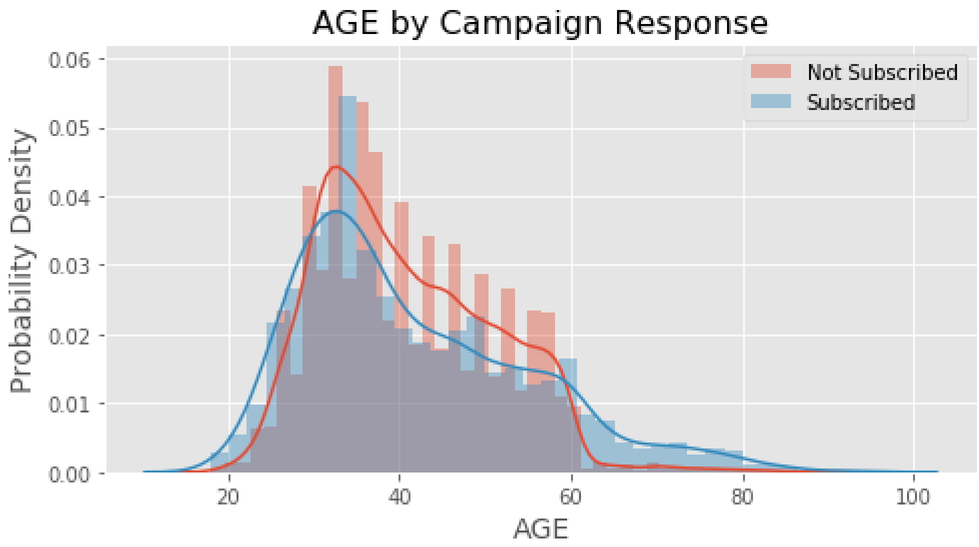



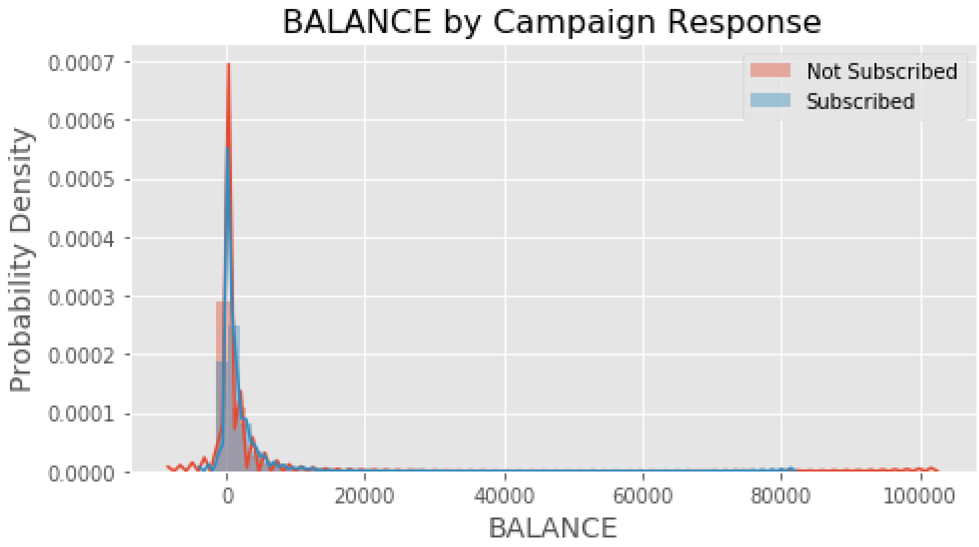

- As expected, targets with loans or default history would have lower discretionary spend and hence are less keen to subscribe to a term deposit

