



# Bank Marketing Analysis

Presented by Keren Vasconcelos

# BUSINESS CONTEXT

- Nowadays, marketing spending in the banking industry is massive, meaning that it is essential for banks to optimize marketing strategies and improve effectiveness.
- Understanding customers' need leads to more effective marketing plans, smarter product designs and greater customer satisfaction.



Keren Vasconcelos. • May.2020

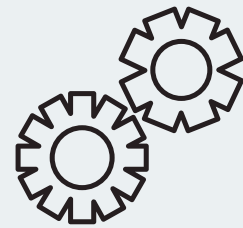
# Project Workflow

**Research on topic/  
Find an  
appropriate  
dataset**



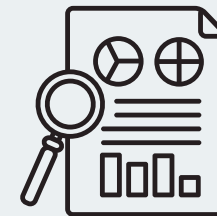
Didactic dataset  
and suitable for  
marketing analytics

**Cleaning and  
Manipulate**



Get Dummies  
Convert response  
into binary

**Exploratory Data  
Analysis**



Understand the  
customers  
Extract KPI's

**Modelling**



Binary response  
Logistic Regression  
Unbalanced data

# DATASET

## BANK TELEMARKETING

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls.

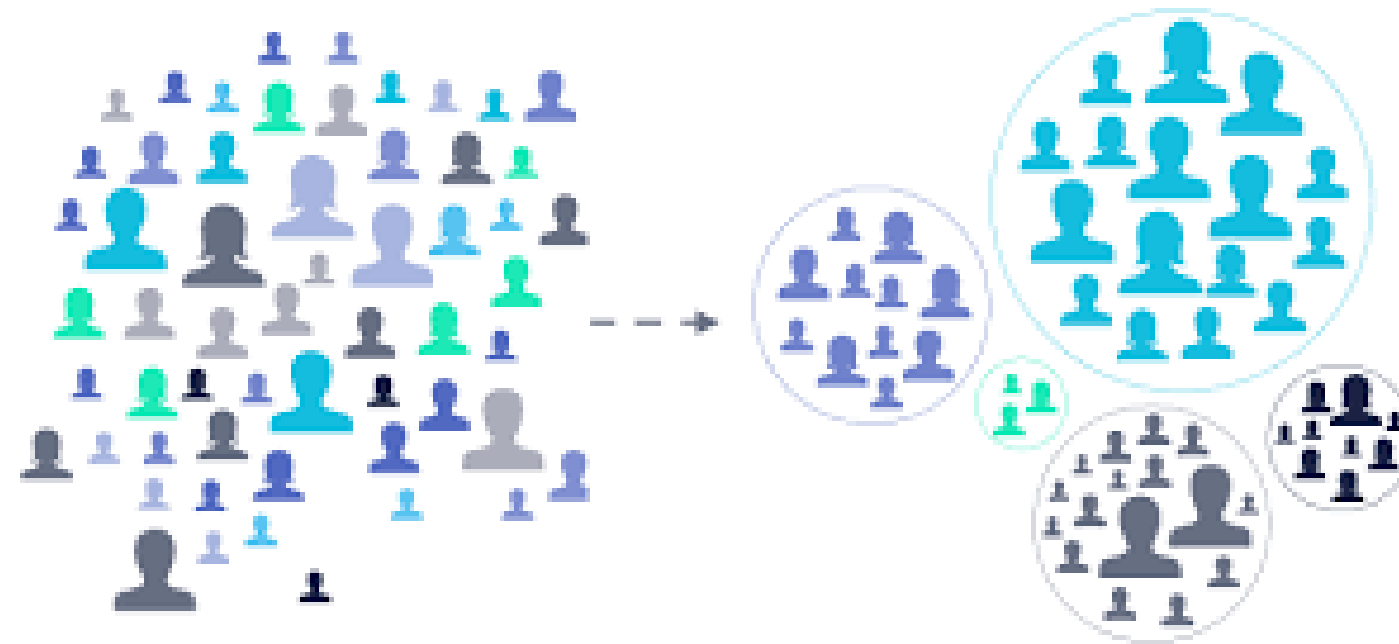
Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be (or not) subscribed.



## DATA SHAPE

- 45211 rows | 17 columns;
- Bank client data (e.g. age, job...);
- Related with the last contact of the current campaign(e.g. duration, day..);
- Desired target: Response - has the client subscribed a term deposit? (binary: "yes","no")

Which customer will  
subscribe for the bank  
deposit?



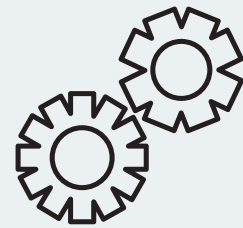
# Project Workflow

**Research on topic/  
Find an  
appropriate  
dataset**



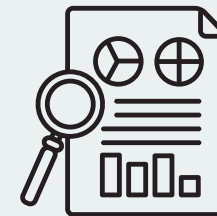
Didactic dataset  
and suitable for  
marketing analytics

