

The background features a complex network of thin grey lines and dots, forming a web-like structure. Scattered throughout are various triangles of different sizes and orientations, some with solid grey dots at their vertices. The overall aesthetic is modern and technical.

Análise T10 - Client Report

EDA

Exploração Básica de Dados

01

Análise Univariada

Análise com Phi K

02

Análise Detalhada das Principais Variáveis

Entendendo quem se inscreve

03

Conteúdo de Hoje

04

Análise Multivariada e Modelos

Predição de clientes!

05

Conclusões Finais

01

EDA

Exploração Inicial





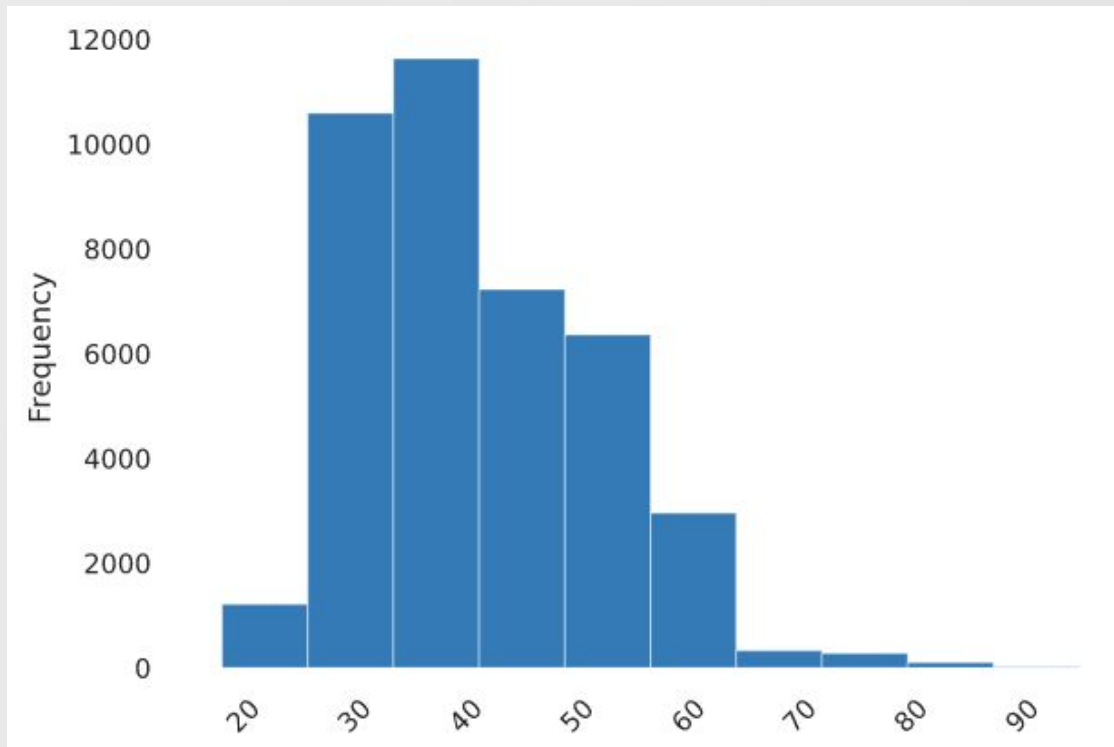
EDA

Nesta etapa, o objetivo é apenas analisar as principais características do dataset e tentar responder algumas perguntas gerais sobre o mesmo.

Este EDA básico foi feito utilizando a biblioteca Pandas Profiling



Qual a Faixa Etária das Pessoas no Dataset?