  → Consider suppressing those who have defaulted in the past

# No discernable pattern from customer age or balance but call duration seems to be correlated with conversion

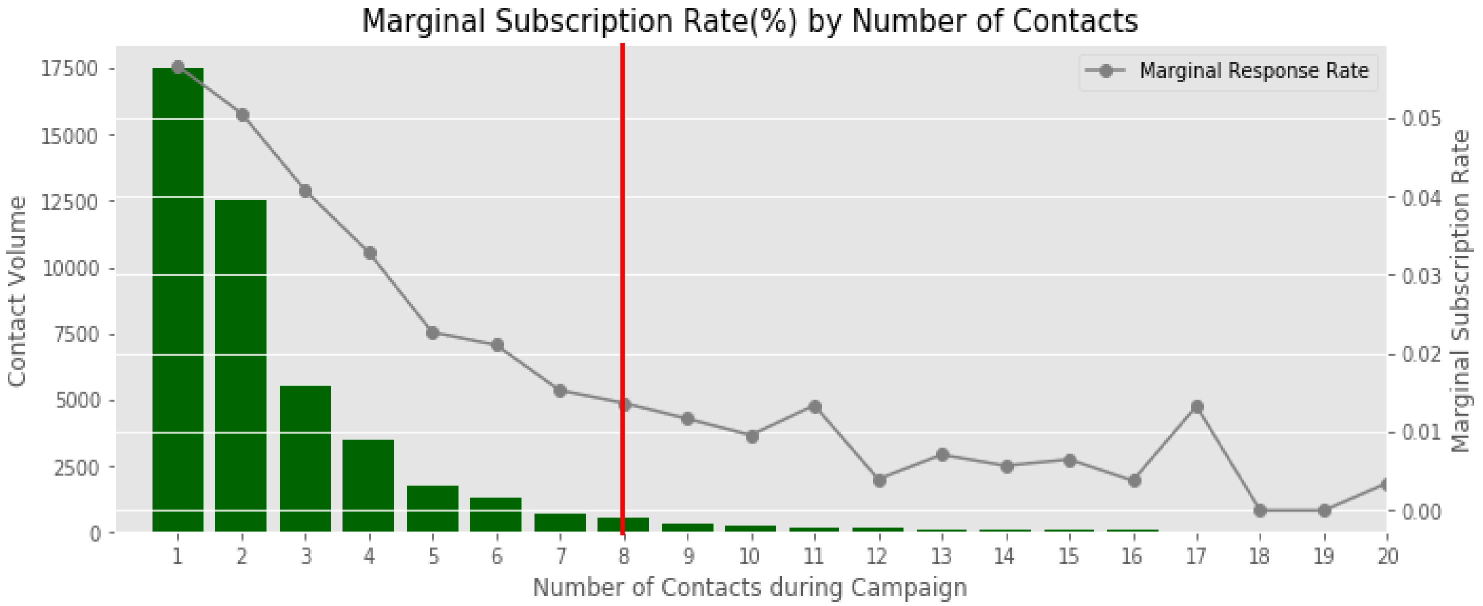



- People who subscribed to a term deposit tend to have a longer call duration
- No significant difference observed from the distributions of Age and Balance

