




Feature Analysis

What features we must focus to
increase money retention?

Matheus Felipe
mathlippo@gmail.com

Summary

- ◆ Objectives.
 - ◆ Columns.
 - ◆ Descriptive Analysis.
 - ◆ Temporal Analysis.
 - ◆ Benchmark Analysis.
 - ◆ Model.
 - ◆ Most important features.
 - ◆ Ending.
- 

Objectives:

As a data analyst, my objectives for this research are to:

- Understand the profile of clients who make term deposits with the bank.
- Analyze metrics and routine variables that may influence productivity.
- Extract the variables that have the most impact on term deposit acquisition.

Dataset: Bank Marketing

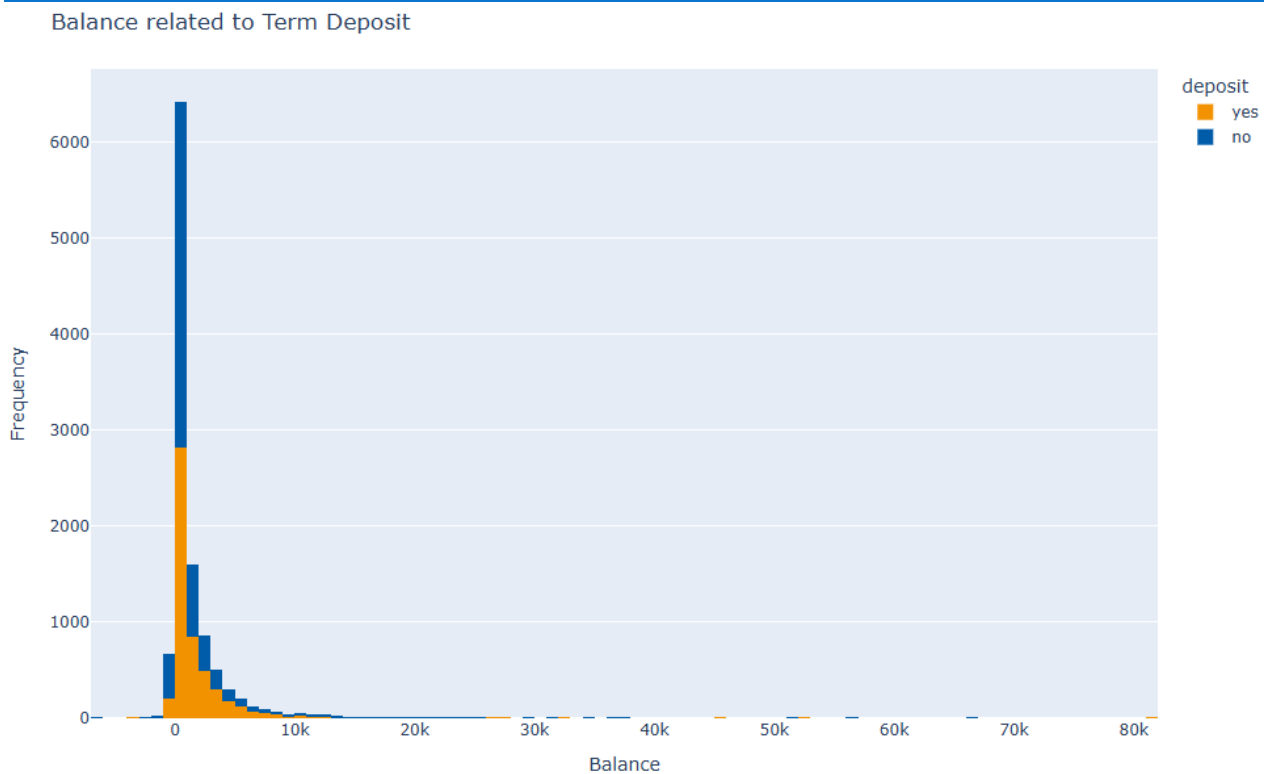
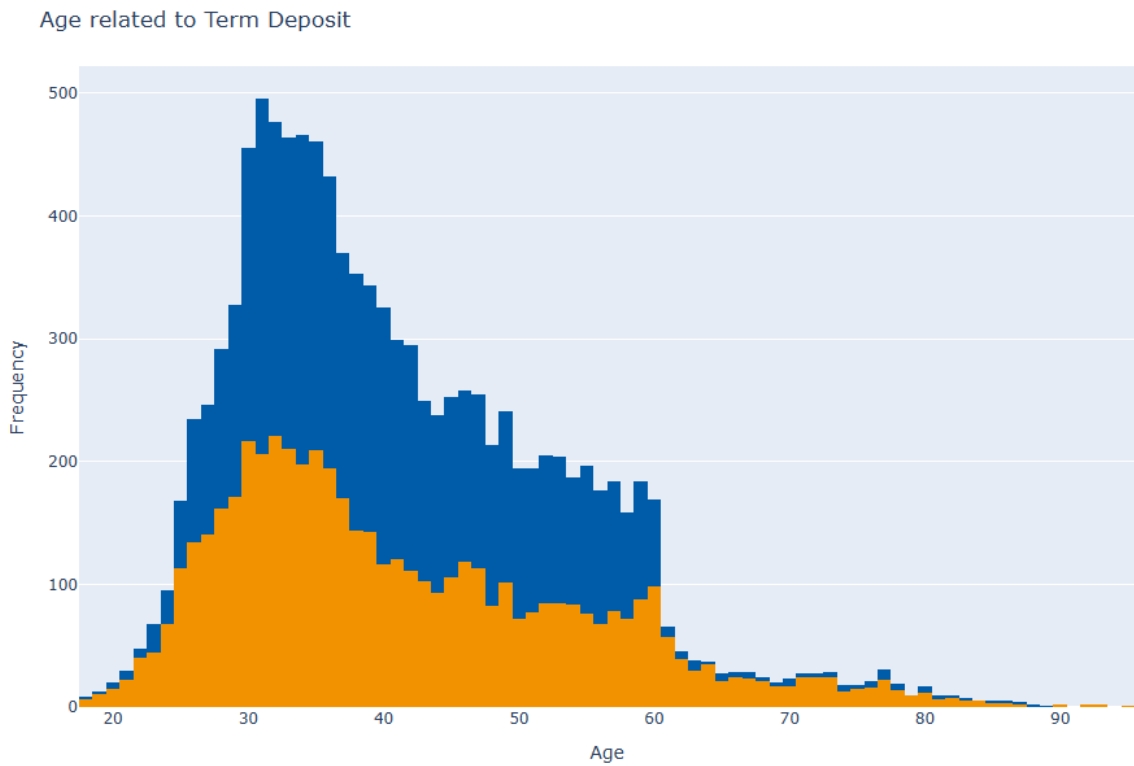
Context: Direct marketing campaigns (phone calls) from a Portuguese banking institution.