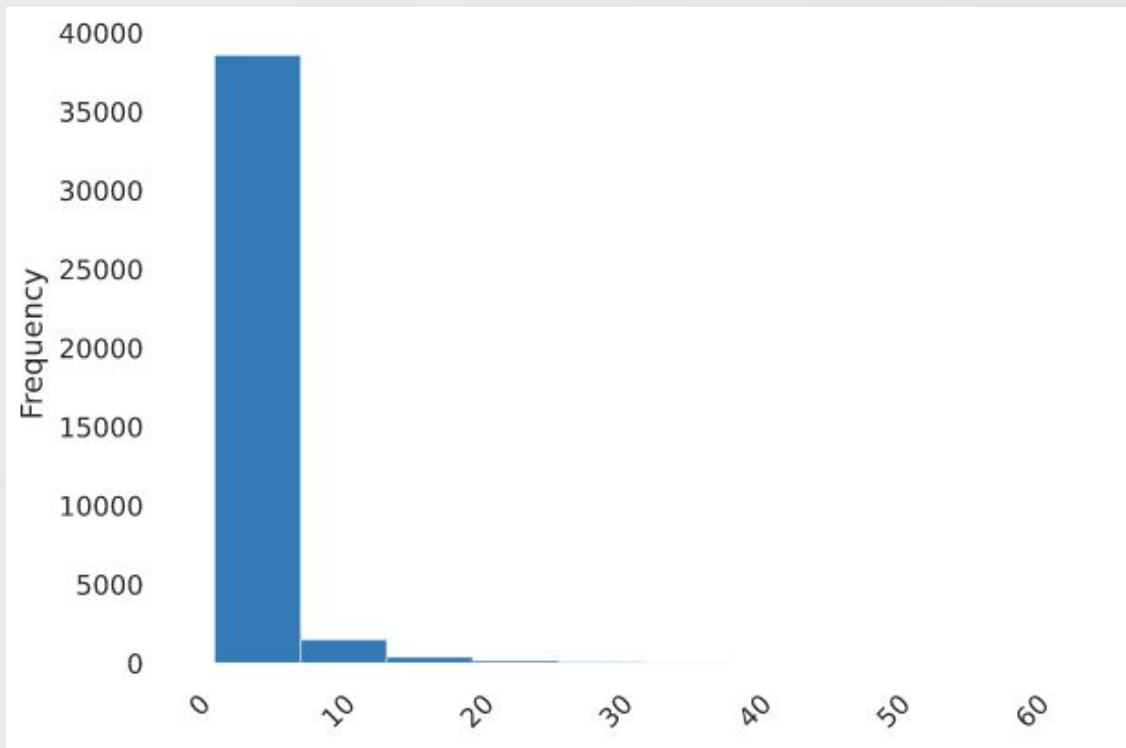
Quais são as profissões mais comuns?

Value	Count	Frequency (%)	
blue-collar	8769	21.6%	
management	8504	20.9%	
technician	6818	16.8%	
admin.	4661	11.5%	
services	3725	9.2%	
retired	2027	5.0%	
self-employed	1427	3.5%	
entrepreneur	1339	3.3%	
unemployed	1193	2.9%	
housemaid	1125	2.8%	
Other values (2)	1102	2.7%	

Nível de escolaridade

Value	Count	Frequency (%)	
secondary	20951	51.5%	
tertiary	11917	29.3%	
primary	6153	15.1%	
unknown	1669	4.1%	



Quantidade de Contatos Feitos Durante a Campanha



Taxa de Sucesso Em Campanhas Anteriores

Value	Count	Frequency (%)	
unknown	33284	81.8%	
failure	4377	10.8%	
other	1648	4.1%	
success	1381	3.4%	