# Marginal Subscription Rate by Number of Contacts



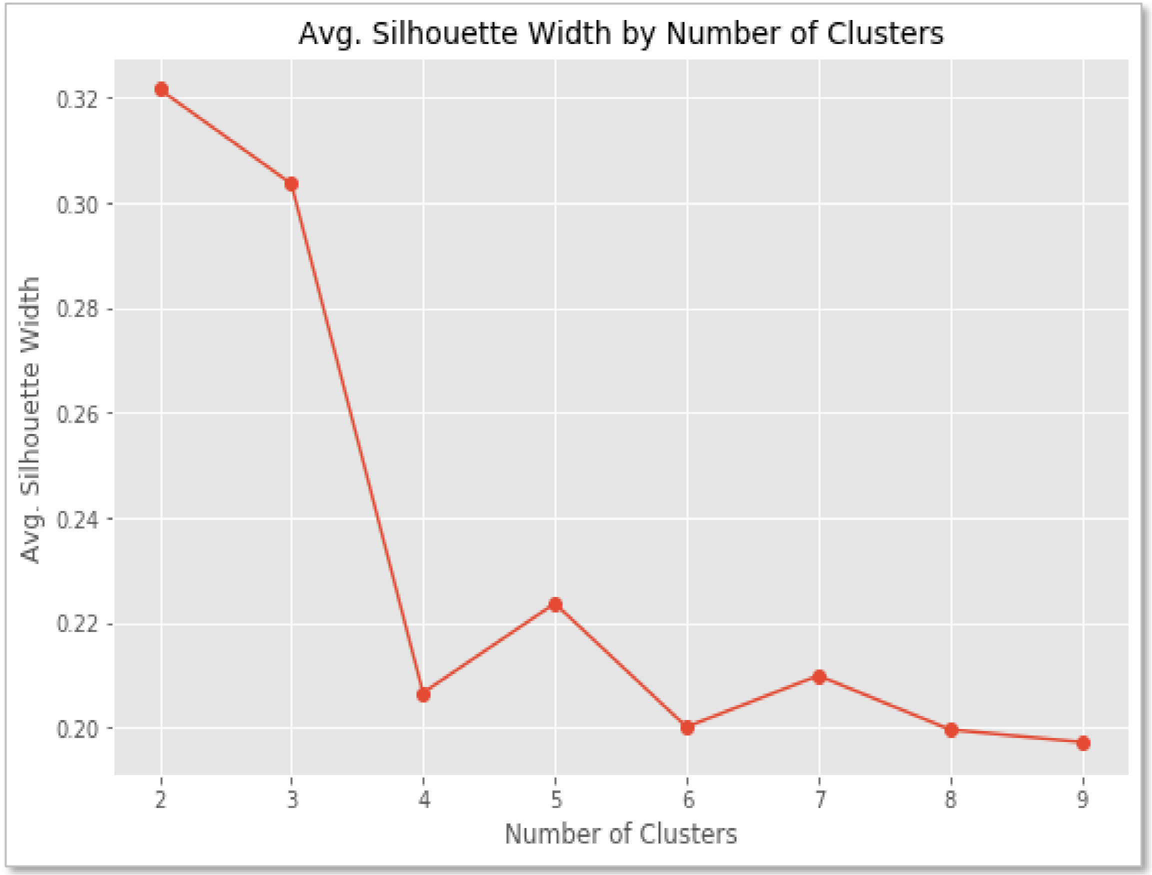

| | # Subscribed | Volume Remaining | Marginal Subscription Rate |
|---|---|---|---|
| 1 | 2561.0 | 45211.0 | 0.0566 |
| 2 | 1401.0 | 27667.0 | 0.0506 |
| 3 | 618.0 | 15162.0 | 0.0408 |
| 4 | 317.0 | 9641.0 | 0.0329 |
| 5 | 139.0 | 6119.0 | 0.0227 |
| 6 | 92.0 | 4355.0 | 0.0211 |
| 7 | 47.0 | 3064.0 | 0.0153 |
| 8 | 32.0 | 2329.0 | 0.0137 |
| 9 | 21.0 | 1789.0 | 0.0117 |
| 10 | 14.0 | 1462.0 | 0.0096 |
| 11 | 16.0 | 1196.0 | 0.0134 |
| 12 | 4.0 | 995.0 | 0.004 |

- Marginal Subscription Rate (%) by the number of contacts (Table) shows how much conversion opportunity remains after contacting X number of time for each target
- The Marginal Subscription Rate (%) flattens out as number of contacts per target increases
- With threshold of 7 contacts, about 98% of the responses are captured. Hence, to save time and effort, no more than 7 contacts should be attempted for each target

# Segmentation Approach - Model Design

| | | |
|---|---|---|
| Algorithm | **K-Medoids** | A classical partitioning technique of clustering, which clusters data into k clusters<br>In contrast to the K-Means, K-Medoids chooses data  points within data as centers |
| Evaluate Metrics | **Gower's Distance** | Dissimilarity measure that also handles both numerical and categorical variables |
| Evaluate Metrics | **Average Silhouette Score** | Measures the quality of clustering achieved ranging from -1 and 1. Values near 0 indicate overlapping clusters (poor quality of clusters) |

