

Predictive Analytics in Marketing: Term Deposit Marketing Campaign

Darren Tay

Data Science Immersive 6



Agenda

A

Business Problem

B

About the Data

C

Models & Results

D

Implications & Next Steps

A1. Banks try to gain maximum utility from their customer base by cross-selling to them



A2. Customer leads are a scarce resource, thus the need to correctly identify customers to be included in a specific product call campaign



Limited Number of Customers

Not everyone in the world is a client of the bank.



Call Fatigue of Customers

Overcalling customers could lead to negative repercussions

Agenda

A

Business Problem

B

About the Data

C

Results & Findings

D

Implications & Next Steps

B1. Data comprises of 4 main groups: Bank client, Current campaign, Socio-Economic and Previous campaign data



Bank Client

- Age
- Job
- Marital Status
- Education
- Loan

Current Campaign

- Contact Info
- Last Contact Duration
- Number of Contacts

Socio-Economic

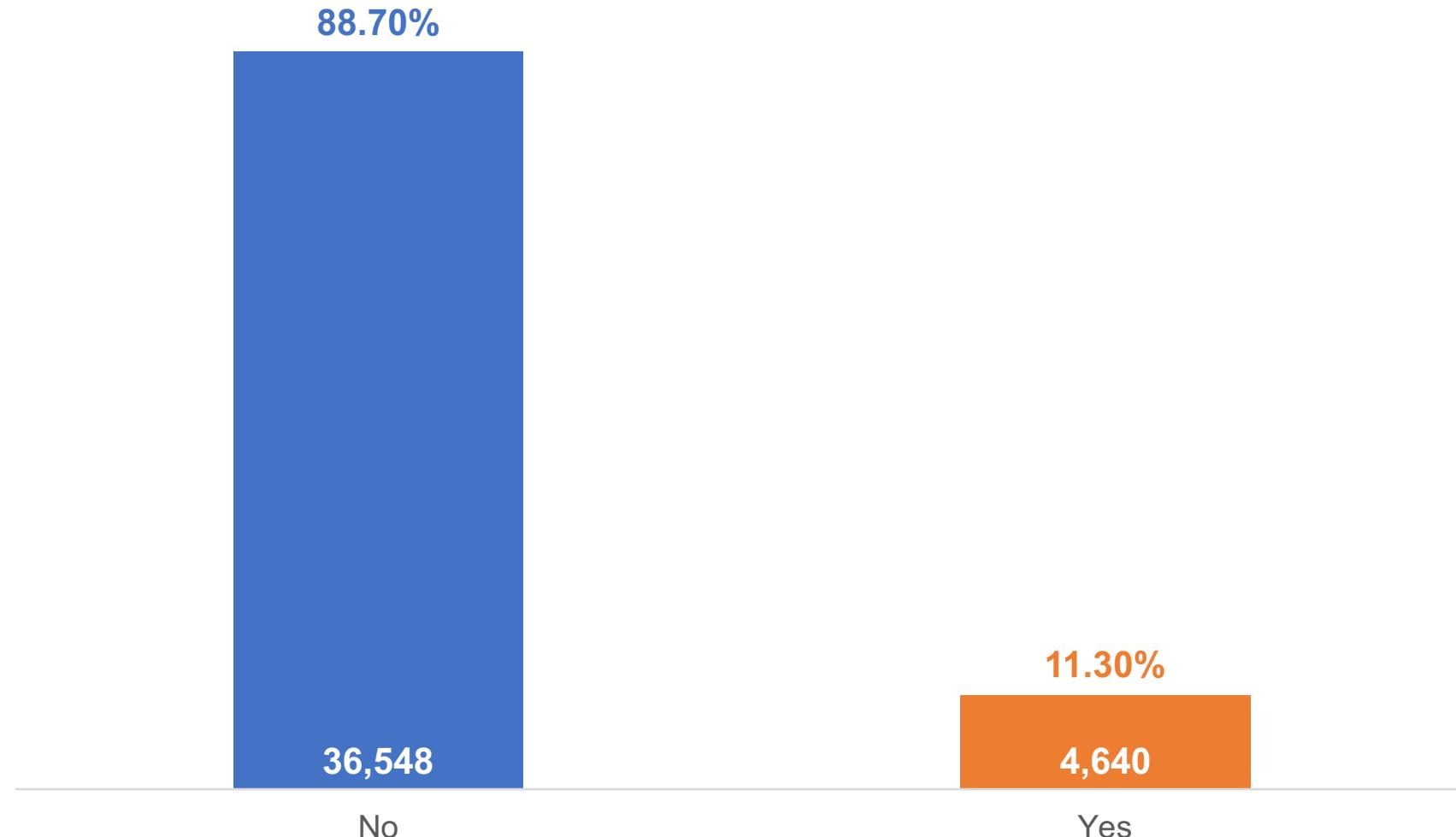
- Employment
- CPI
- Interest Rate

Previous Campaign

- Outcome
- Days since last campaign

B2. Out of 41,188 customers, only 4,640 signed up for a TD (conversion rate of 11.3%)