**Cleaning and  
Manipulate**



Get Dummies  
Convert response  
into binary

**Exploratory Data  
Analysis**



Understand the  
customers  
Extract KPI's

**Modelling**



Binary response  
Logistic Regression  
Unbalanced data

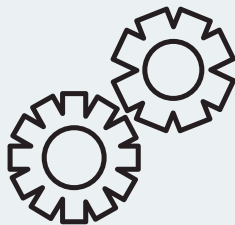
# Project Workflow

Research on topic/  
Find an  
appropriate  
dataset



Didactic dataset  
and suitable for  
marketing analytics

Cleaning and  
Manipulate



Get Dummies  
Convert response  
into binary

	age	balance	day	duration	campaign	pdays	previous	y	response	d__admin.	d__blue-collar
0	58.0	2143.0	5.0	4.35	1.0	-1.0	0.0	no	0	0	0
1	44.0	29.0	5.0	2.52	1.0	-1.0	0.0	no	0	0	0
2	33.0	2.0	5.0	1.27	1.0	-1.0	0.0	no	0	0	0
3	47.0	1506.0	5.0	1.53	1.0	-1.0	0.0	no	0	0	1
4	33.0	1.0	5.0	3.30	1.0	-1.0	0.0	no	0	0	0
5	35.0	231.0	5.0	2.32	1.0	-1.0	0.0	no	0	0	0
6	28.0	447.0	5.0	3.62	1.0	-1.0	0.0	no	0	0	0
7	42.0	2.0	5.0	6.33	1.0	-1.0	0.0	no	0	0	0
8	58.0	121.0	5.0	0.83	1.0	-1.0	0.0	no	0	0	0
9	43.0	593.0	5.0	0.92	1.0	-1.0	0.0	no	0	0	0

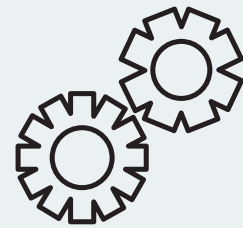
# Project Workflow

**Research on topic/  
Find an  
appropriate  
dataset**



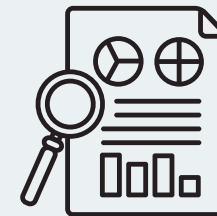
Didactic dataset  
and suitable for  
marketing analytics

**Cleaning and  
Manipulate**



Get Dummies  
Convert response  
into binary

**Exploratory Data  
Analysis**



Understand the  
customers  
Extract KPI's

**Modelling**



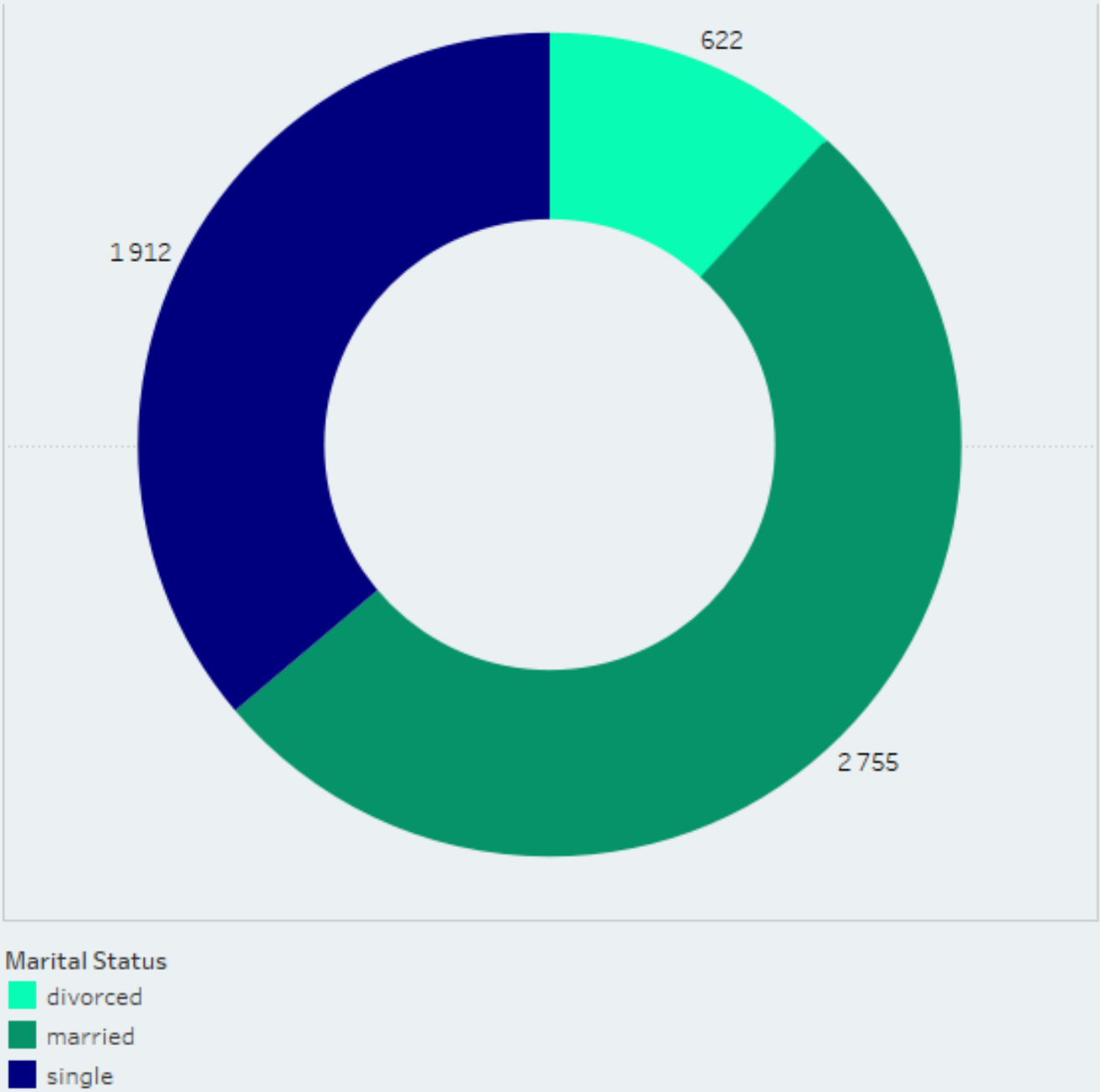
Binary response  
Logistic Regression  
Unbalanced data