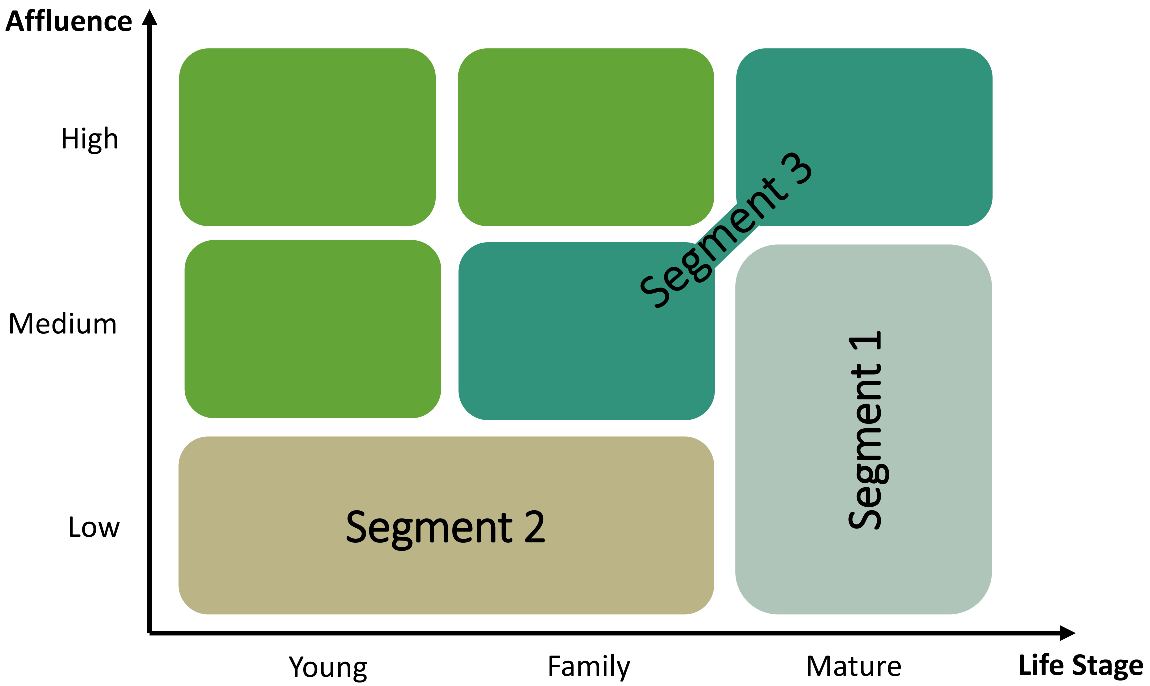
# Resulting Silhouette Score Indicates overlapping clusters




- Average Silhouette Coefficient indicates the quality of clustering achieved
  - e.g. Higher Silhouette index indicates a better quality of clustering
- Practical threshold of reasonable quality clusters is 0.5
- However, 0.30-0.32 was the best with given data set indicating overlapping clusters. Although not ideal, further analysis in the next two slides is conducted based on this result.

# Profiles by Segment

| Target Audience | | Segment 1 | Segment 2 | Segment 3 |
|---|---|---|---|---|
| | **Volume** | 19,396 | 11,336 | 12,461 |
| | **Avg. Age** | 45 | 33.5 | 40.7 |
| **Marital Status** | **divorced** | 11.63% | 11.09% | 12.17% |
| | **married** | 86.42% | 10.6% | 64.06% |
| | **single** | 1.95% | 78.31% | 23.78% |
| **Education** | **primary** | 27.44% | 6.19% | 6.23% |
| | **secondary** | 72.38% | 77.38% | 2.58% |
| | **tertiary** | 0.19% | 16.43% | 91.2% |
| **Occupation** | **admin.** | 14.14% | 16.23% | 3.35% |
| | **blue-collar** | 39.72% | 13.33% | 0.51% |
| | **entrepreneur** | 2.88% | 1.7% | 5.29% |
| | **housemaid** | 3.95% | 1.25% | 2.3% |
| | **management** | 2.54% | 2.74% | 67.51% |
| | **retired** | 8.71% | 0.68% | 3.04% |
| | **self-employed** | 2.71% | 3.14% | 5.29% |
| | **services** | 12.74% | 11.55% | 1.8% |
| | **student** | 0.1% | 6.33% | 0.3% |
| | **technician** | 9.22% | 39.83% | 8.44% |
| | **unemployed** | 3.29% | 3.21% | 2.18% |
| **Loans and Balances** | **housing** | 59.79% | 57.89% | 49.23% |
| | **default** | 1.8% | 2.03% | 1.63% |
| | **loan** | 18.75% | 15.28% | 13.96% |
| | **Median Balance ($)** | $425.0 | $367.0 | $567.0 |