Clients who Subscribed to Term Deposit



Agenda

A

Business Problem

B

About the Data

C

Results & Findings

D

Implications & Next Steps

C1. Current call campaign information to be excluded from modelling as this information is not available when selecting customers & is highly predictive of the outcome



Current Campaign

- Last call duration highly affects the outcome
- People who are not interested would not speak long with the representative

C2. Best model: Random Forest after applying Recursive Feature Elimination

Logistic Regression

- AUC: **0.7824**
- Stacked Model (AUC): **0.7965**

Random Forest

- AUC: **0.8231**
- Stacked Model (AUC): **0.7982**



K-Nearest Neighbours

- AUC: **0.7643**
- Stacked Model (AUC): **0.7928**

XGBoost

- AUC: **0.8062**
- Stacked Model (AUC): **0.7983**

C3. The most important features for predicting TD subscription are socio-economic factors, age and time since previous campaign



Important Features

3M Euribor Rate

Market Interest Rate

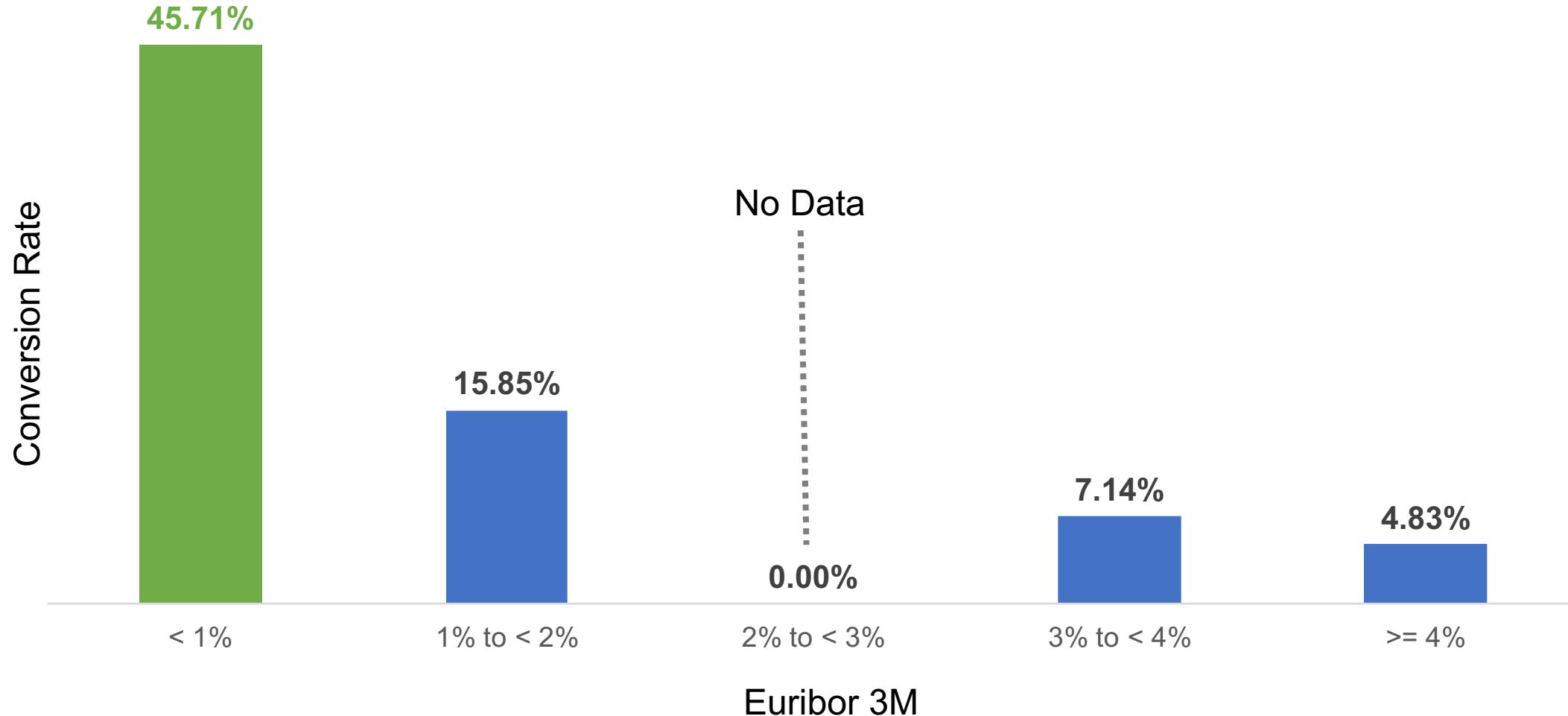
Number of Employees

Quarterly average of employed citizens

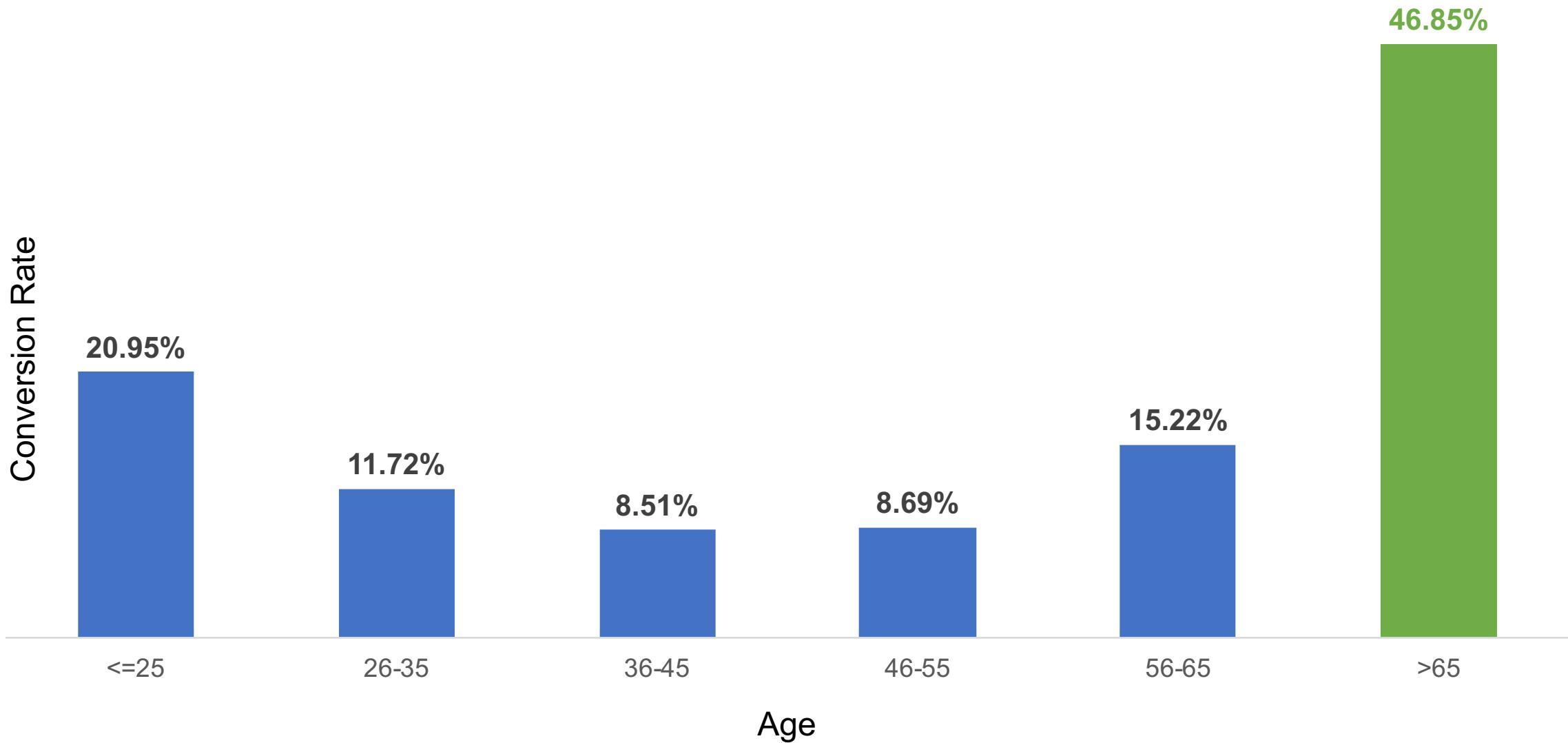
Number of **days after last contact** from **previous** campaign

