

RETHINKING  
THE BOUNDARIES

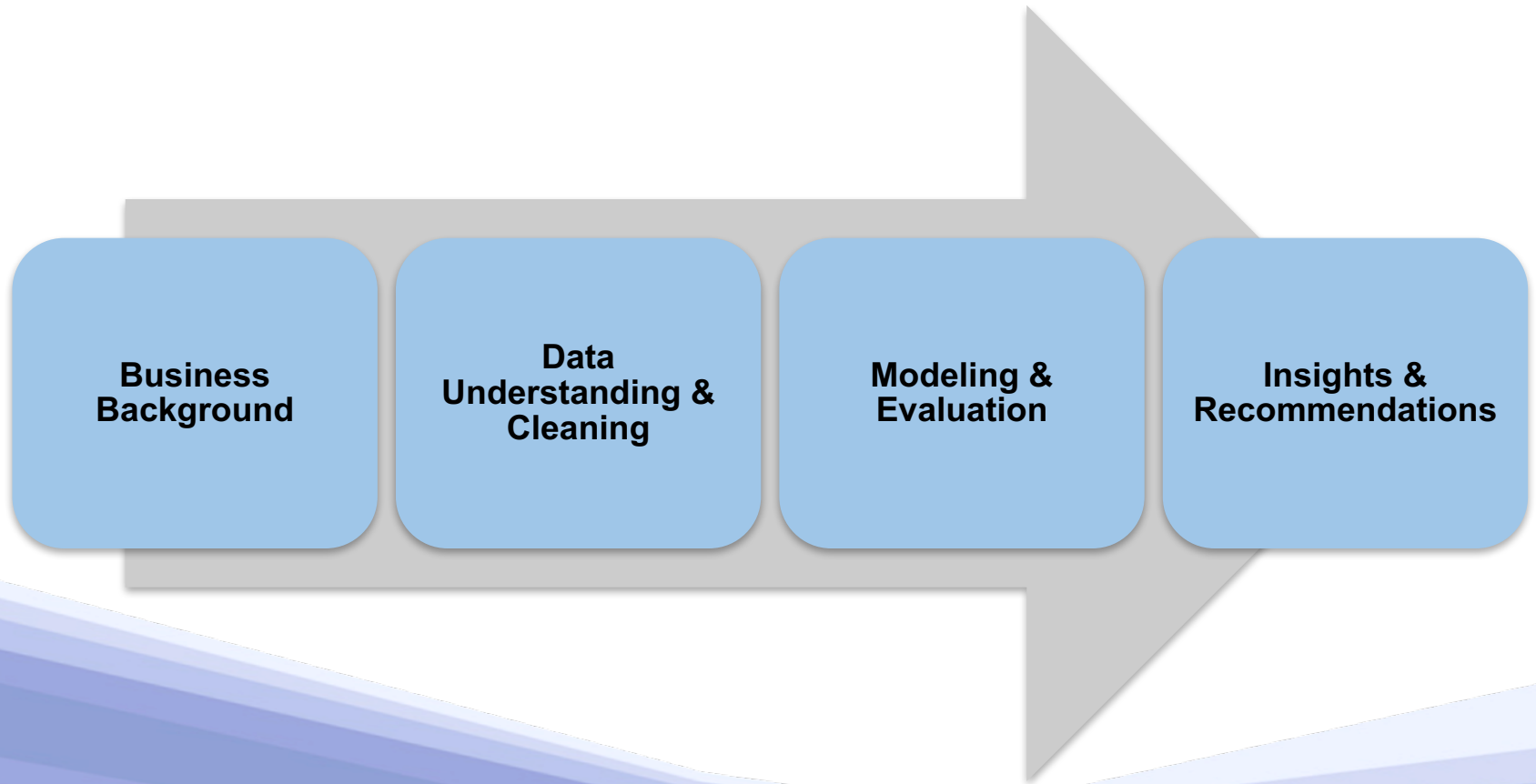


# Data Science for Business

Aaron Rodriguez  
Yin-Ta Pan  
Davide Sgarbi  
Flora Yang



# Agenda



# Business Background

- Given data regarding a Portuguese banking institution's direct marketing campaign
- For any company, being able to evaluate the effectiveness of a marketing campaign is crucial
  - Increase revenue generated from customers
  - Cut the cost of wasted calls
- 3 things to evaluate
  - Customers with highest probability of signing up through a marketing phone call campaign
  - the probability of a customer signing up required to have a positive expected value for the call
  - other miscellaneous insights related to marketing campaigns that can be derived from the data