**Segment 1**

- **Mature lifestage Low/Medium Affluent Segment**

→ Oldest segment, mostly married, less education, more loans and medium deposit balances

**Segment 2**

- **Young/Emerging Professionals Segment**

→ Youngest segment, mostly single, lowest deposit balance, majority secondary school

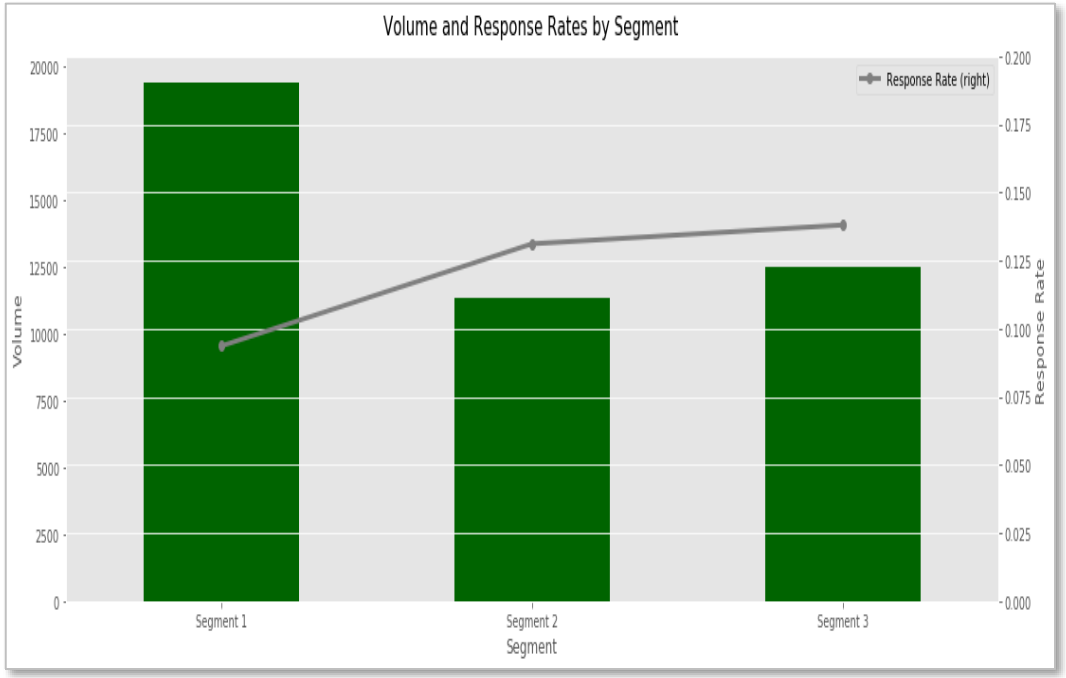
**Segment 3**

- **Family/Mature lifestage Medium/High Affluence Segment**

→ Pursued higher education, majority married, majority 'management' job

# Responder Profiles and Response Rates by Segment

| Responders | | Segment 1 | Segment 2 | Segment 3 |
|---|---|---|---|---|
| | Volume | 1,815 | 1,486 | 1,720 |
| | Avg. Age | 49.5 | 31.9 | 41.3 |
| Marital Status | divorced | 15.54% | 7.6% | 11.8% |
| | married | 82.42% | 7.47% | 57.91% |
| | single | 2.04% | 84.93% | 30.29% |
| Education | primary | 24.85% | 5.38% | 3.31% |
| | secondary | 75.04% | 70.52% | 1.8% |
| | tertiary | 0.11% | 24.09% | 94.88% |
| Occupation | admin. | 16.53% | 16.49% | 3.95% |
| | blue-collar | 27.66% | 11.1% | 0.47% |
| | entrepreneur | 2.37% | 1.62% | 2.85% |
| | housemaid | 3.31% | 1.62% | 1.22% |
| | management | 2.42% | 1.55% | 68.95% |
| | retired | 20.72% | 0.47% | 5.99% |
| | self-employed | 1.93% | 4.64% | 4.53% |
| | services | 10.3% | 9.76% | 1.05% |
| | student | 0.28% | 14.33% | 0.47% |
| | technician | 9.53% | 34.05% | 8.02% |