Taxa de Inscritos e Não Inscritos

Value	Count	Frequency (%)	
no	35903	88.2%	
yes	4787	11.8%	



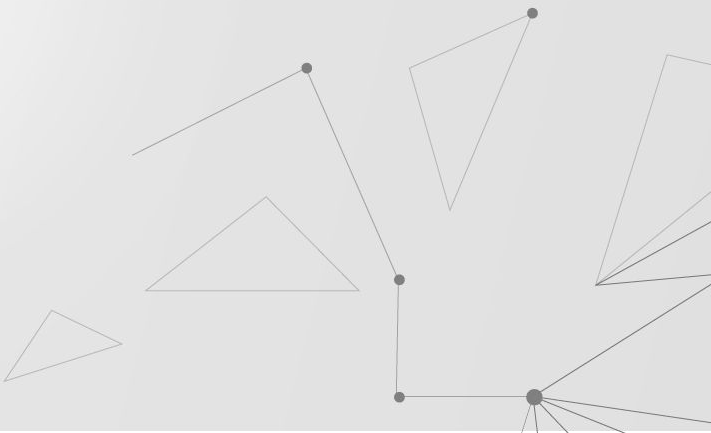
02

Análise Univarida



Análise Univariada

Vamos analisar como as variáveis se relacionam
com o resultado de forma única



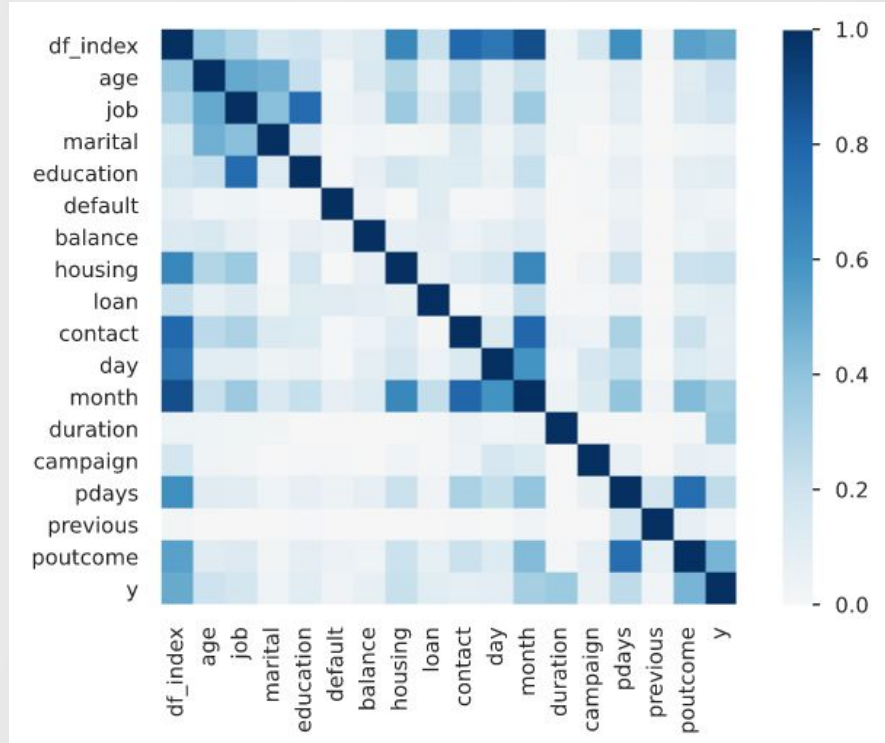
Metodologia

Para realizar esta análise vamos utilizar o coeficiente de **Phi K**.

Este coeficiente consegue **analisar variáveis categóricas e numéricas ao mesmo tempo**.



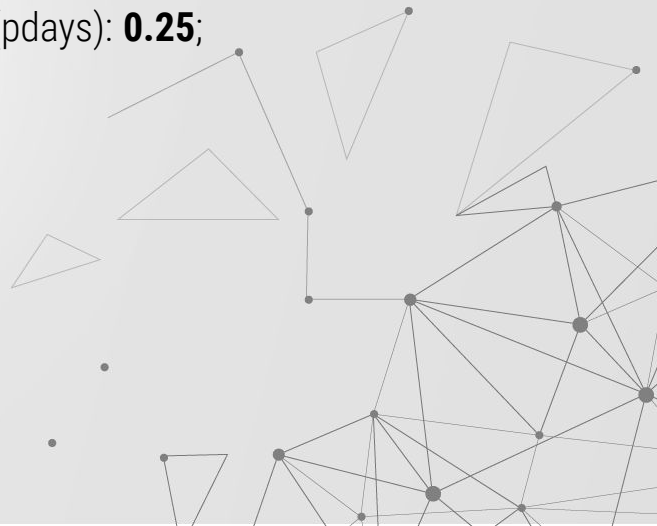
Heatmap Phi K



Variáveis que Mais Se Relacionam com o Resultado (y)

As variáveis que mais se relacionam com o resultado da campanha são:

- Resultado da última campanha de marketing realizada (poutcome): **0.46**;
- Duração do último contato (duration): **0.36**;
- Mês do último contato (month): **0.33**;
- Número de dias que se passaram desde o último contato (pdays): **0.25**;
- Se a casa é financiada (housing): **0.21**;
- A idade (age): **0.20**