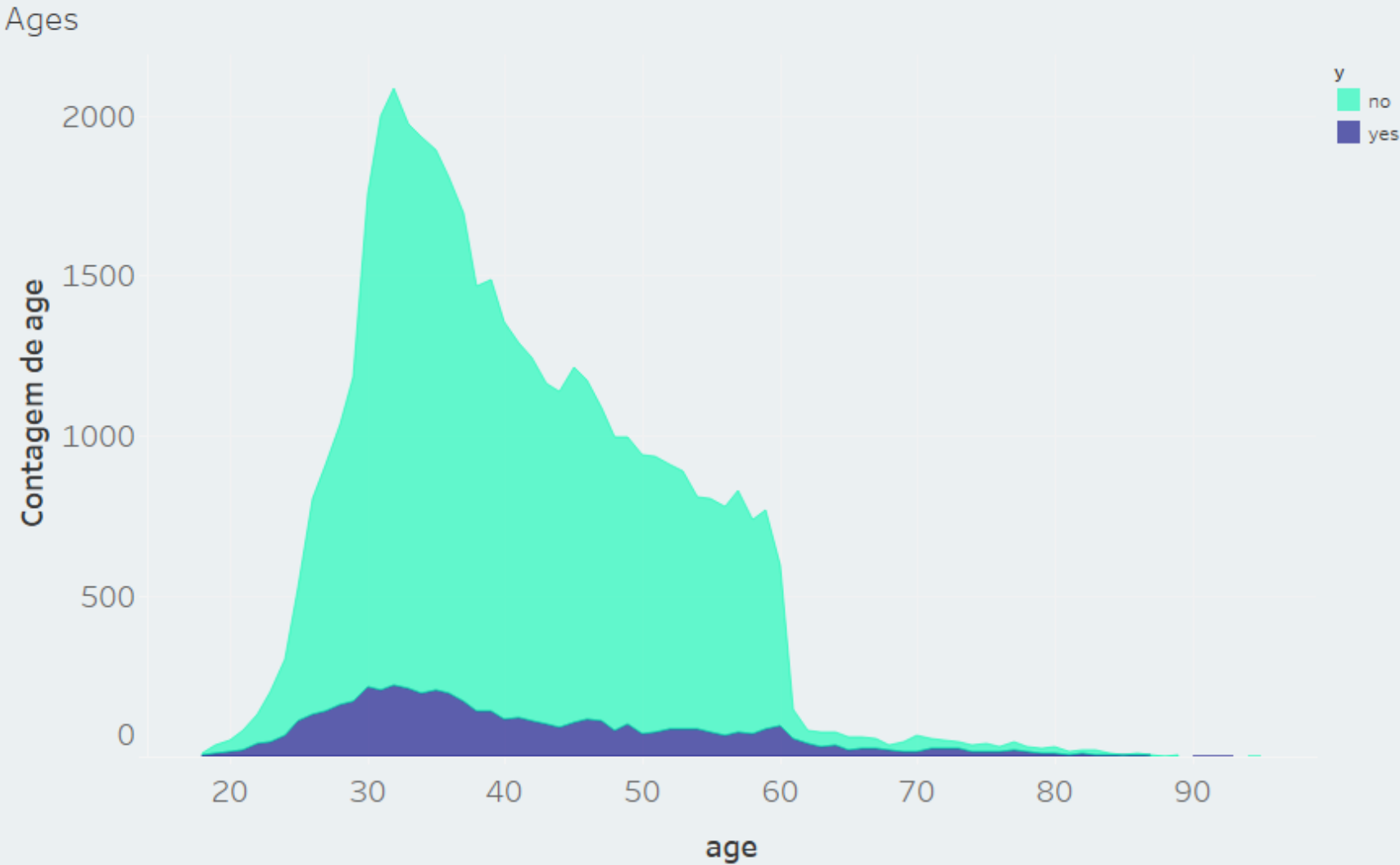
# WHO ARE THE COSTUMERS?