# Data Understanding & Cleaning

Customer Information

variables	Type	Note
age	int	
job	factor	
marital	factor	
education	factor	
default	factor	has credit card in default?
housing	factor	has housing loan?
loan	factor	has personal loan?

Social Index

variables	Type	Note
emp.var.rate	num	employment variation rate
cons.price.idx	num	consumer price index
cons.conf.idx	num	consumer confidence index
euribor3m	num	euribor 3 month rate
nr.employed	num	number of employees

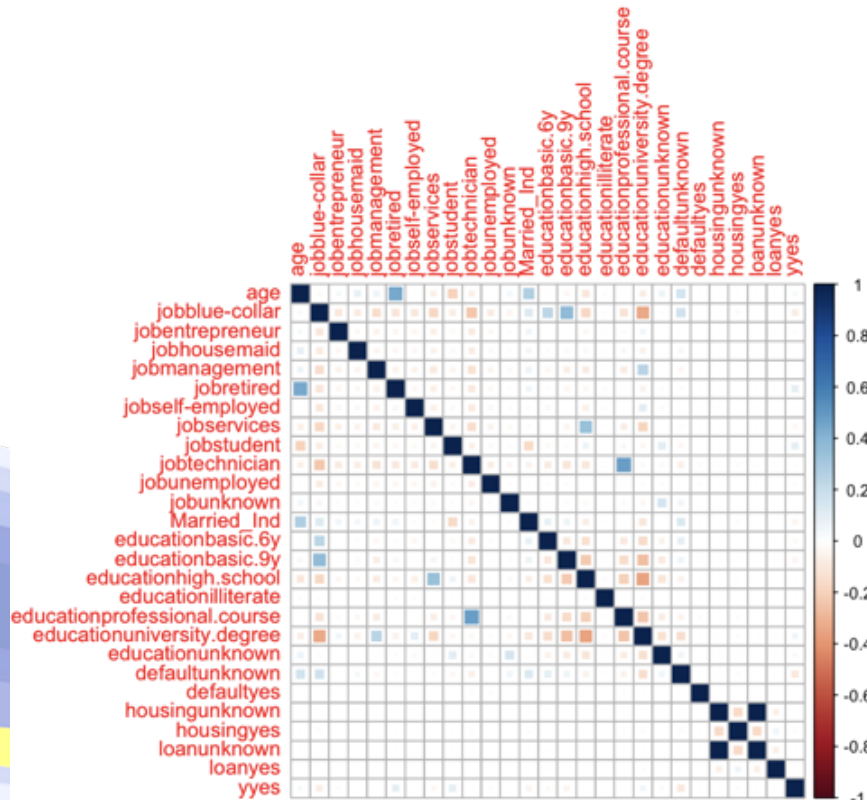
Information about last call

variables	Type	Note
contact	factor	contact communication type (cellular/ telephone)
month	factor	last contact month of the year
day_of_week	factor	last contact day of the week
duration	int	last contact duration (afterward index)
campaign	int	number of contacts performed during this campaign and for this client
previous	int	number of contacts performed before this campaign and for this client
poutcome	factor	outcome of the previous marketing campaign