Columns:

- age
- job
- marital
- education
- default
- balance
- housing
- loan
- contact
- day
- month
- duration
- campaign
- pdays
- previous
- poutcome
- deposit

- Customer's age
- Type of job
- Marital status
- Education level
- Has credit in default?
- Average yearly balance (in euros)
- Has a housing loan?
- Has a personal loan?
- Contact communication type
- Last contact day of the week
- Last contact month of the year
- Last contact duration, in seconds (integer)
- Number of contacts performed during this campaign for this client
- Days passed since last contact from a previous campaign
- Number of contacts before this campaign for this client (integer)
- Outcome of the previous marketing campaign
- Has the client subscribed to a term deposit?

Descriptive Analysis: numerical variables



Descriptive variables:

	count	mean	std	min	25%	50%	75%	max
age	11162.0	41.231948	11.913369	18.0	32.0	39.0	49.0	95.0
balance	11162.0	1528.538524	3225.413326	-6847.0	122.0	550.0	1708.0	81204.0

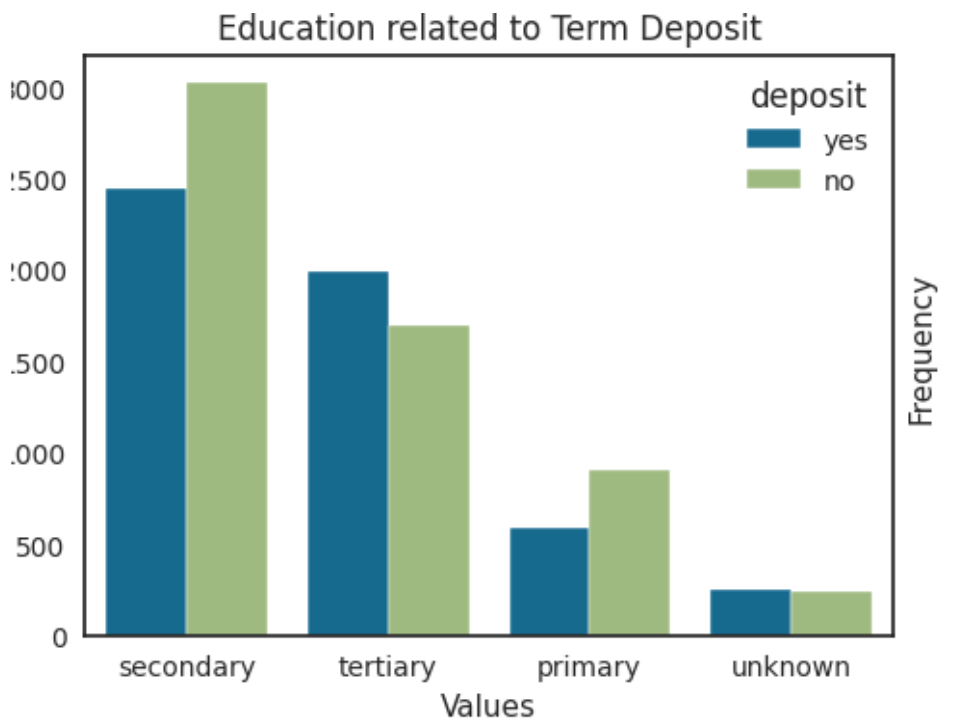
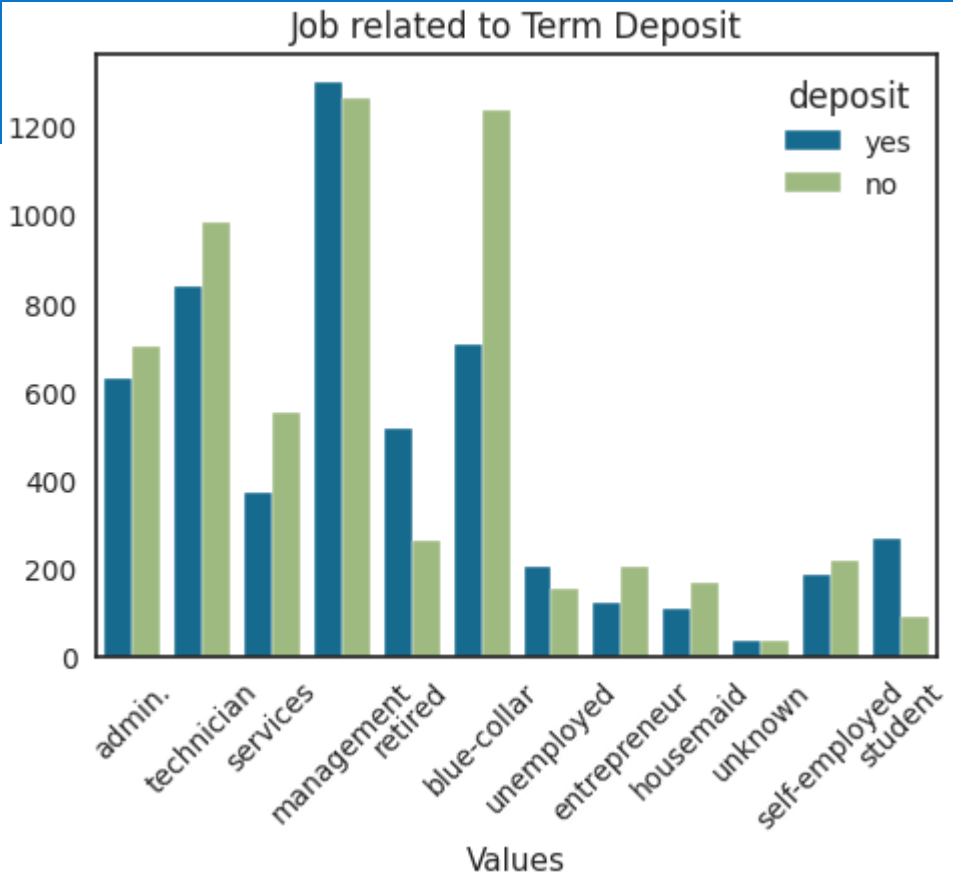
Correlation:

	age	balance
age	1.0000	0.1123
balance	0.1123	1.0000

There's no apparent correlation between age and balance.

People under 30 or over 60 have a higher chance of making a deposit.
For those aged 30 to 60—which represents the majority of the bank's clients—the chances of not making a deposit are slightly higher.
For people with over 1000 euros in their account, the chances of a deposit progressively increase. This is expected, as the median balance is 550 euros, suggesting that most people will not deposit part of their capital into a term deposit.

Descriptive analysis: binary and categorical variables (1/2)



Key Findings:

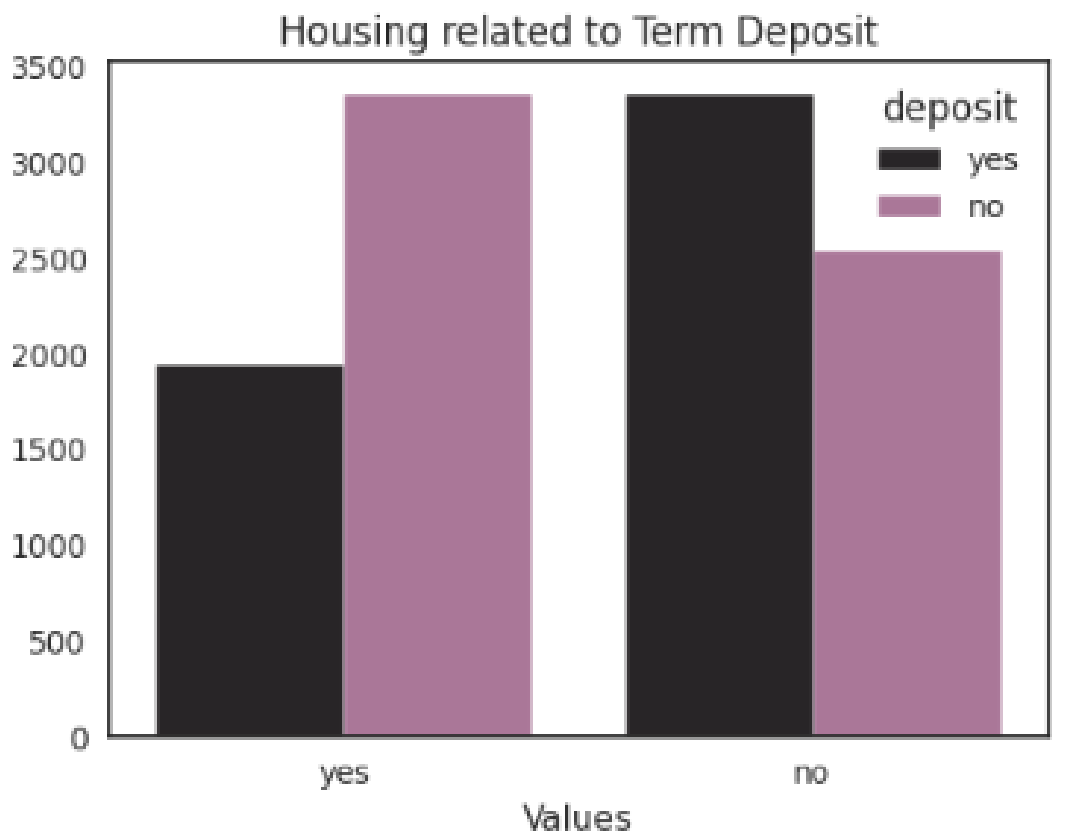
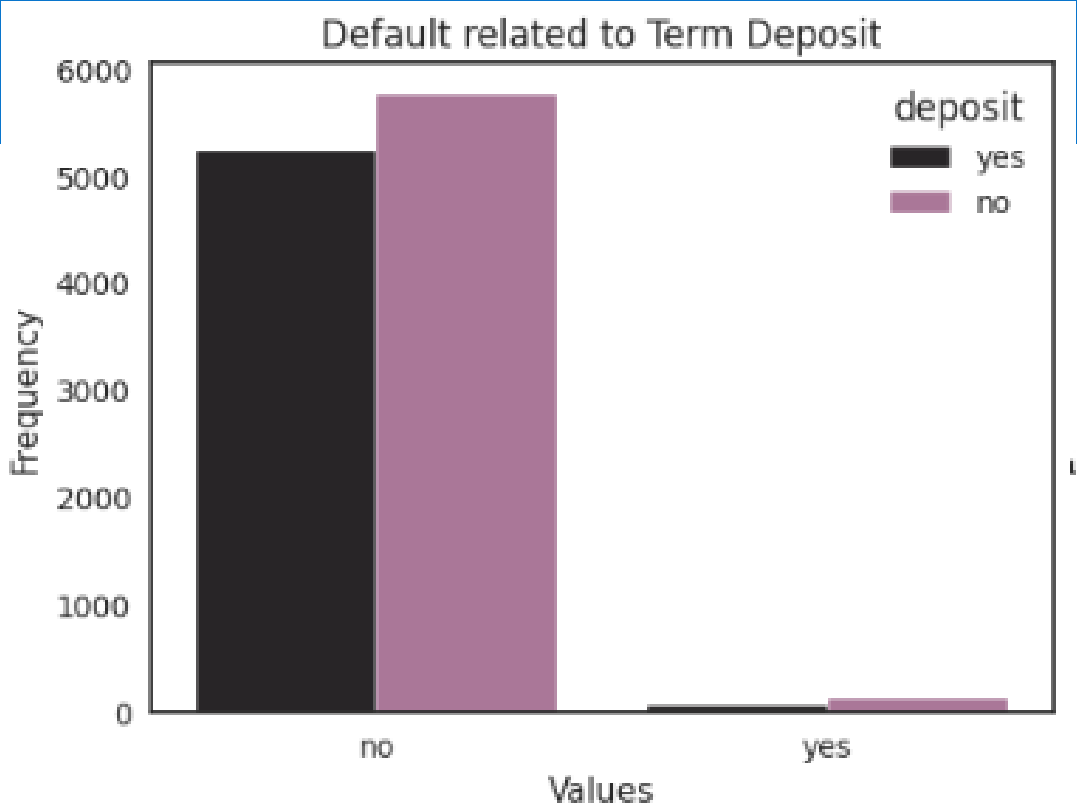
There is a considerable impact from a client's profession, education level, and whether they have a housing loan on their decision to make a deposit. Regarding default rates, there