## EXPLORATORY DATA ANALYSIS

Marital Status



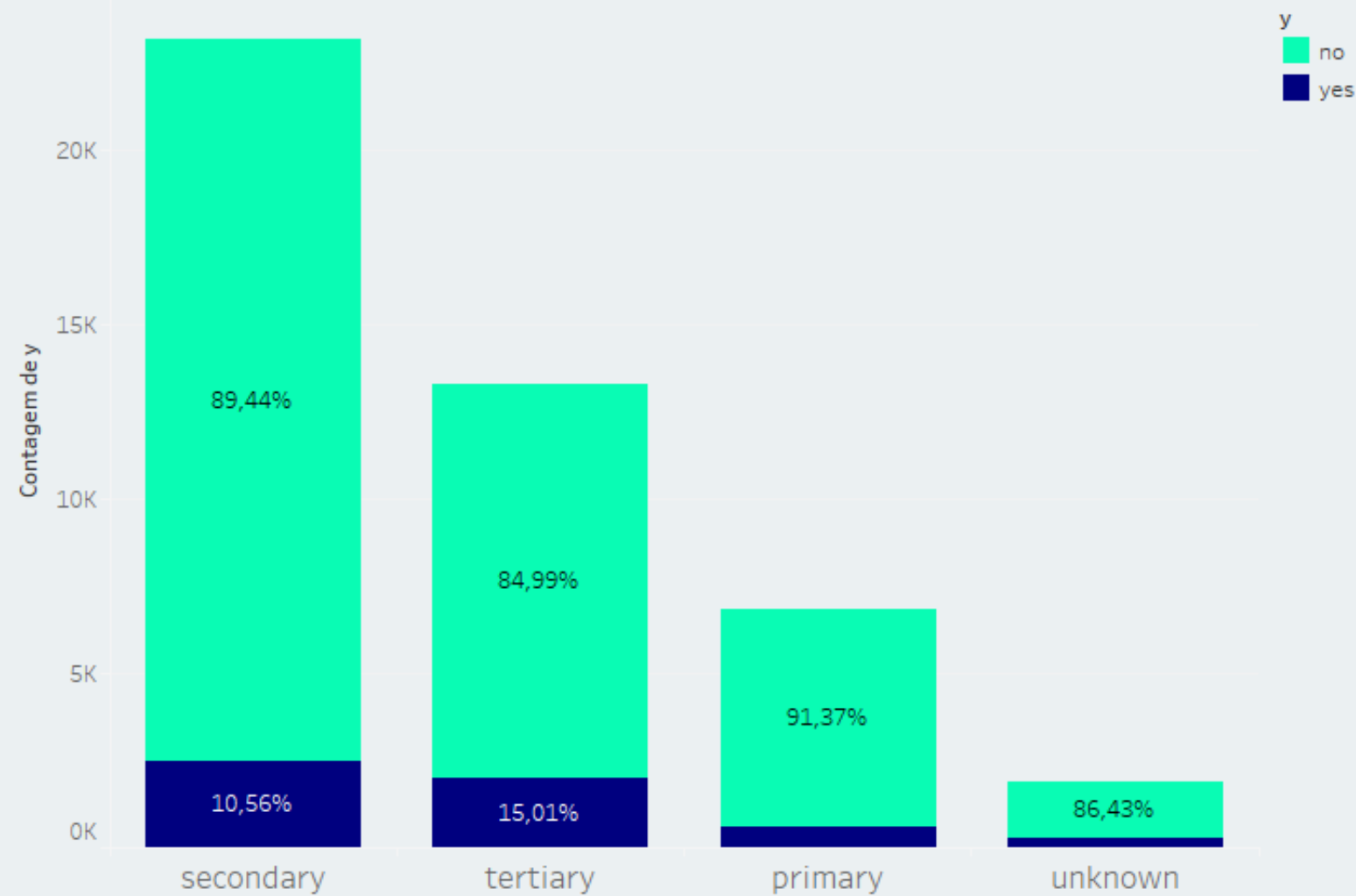
Ages



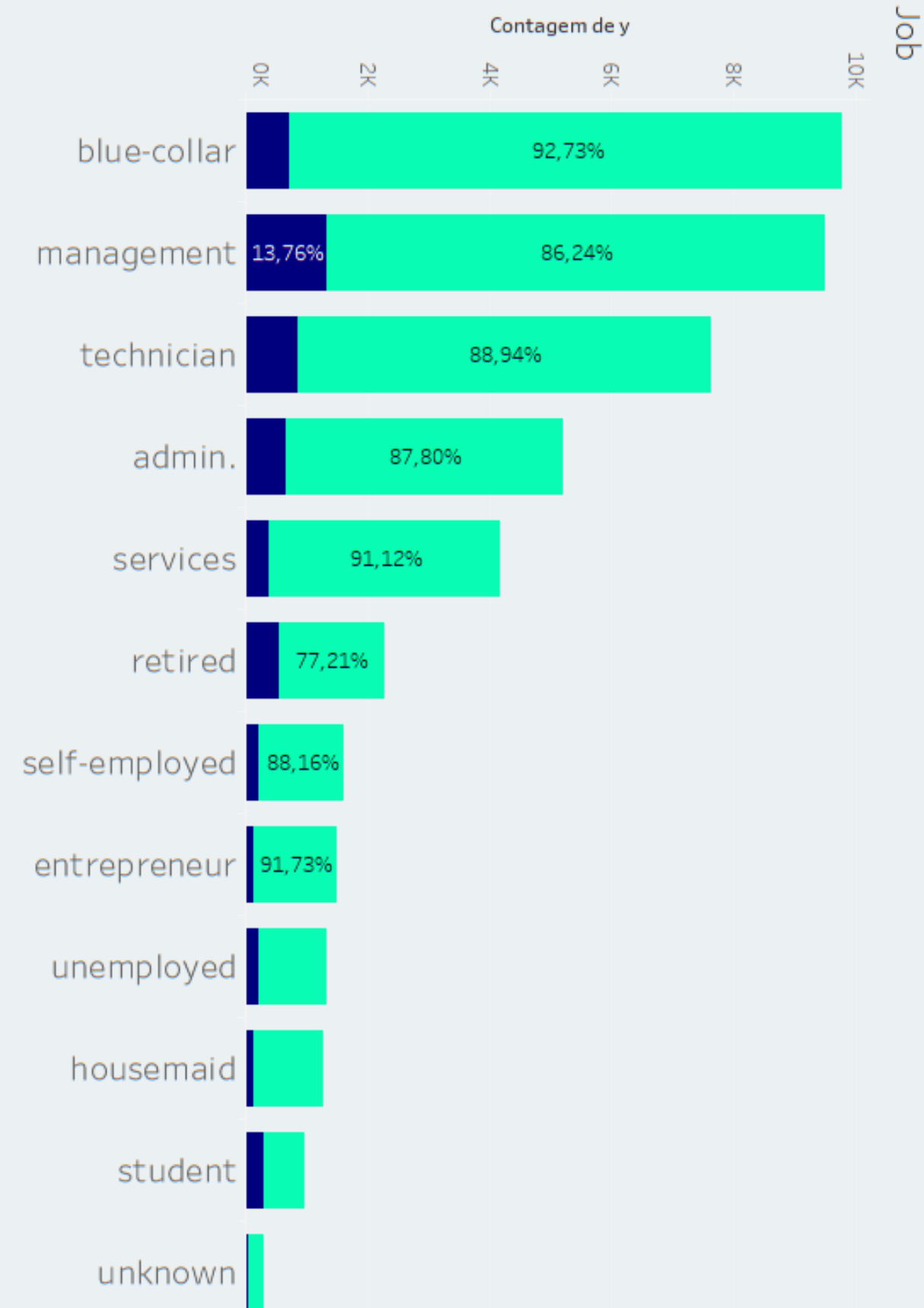
# WHO ARE THE COSTUMERS?