| | unemployed | 4.96% | 4.37% | 2.5% |
| Loans and Balances | housing | 40.22% | 40.11% | 31.86% |
| | default | 0.83% | 1.28% | 0.81% |
| | loan | 11.96% | 9.22% | 6.92% |
| | Median Balance ($) | $802.0 | $531.0 | $925.5 |




**Segment 1**

- **Mature lifestage Low/Medium Affluent Segment**

→ Oldest segment, mostly married, less education, more loans and medium deposit balances

**Segment 2**

- **Young/Emerging Professionals Segment**

→ Youngest segment, mostly single, lowest deposit balance, majority secondary school

**Segment 3**

- **Family/Mature lifestage Medium/High Affluence Segment**

→ Pursued higher education, majority married, majority ‘management’ job

# Classification Model Approach - Model Design

## Algorithm

**Logistic Regression**

Extension of linear regression model for classification

**Random Forest with SMOTE**

Bagging-based algorithm using a subset of features selected random to build a forest of trees

**XGBoost**

Tree-based ensemble algorithm using gradient boosting framework

## Evaluate Metrics

**AUC (Area Under Curve)**

Evaluation metric that tells how much model is capable of distinguishing between classes

**Decile Lift**

Predicted Probability by Decile / Average Rate

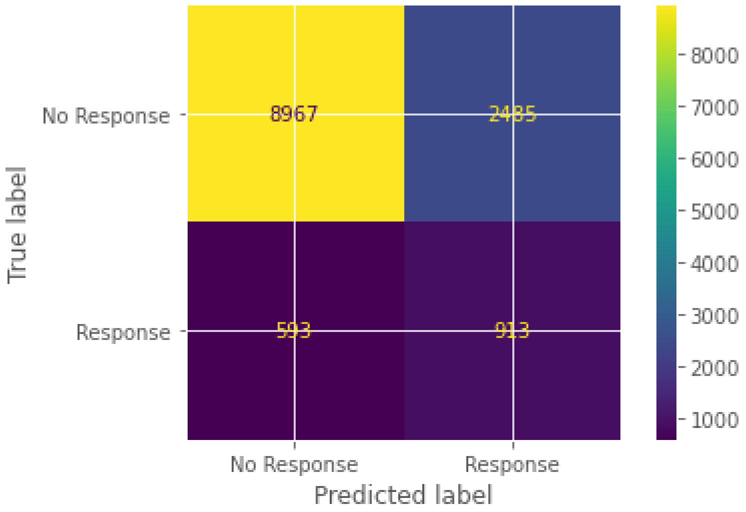
## Interpretation

**Shapley Value**

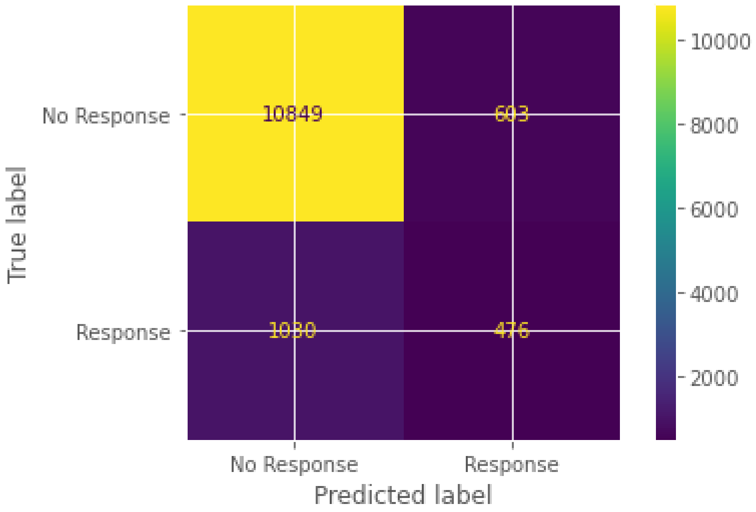
Average marginal contributions across all permutations