is not a sufficient sample of defaulting clients to make a definitive assessment.

Impact Variables:

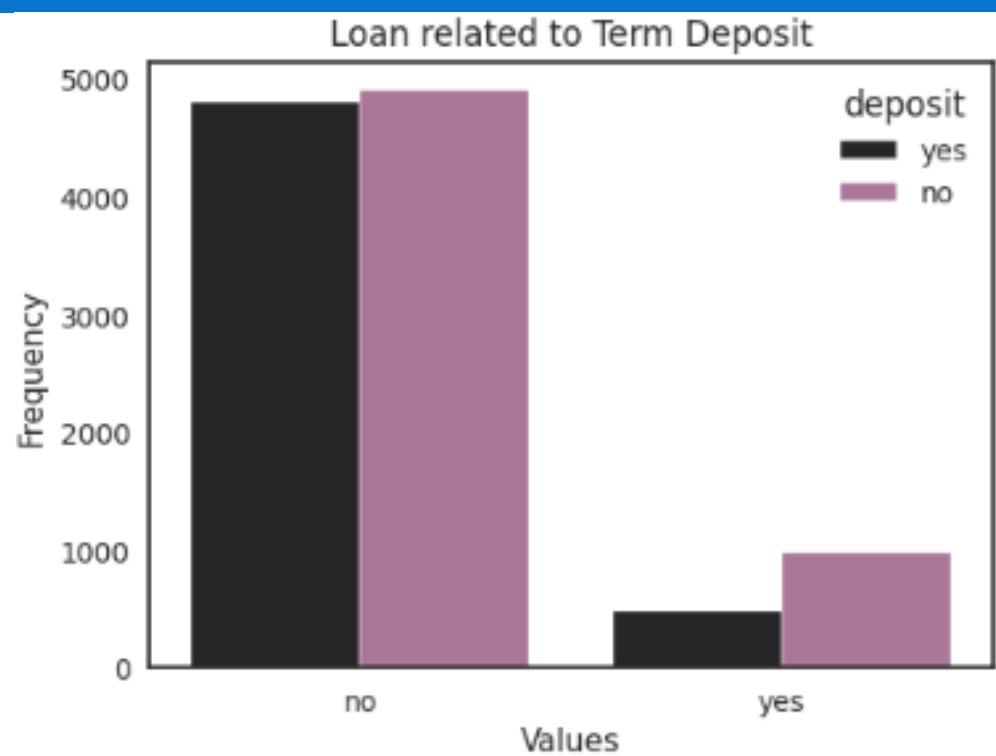
Job: A high deposit rate is seen among students and retirees, while a low rate is observed among blue-collar workers.