## EXPLORATORY DATA ANALYSIS

### Education

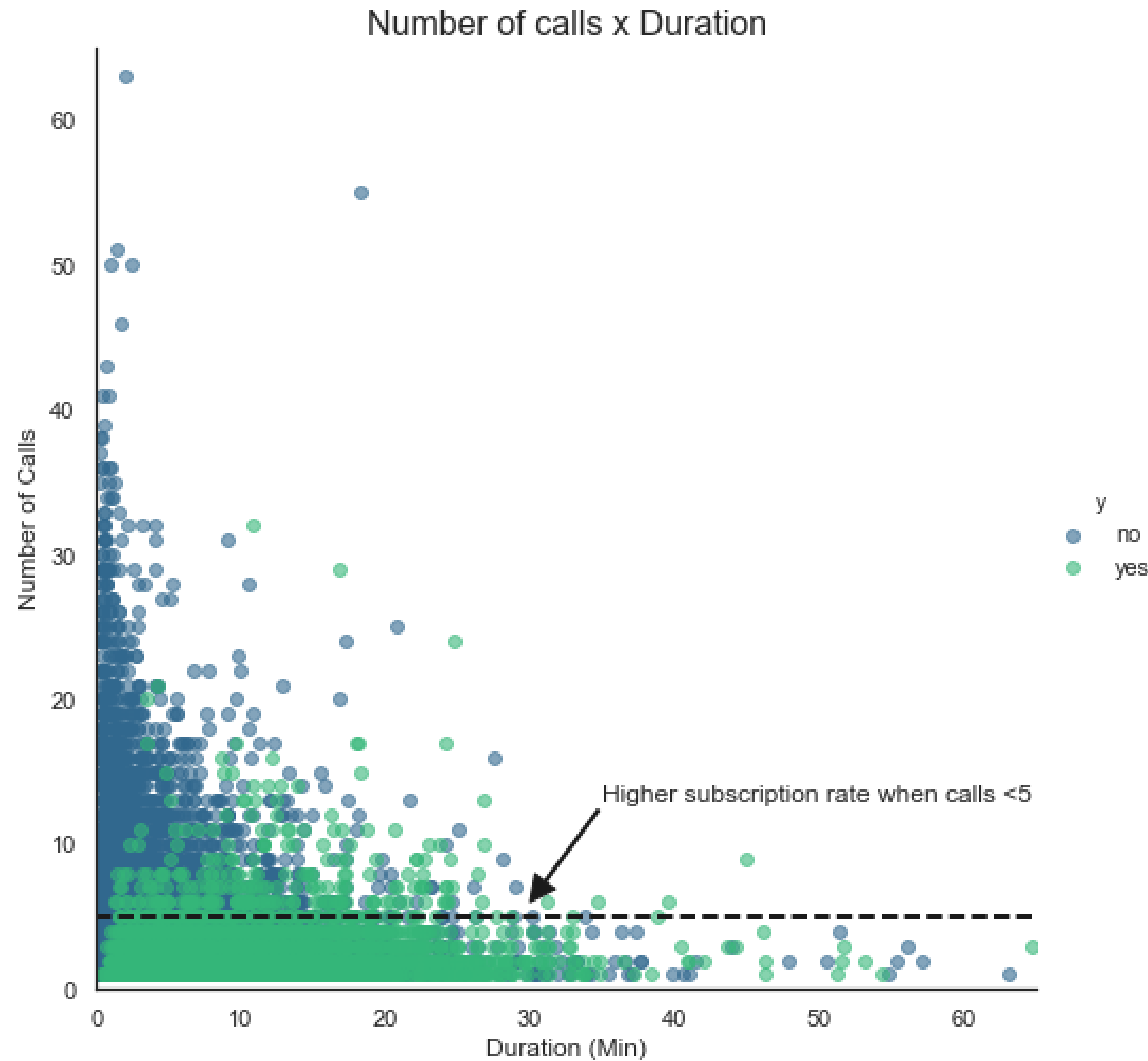


### Job



# BUSINESS INSIGHTS

## POSITIVE RESULTS



- Compared to “no” clients”, “yes” clients were contacted by fewer times and had longer call duration.
- More importantly, after five campaign calls, clients are more likely to reject the term deposit unless the duration is high.
- Most “yes” clients were approached by less than 10 times.