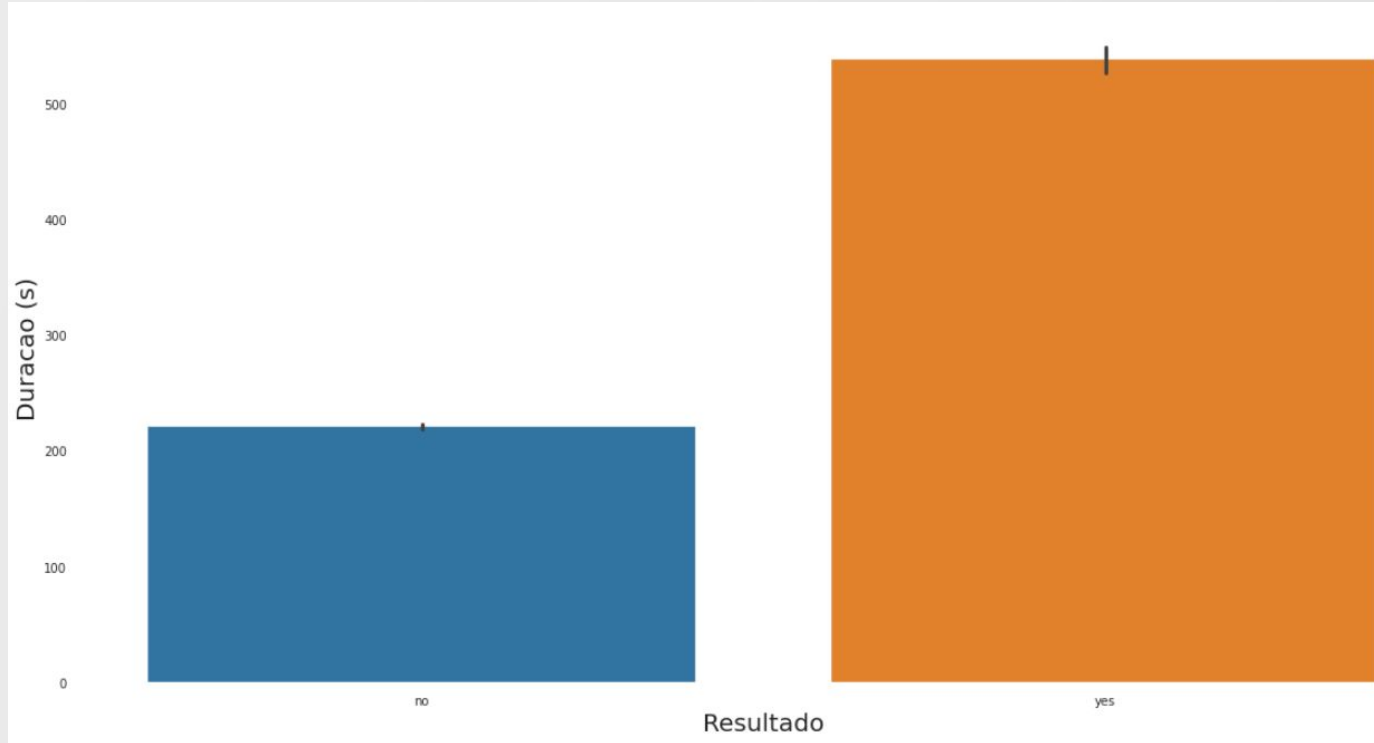


03

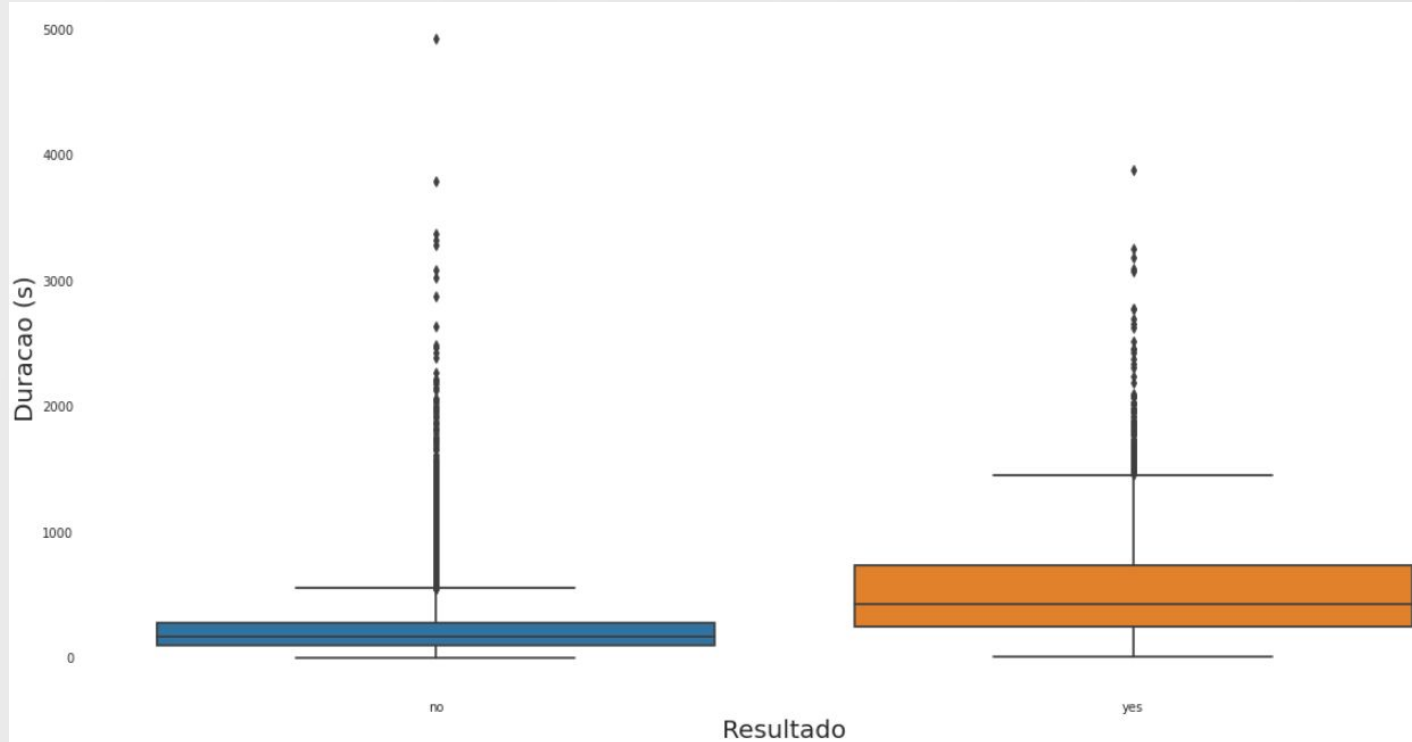
Análise das Principais