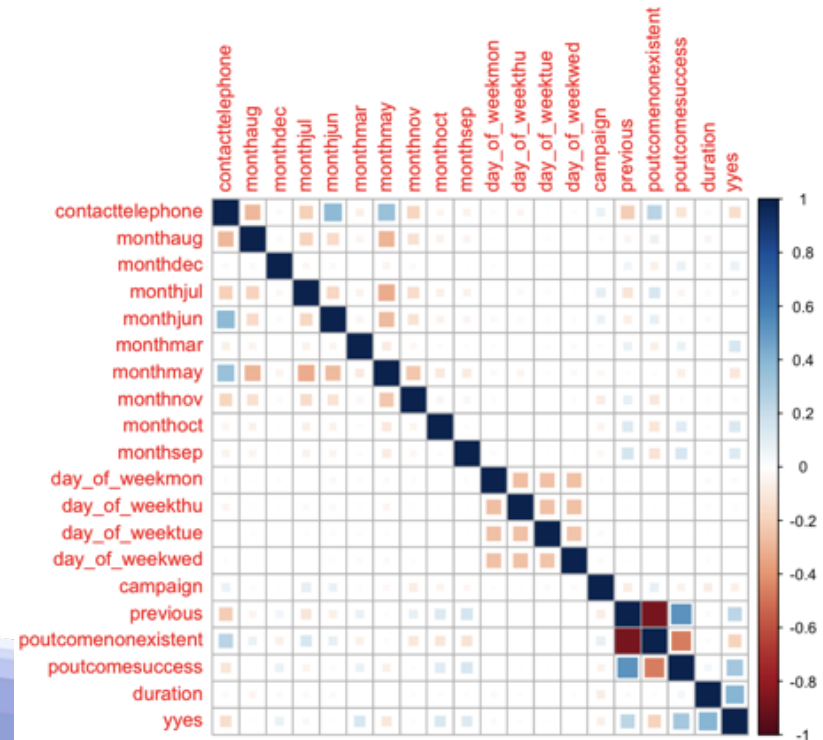
# Exploratory Data Analysis

## General Correlation

### Customers' Characteristics



### Information about last call



# Exploratory Data Analysis

## General Correlation

### Economic Indexes



# Modeling & Evaluation

## Simple Model

## Complex Model



**Logistic  
Model**

**Decision  
Tree**

**Logistic  
with  
interactions  
regularized  
via Lasso**

**Principal  
Component  
Analysis  
(PCA)**

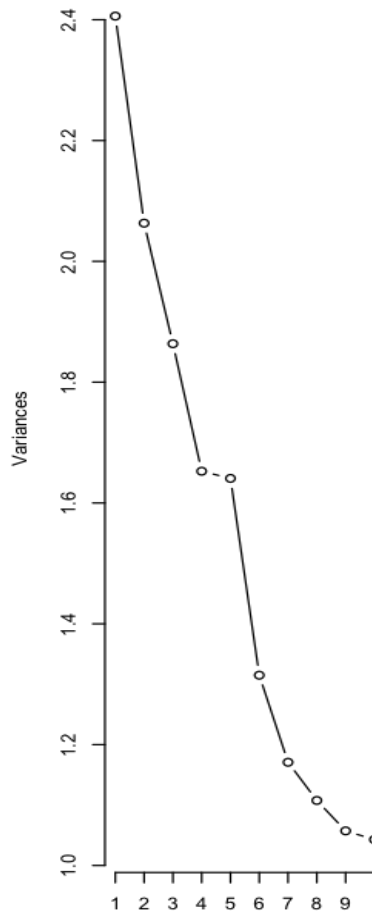
**We could be sure of neither overfit nor underfit the data when running cross validation!**