# BUSINESS INSIGHTS

POSITIVE RESULTS



**9 MIN**

AVERAGE CALL  
DURATION

**2**

AVERAGE OF  
CONTACTS  
PERFORMED  
DURING THIS  
CAMPAIGN

# BUSINESS INSIGHTS

POSITIVE RESULTS



4868

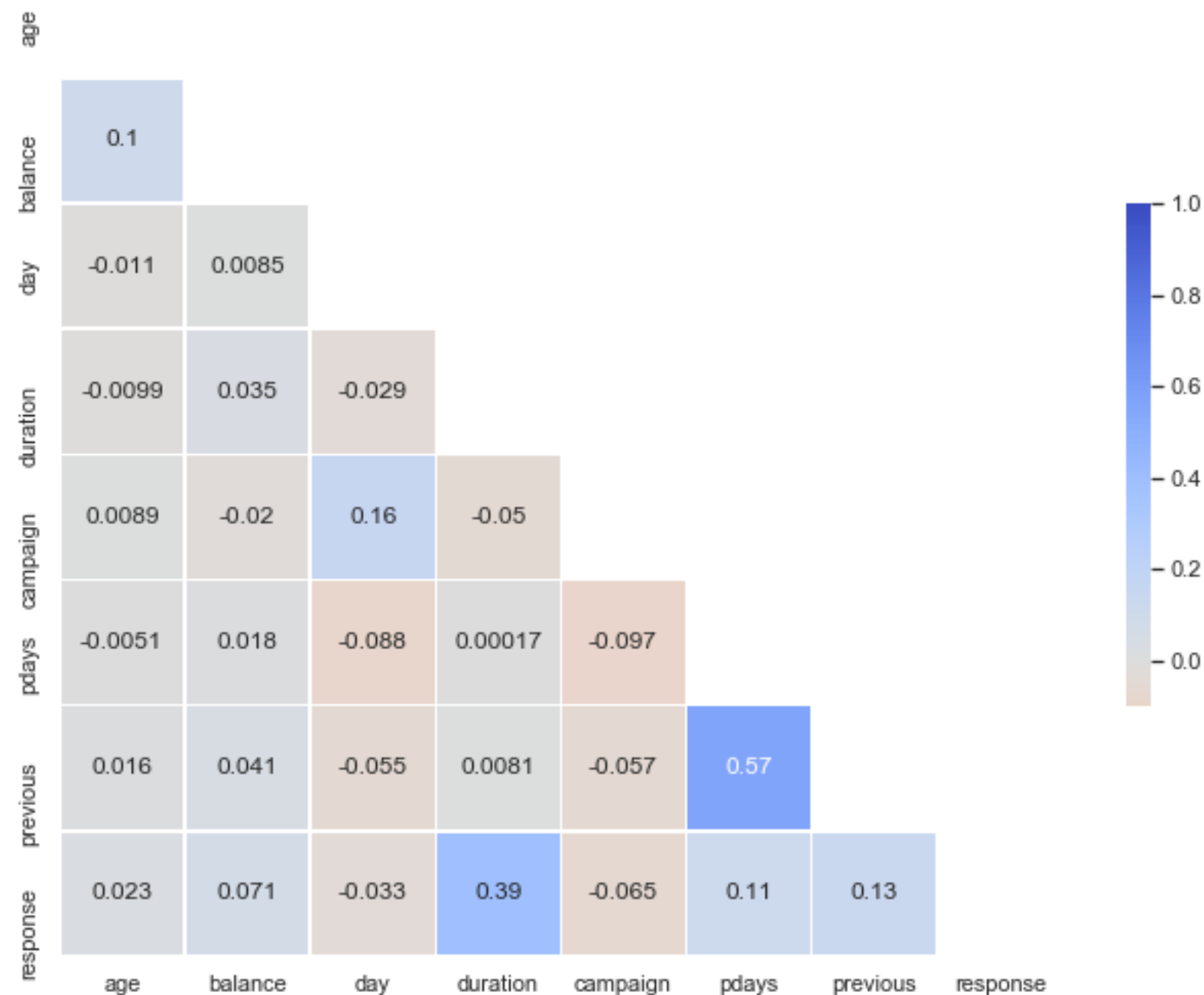
CUSTOMERS  
DEPOSITED AT  
THE BANK.

12.6%

CAMPAIGN  
SUCCESS RATE

# WHAT ARE THE MOST CORRELATED VARIABLES?

## EXPLORATORY DATA ANALYSIS

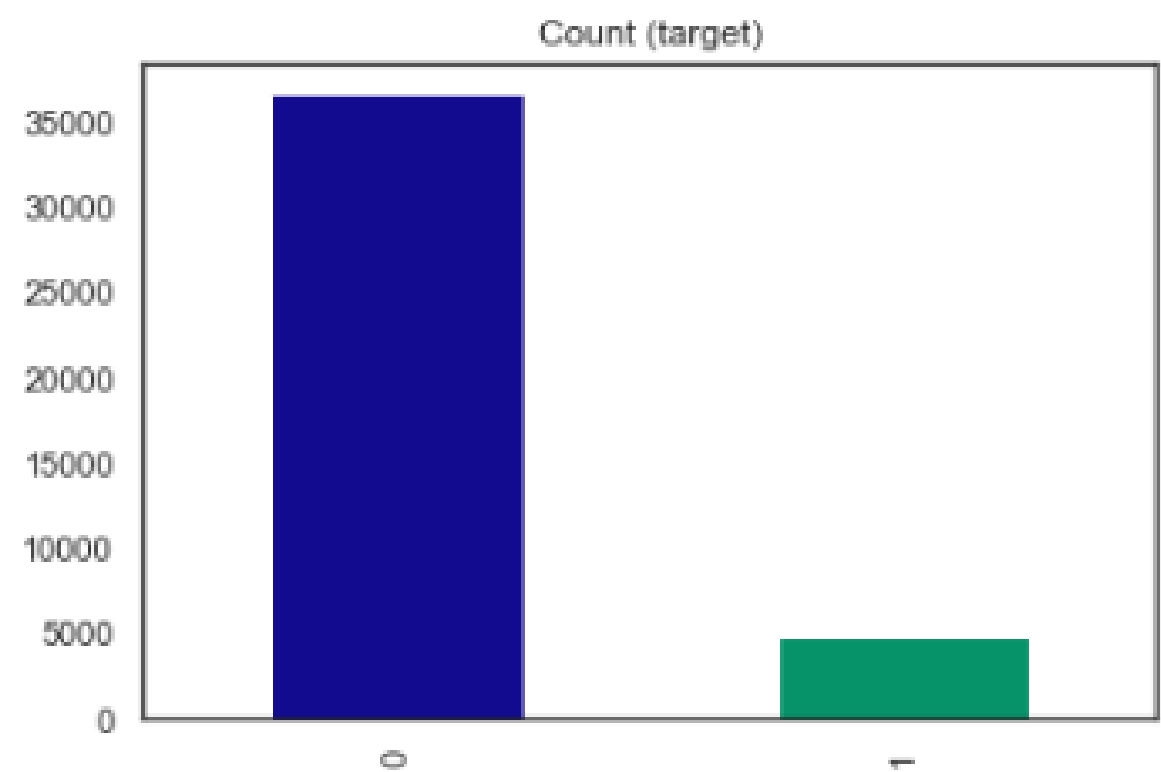


- Clearly the variable '**duration**' has the strongest correlation with the target variable 'response'
- 'campaign ' has a strong correlation with 'duration";
- 'previous' and 'pdays' also show correlation.