# Model Comparison

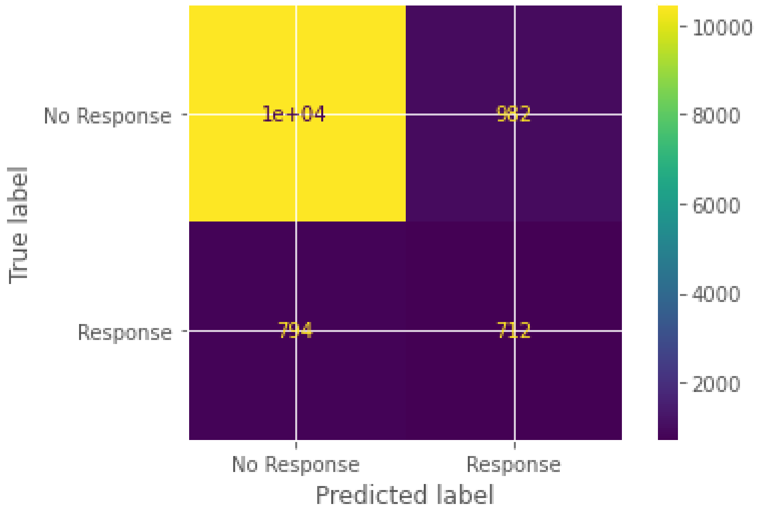
**Logistic Regression**




**Random Forest with SMOTE**



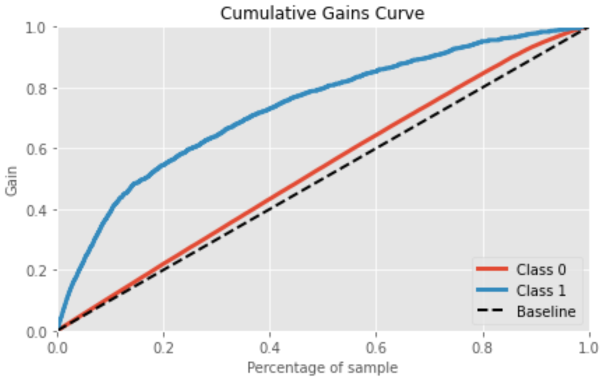

**XGBoost**

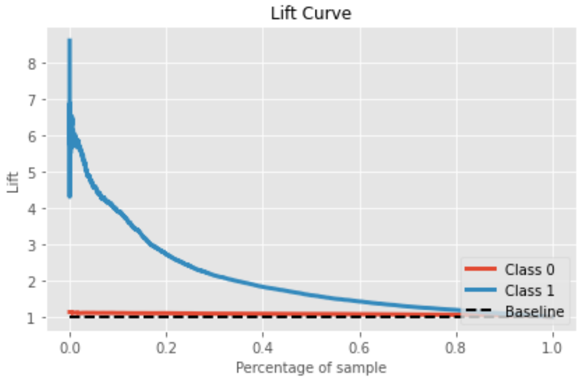



| | Logistic Regression | Random Forest with SMOTE | XGBoost |
|---|---|---|---|
| **AUC (Area Under ROC Curve)** | 0.695 | 0.632 | 0.694 |
| **Time Took** | 12.1 sec | 7.9 min | 4.6 min |
| **Recall** | 0.69 | 0.32 | 0.47 |
| **Precision** | 0.27 | 0.44 | 0.42 |