Variáveis

Entendendo quem se increce

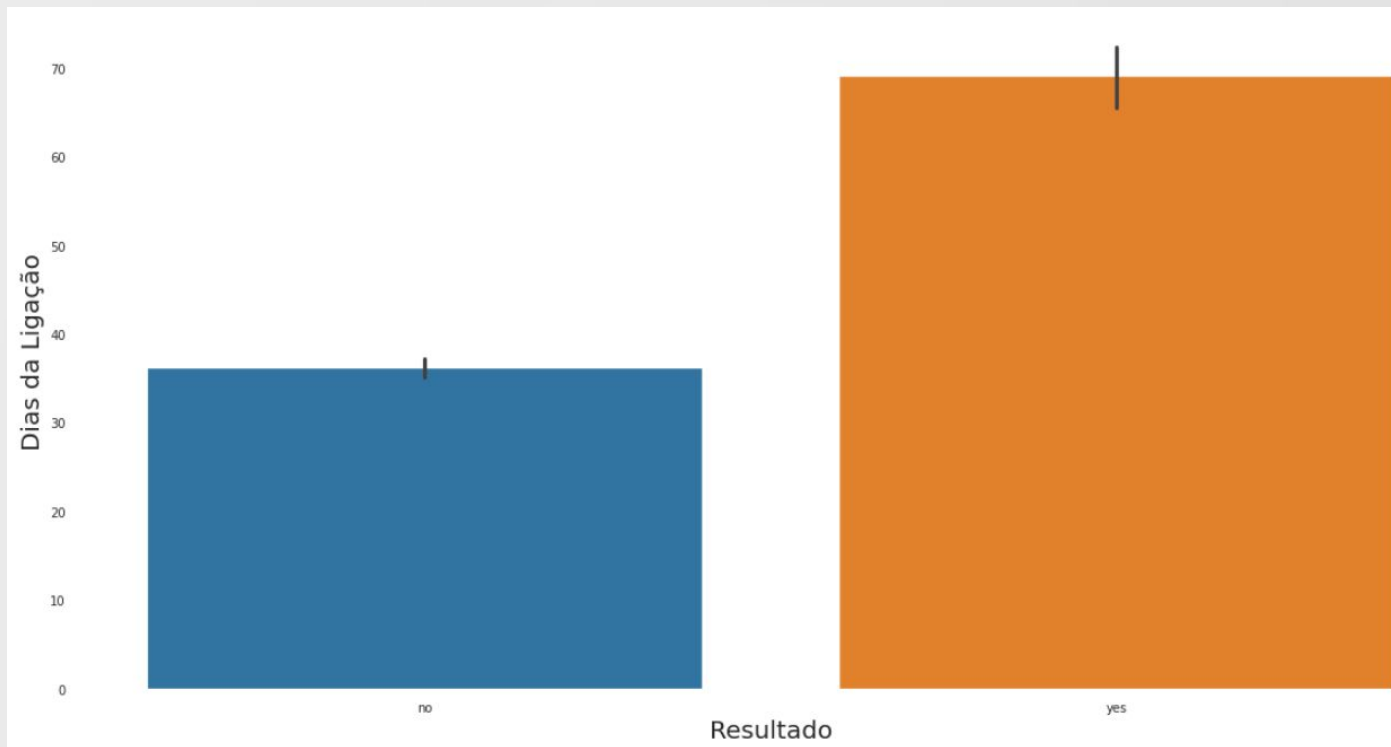
Duração do Último