# MODEL BUILDING

## UNBALANCED DATA

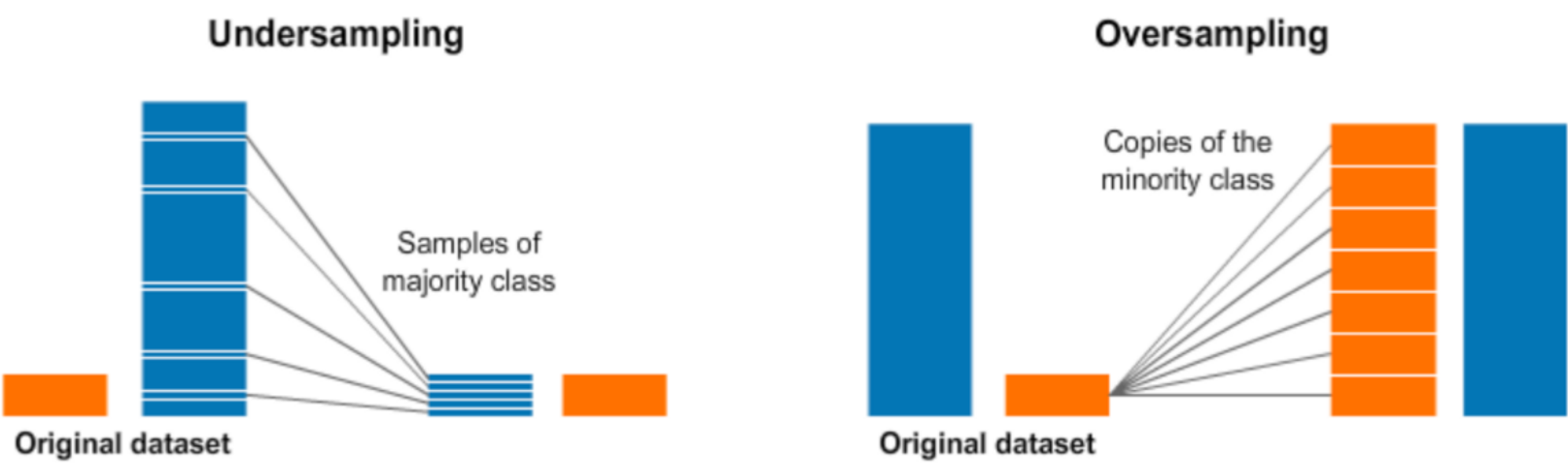
Class 0: 36661  
Class 1: 4868  
Proportion: 7.53 : 1



	precision	recall	f1-score	support
0	0.90	0.98	0.94	11215
1	0.57	0.18	0.27	1453
accuracy			0.89	12668
macro avg	0.74	0.58	0.61	12668
weighted avg	0.86	0.89	0.86	12668

## LOGISTIC REGRESSION MODEL

- `sklearn.linear_model.LogisticRegression`
- Fits a sigmoid function to a data
- Outputs probability which is in  $[0,1]$  range unlike linear models.
- Prepare, Train and Predict