Age of customer

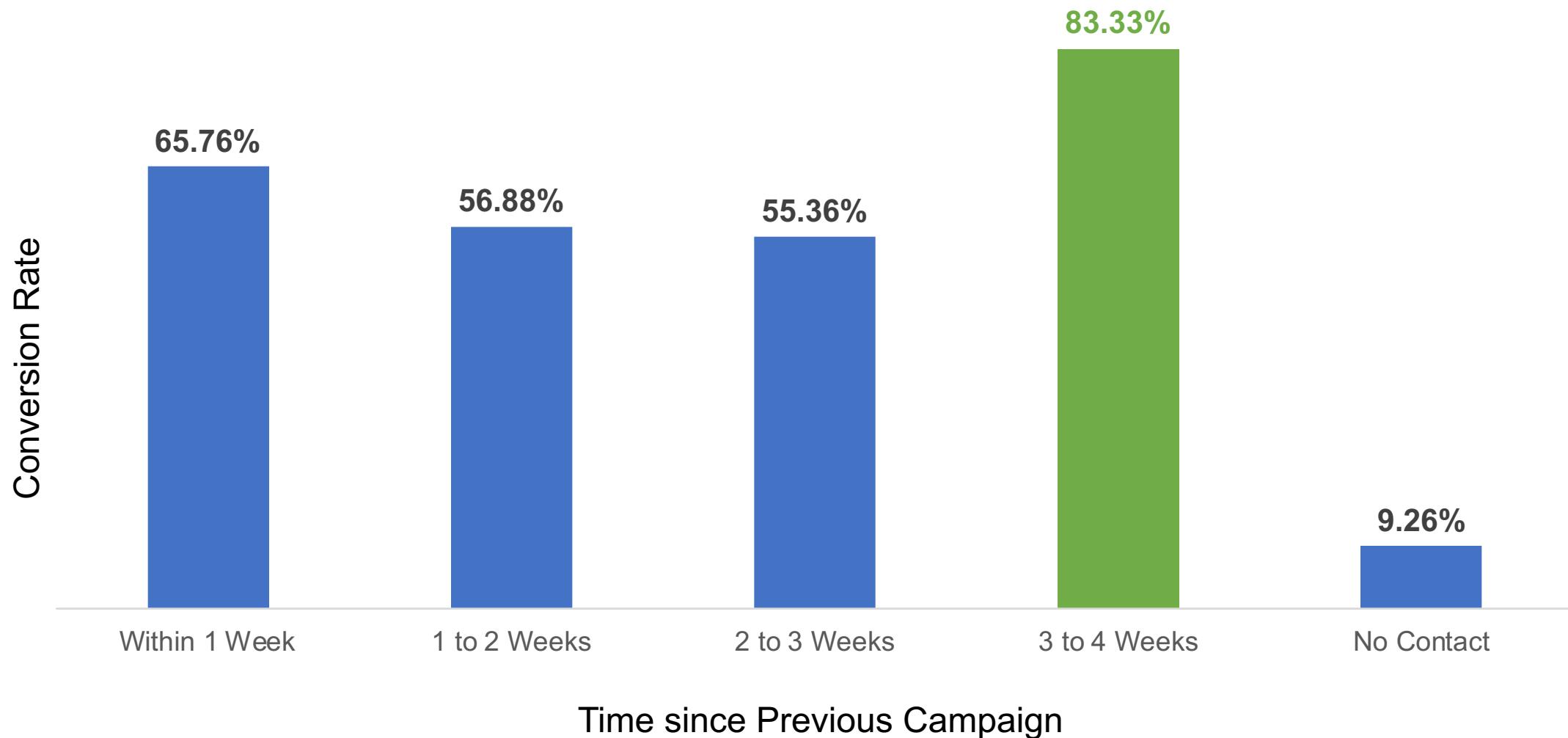
C4. Higher conversion rate when market interest rates are low



C5. Retired customers have the highest conversion rate, followed by young adults



C6. Customers who have at least 3 weeks gap since being last contacted on a previous campaign have the highest conversion rate



Agenda

A

Business Problem

B

About the Data

C

Results & Findings

D

Implications & Next Steps

D1. After tuning the cut-off, the model gives a precision of 0.36 & recall of 0.66

		Predicted	
		Not Converted	Converted
Actual	Not Converted	7,793	1,344
	Converted	393	767

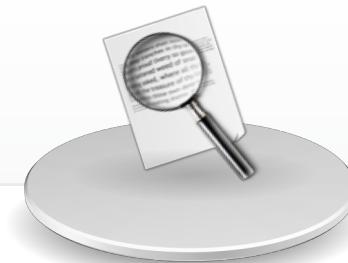
D1a. Increasing the precision to 0.84, reduces recall to 0.04. However, this leads to a massive loss of leads to call

		Predicted	
		Not Converted	Converted
Actual	Not Converted	9,127	10
	Converted	1,108	52

D2. Applying the model would enable the bank to save costs while minimizing profit loss

	Before	After
Customer Count	41.1k	8.4k
Calling Costs	36.5k	5.3k
Opportunity Costs	4.6k	3.1k
Profit	n/a	1.5k
Loss of Profit	n/a	32.7k
Leads saved	n/a	

D3. Exploring other pieces of information that could be useful in predicting TD subscription

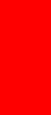


Assets & Loans Under Management

People with more assets with the bank tend to sign up for TD?

Risk Appetite

TD appeals more to customers with lower risk appetite?



Thank you