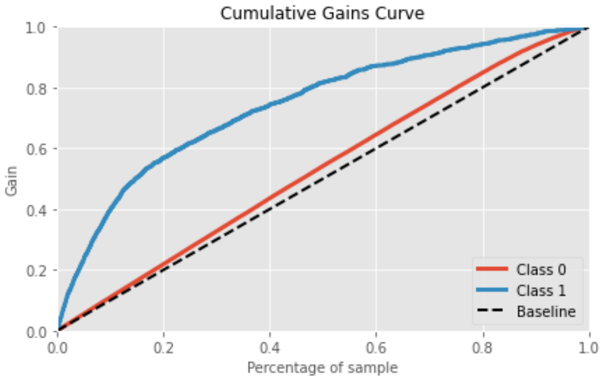
# Model Comparison - Continued

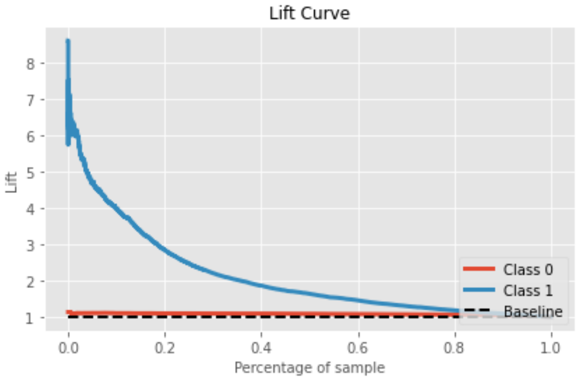
**Logistic Regression**



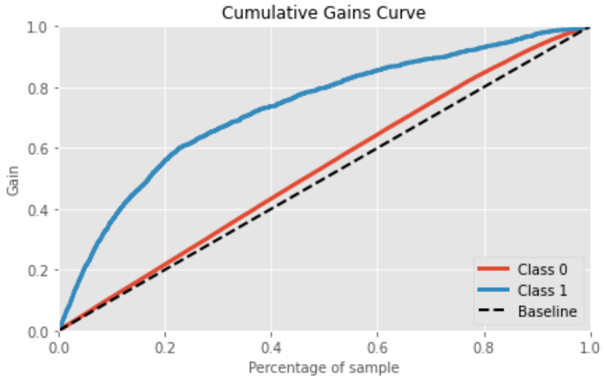



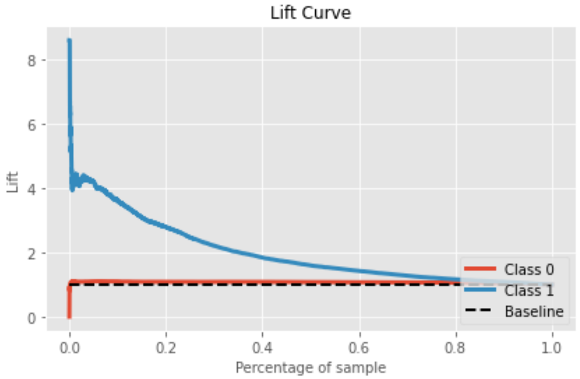

**XGBoost**




**Random Forest with SMOTE**




- All three models show very similar patterns in both plots.
  - ~70% of the responses are captured within 3 deciles
  - ~300% lift around decile 2

# Final Thoughts and Next Step

- A rule-based engine based on the responder profiles and exploratory analysis would provide the flexibility in the target selection. However, it may require some time and energy to reach the (sub-)optimal rule that satisfies ROI goals of the campaign.

- Propensity model gives the option to score each individual customers and choose how many of them we want to target (e.g. up to 20% or 30%?).
  - As the lift curve gradually flattens out around decile 3 and 4, we may want to consider incentivizing for those population to make the offer more attractive.
  - Consider expanding this to other banking products and services to help the contact center to prioritize the calling effort.

- With the limited data, segmentation approach was not very effective due to large amount of overlaps across clusters. More exploration needed to capture different response behaviors across different customer segments
  - Propensity models were fitted for each segment identified from segmentation exercise as an experiment. However, it did not show any noticeable improvement in the model performances