Education: There is little difference in frequency, with college-educated individuals making the most deposits.

Housing Loan: Clients without a housing loan are more likely to make a capital deposit.



Análise Descritiva: binary and categorical variables (2/2)



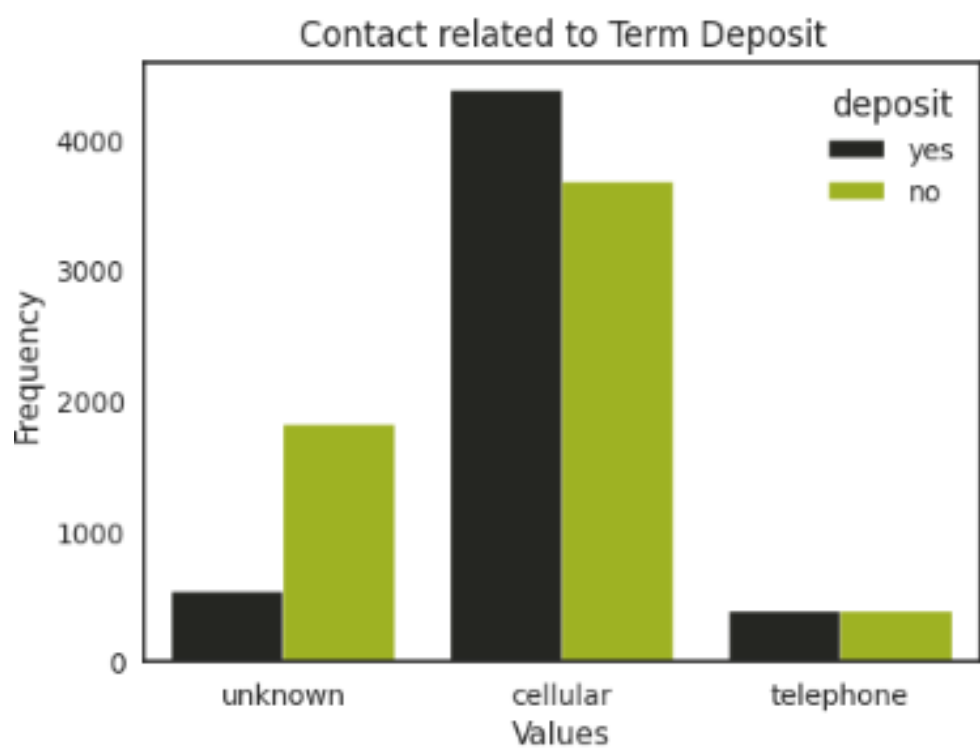
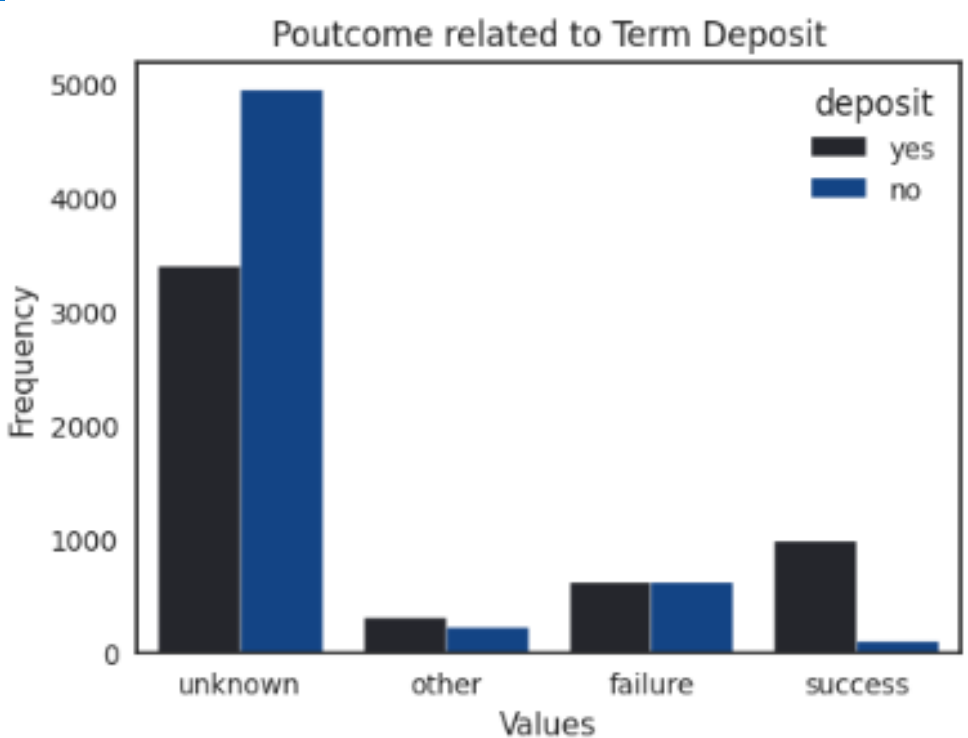
Analysis:

There is a low success rate for clients who already have a loan.

The primary contact method between employees and clients is via mobile phone.

A significant portion of campaign results are not documented.

The data between depositors and non-depositors is balanced, which allows for a solid analysis of the key variables.



deposit count

0	no	5873
1	yes	5289

Temporal analysis:

The proportions of clients who do not make a deposit relative to the daily or monthly total are random, with some months seeing higher deposit acquisition than others.

Months and their proportion (divided by 100) of clients who DID NOT make a deposit:

month	deposit	proportion
dec	no	0.090909
mar	no	0.101449
sep	no	0.156740
oct	no	0.176020
apr	no	0.374865
feb	no	0.431701
aug	no	0.547070
jun	no	0.553191
nov	no	0.572641
jul	no	0.585865
jan	no	0.587209
may	no	0.672450

Days with the lowest chances of a client refusing a term deposit offer, based on their proportion (divided by 100):

		proportion
day	deposit	
10	no	0.257669
1	no	0.262295
25	no	0.406250
3	no	0.418301
22	no	0.427509
4	no	0.427861
30	no	0.433054
12	no	0.451685
2	no	0.455090
13	no	0.467991

Benchmark Analysis:

•**Routine variables** measure the time of contact (or lack thereof) with the client and the number of offers made to them.

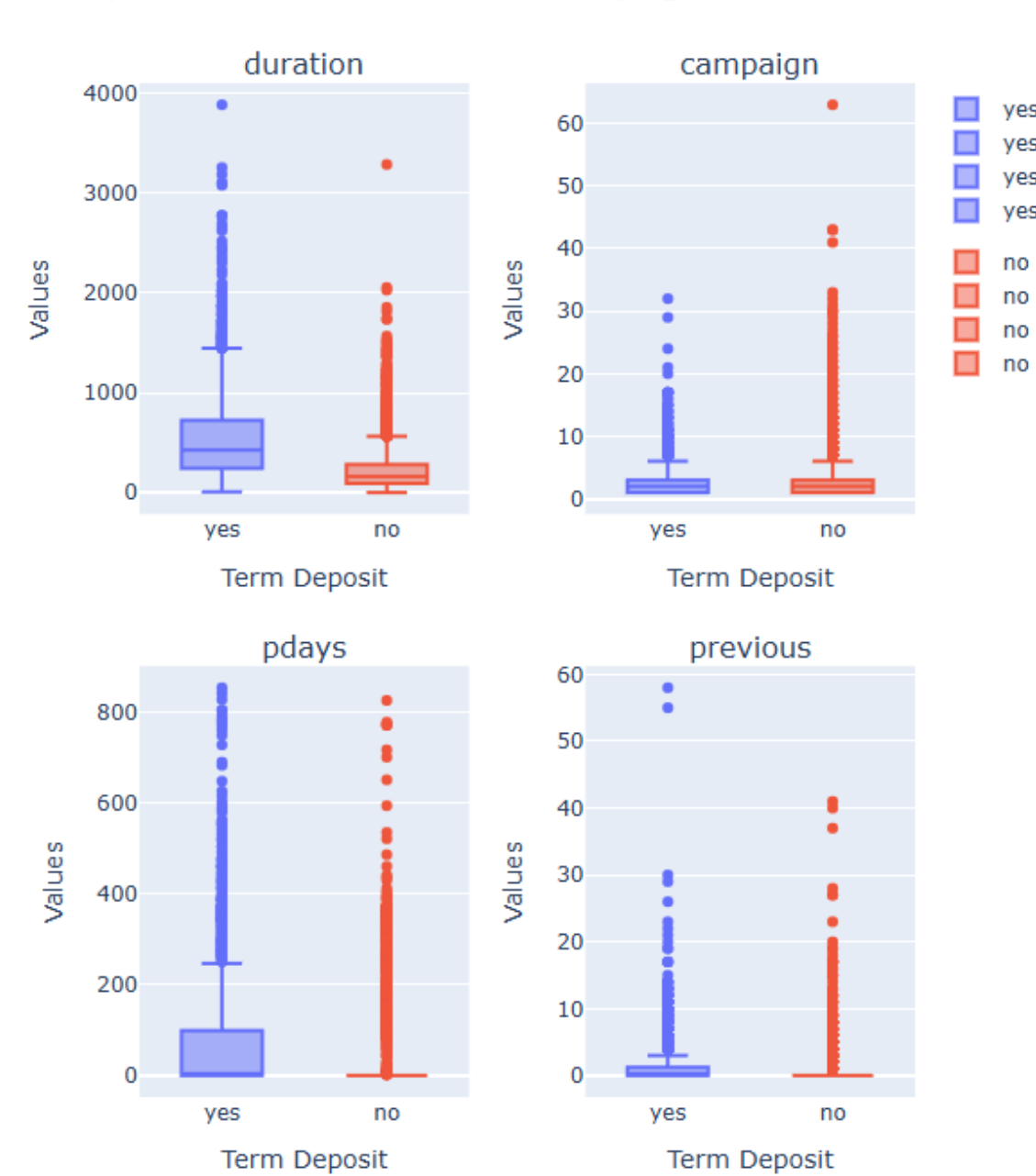
•Most clients decide to sign a term deposit contract during their first contact with the salesperson, represented by -1 on the boxplot.