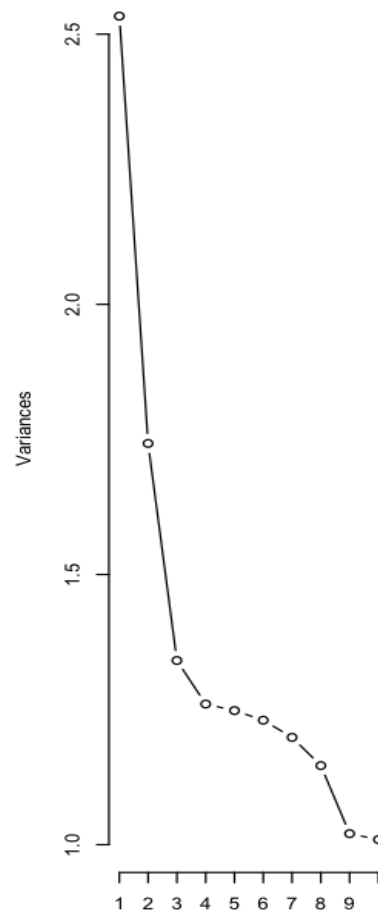
# Modeling & Evaluation II

## PCA

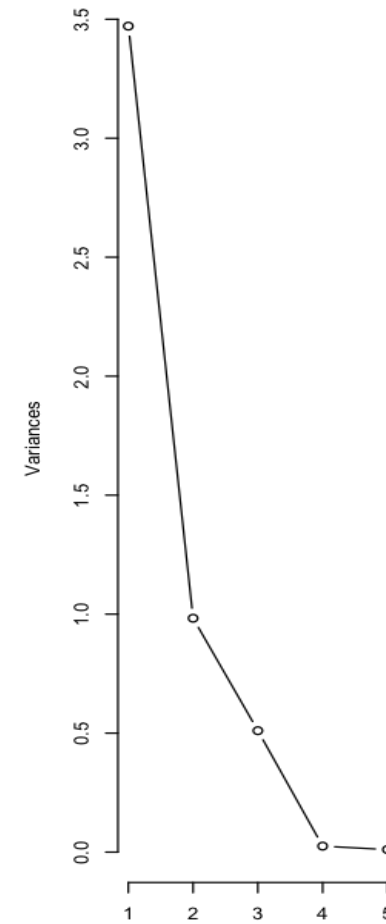
Customers' characteristic



Information about last call



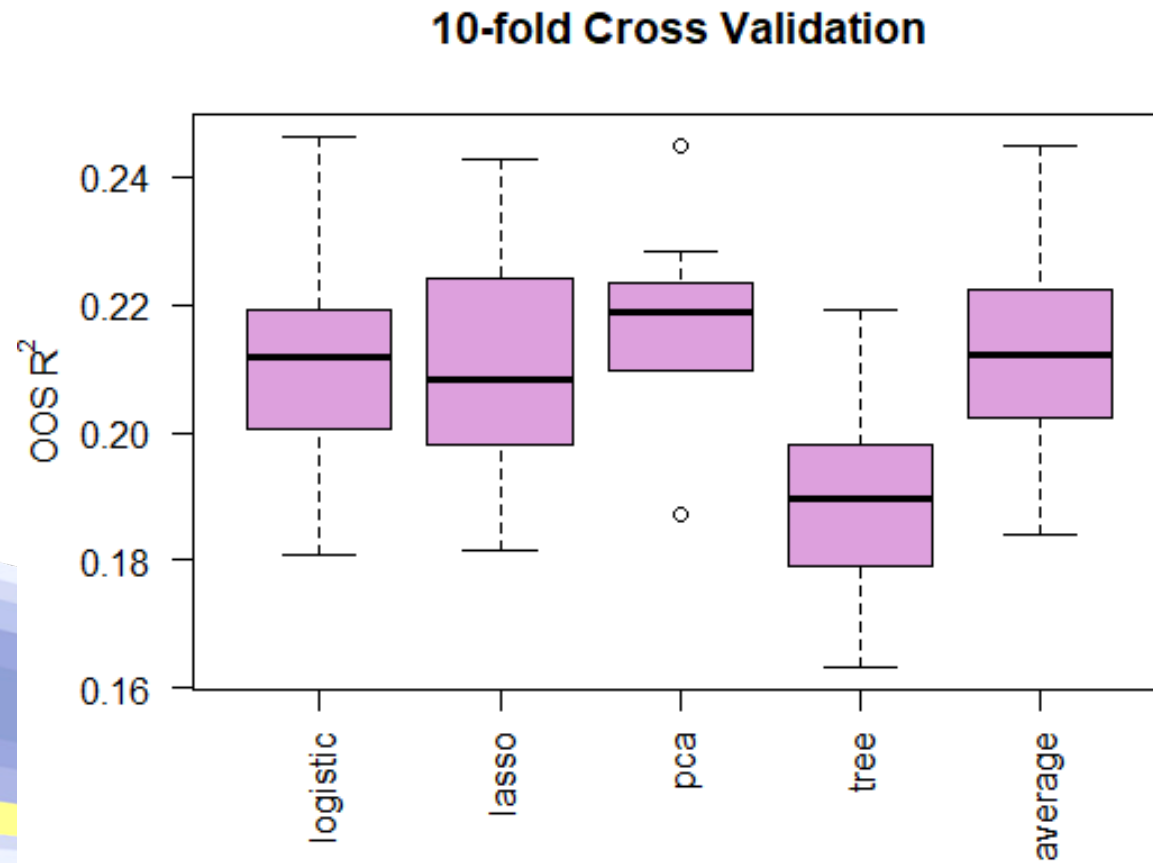
Economic indexes





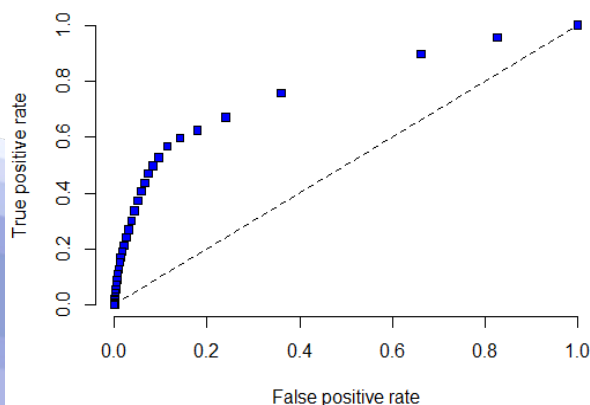
# Modeling & Evaluation III

## 10-fold Cross Validation



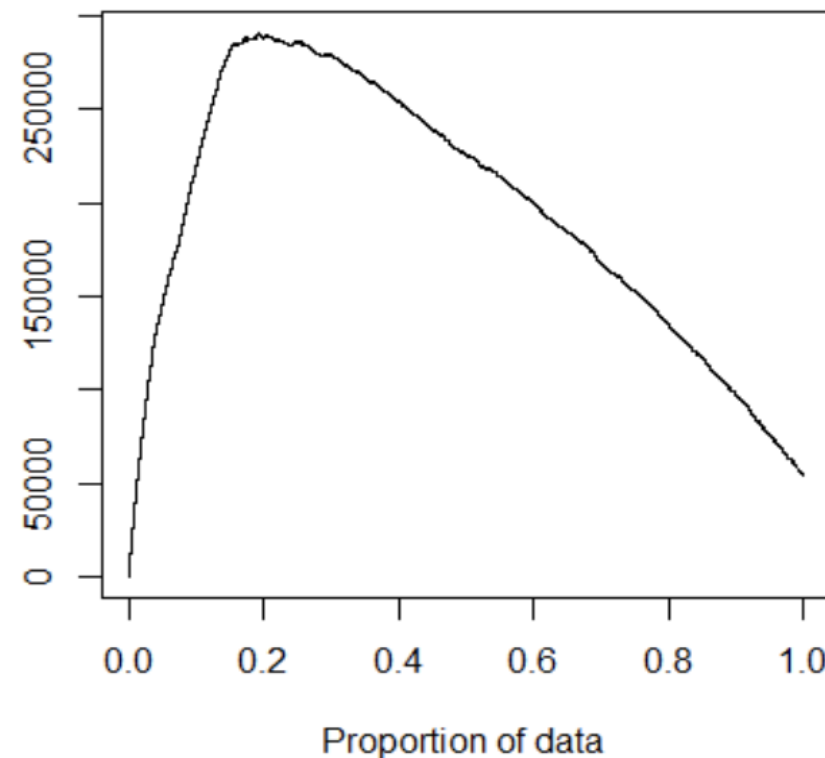
# Modeling & Evaluation IV

	Y = 1 (P)	Y = 0 (N)
$\hat{Y} = 1$	130	-15
$\hat{Y} = 0$	0	0



Profit

### Profit Curve



# Insights and Recommendations

I

By using the predictive model, calls will be successful 5 out of 7 times (instead of 1 out of 9)

II

Huge cost savings for both operating costs and OH; increasing profits from a baseline scenario of \$54,315 to a much higher **\$265,680**

III

Cold-call campaigns will not be that boring in the (near) future after all...

**THANK  
YOU...**