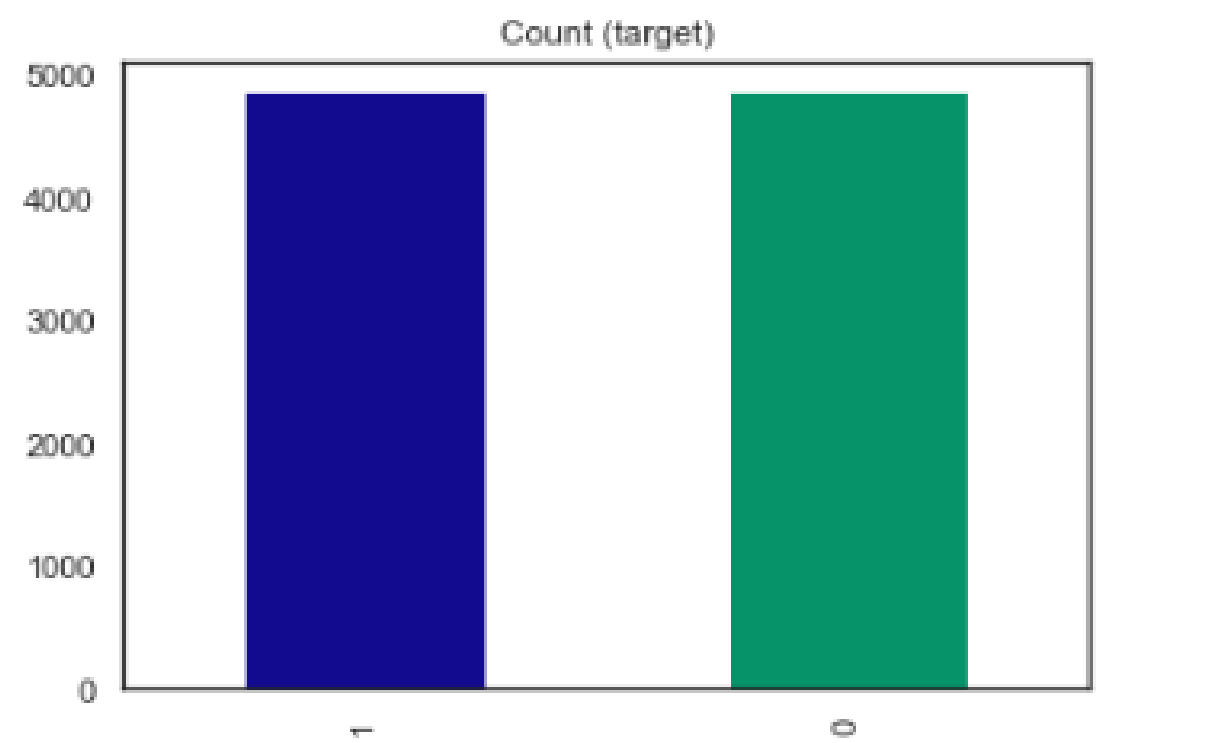


# LOGISTIC REGRESSION MODEL

UNBALANCED DATA

## UNDERSAMPLING

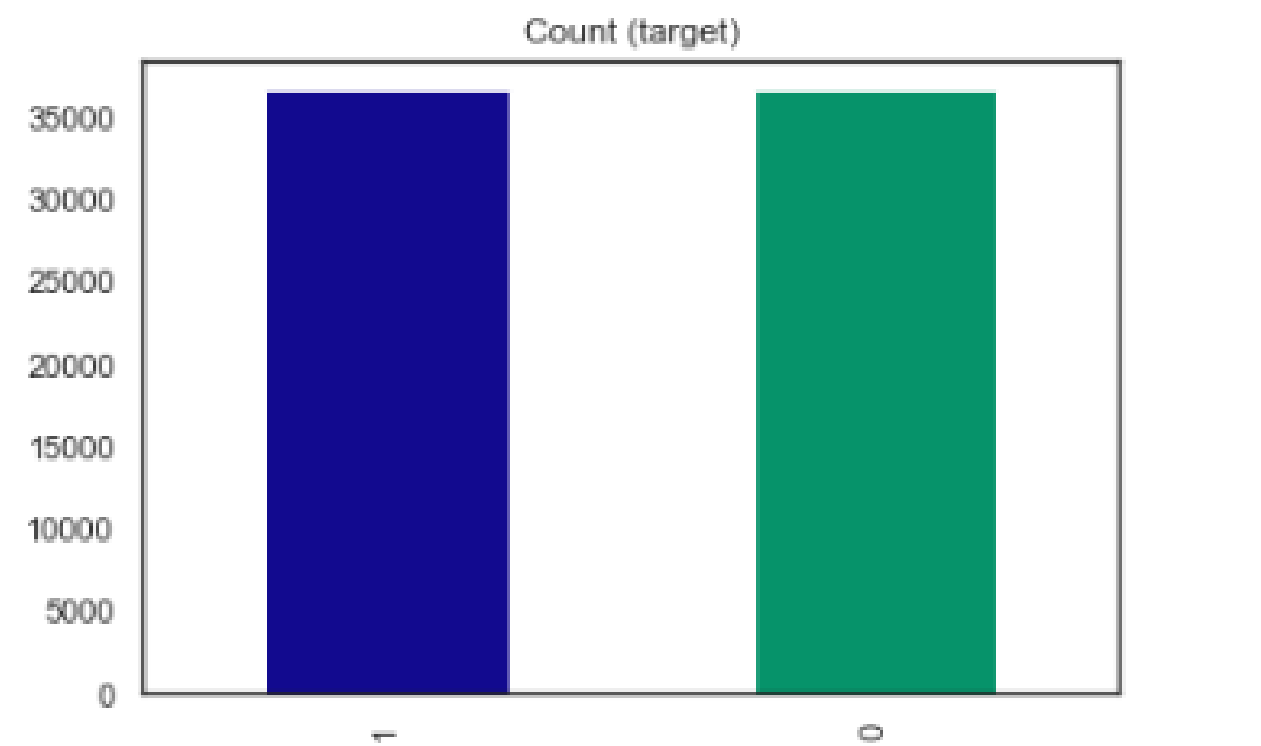
```
Random under-sampling:  
1    4868  
0    4868  
Name: response, dtype: int64
```



	precision	recall	f1-score	support
0	0.79	0.79	0.79	1448
1	0.80	0.79	0.79	1473
accuracy			0.79	2921
macro avg	0.79	0.79	0.79	2921
weighted avg	0.79	0.79	0.79	2921

## OVERSAMPLING

```
Random over-sampling:  
1    36661  
0    36661  
Name: response, dtype: int64
```



	precision	recall	f1-score	support
0	0.79	0.81	0.80	10891
1	0.81	0.79	0.80	11106
accuracy			0.80	21997
macro avg	0.80	0.80	0.80	21997
weighted avg	0.80	0.80	0.80	21997



# Concluding remarks

## RECOMENDATIONS TO THE MARKETING TEAM

- The best customer profile to contact is married people, with age between 40 and 60 years old, and terciary education level.
- Try to engage customers and make calls for a maximum of 9 minutes.
- This analysis sugests that the bank should resist calling a client for more than five times, which can be disturbing and increase dissatisfaction.

## EFFICIENCY IMPROVMENT

By selecting only the most likely buyers, the proposed analysis creates value for the bank telemarketing managers in terms of campaign efficiency improvement (e.g., reducing client intrusiveness and contact costs).

# FURTHER WORKS



## DIMENSIONALITY REDUCTION

Using another ML techniques

## UNBALANCED DATA

Test more advanced methods like SMOTE  
and AllKNN

## MEDIUM POST

Writing about projects and build a  
portfolio

# References

P [Moro et al., 2011] S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.