•*Note: The boxplots for 'previous days' and 'previous campaign' are only represented by the number 1.*

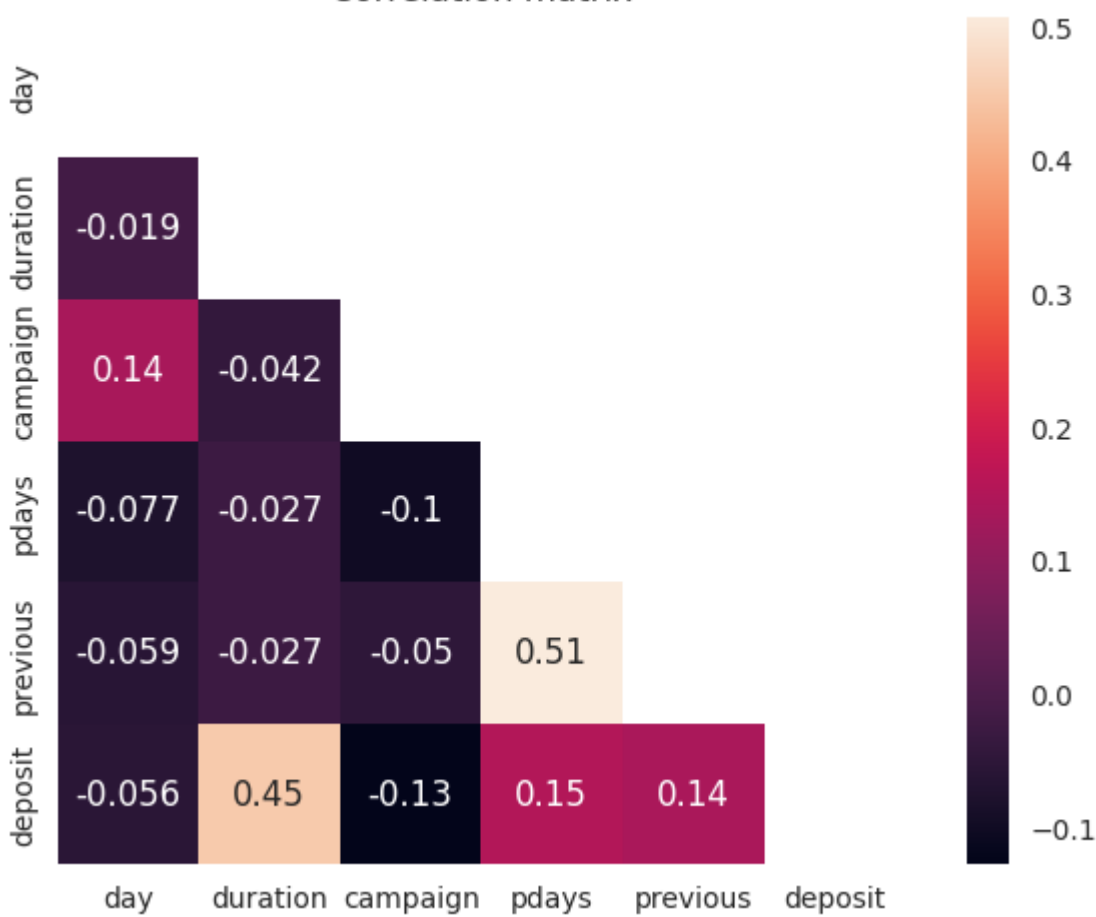
These variables show an equal balance between deposit rates, with the exception of 'call duration.' Naturally, calls resulting in a deposit will last longer, as they involve discussing product details and other service offerings.

.

Boxplots for variables related to campaign and customer service



Correlation matrix



Overall, there is almost no relation between chances of deposit, only the duration of the call being relevant

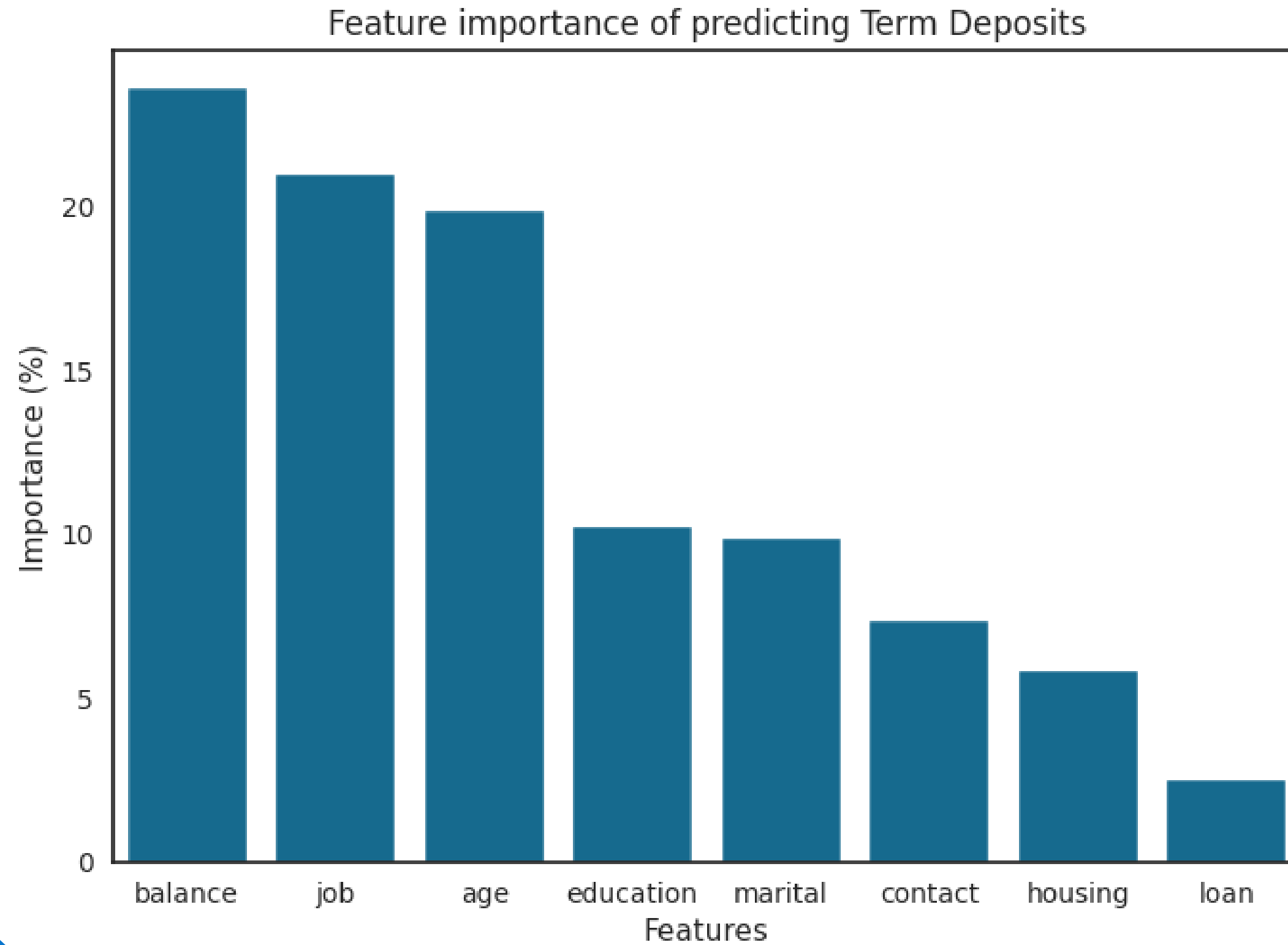
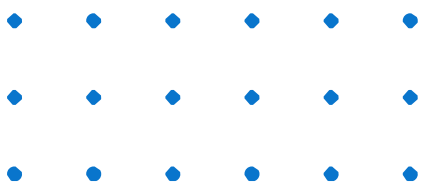
A 3x6 grid of blue diamond shapes. There are 3 rows and 6 columns of these shapes, totaling 18 shapes.

1. Many categorical variables, reducing the need of manually encoding, thus avoiding target leakage or arbitrary attribution;
2. Outliers in the balance column;
3. It's easy to get the coefficients.

Model Features:

- **Independent Variables:** age, job, marital, education, balance, housing loan, personal loan, and contact type.
- **Dependent Variable:** deposit.
- Other variables were removed because they were imbalanced (e.g., defaulted), showed a low correlation with the 'deposit' variable, or contained biases that could cause model overfitting

Most importante features:



Most important:

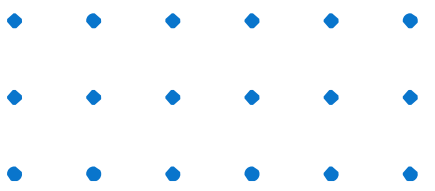
– BALANCE

– JOB

– AGE

Ending

(according to mentioned objectives):



- The broadest client profiles the bank has are young or elderly individuals with a good amount of capital saved and no housing loan.
- A good portion of contracts are either signed or not signed during the first contact with the client.
- By knowing a client's age, job, and financial balance, we can get a good estimate of their likelihood of investing through the bank.

THANK YOU!

Matheus Felipe, 09/04/2025