Contato por Resultado



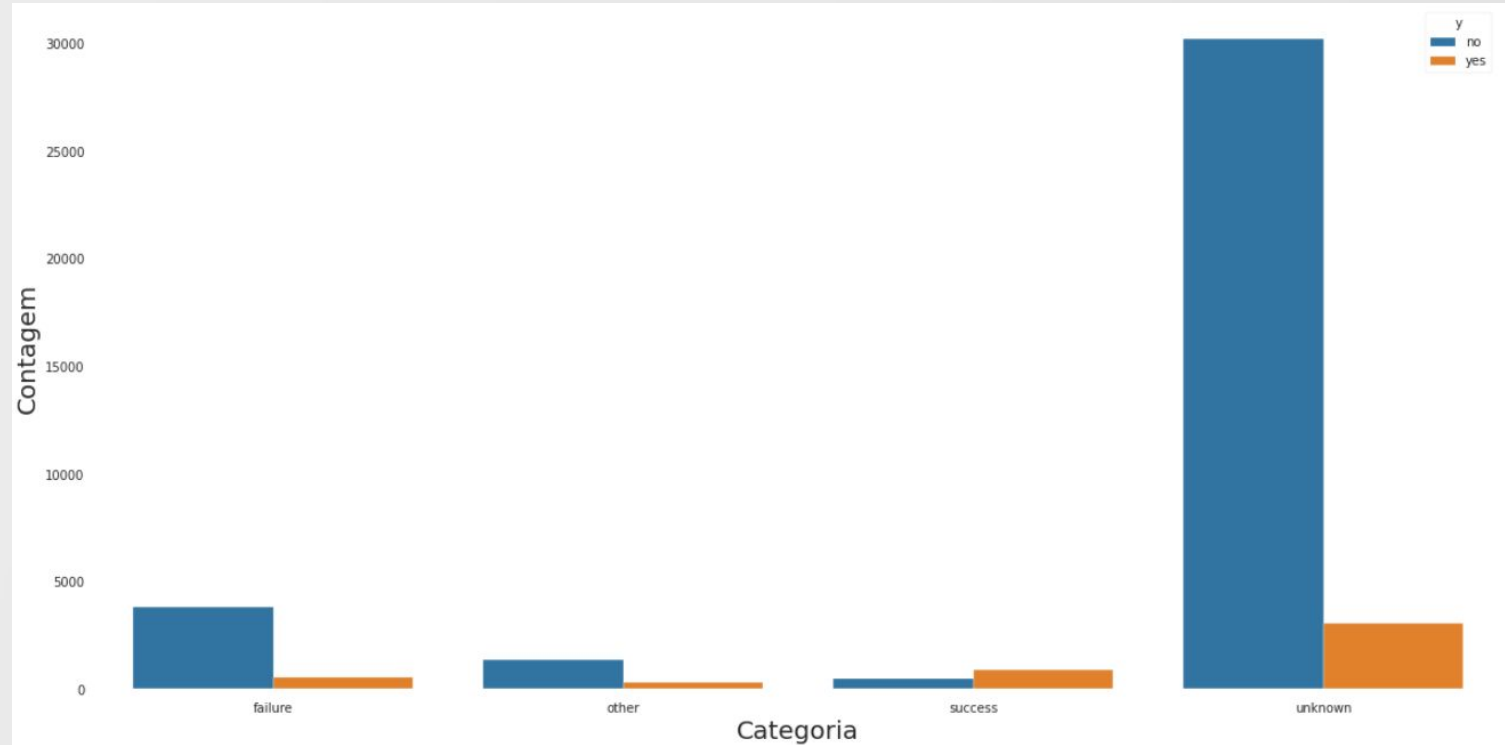
Duração do Último Contato por Resultado



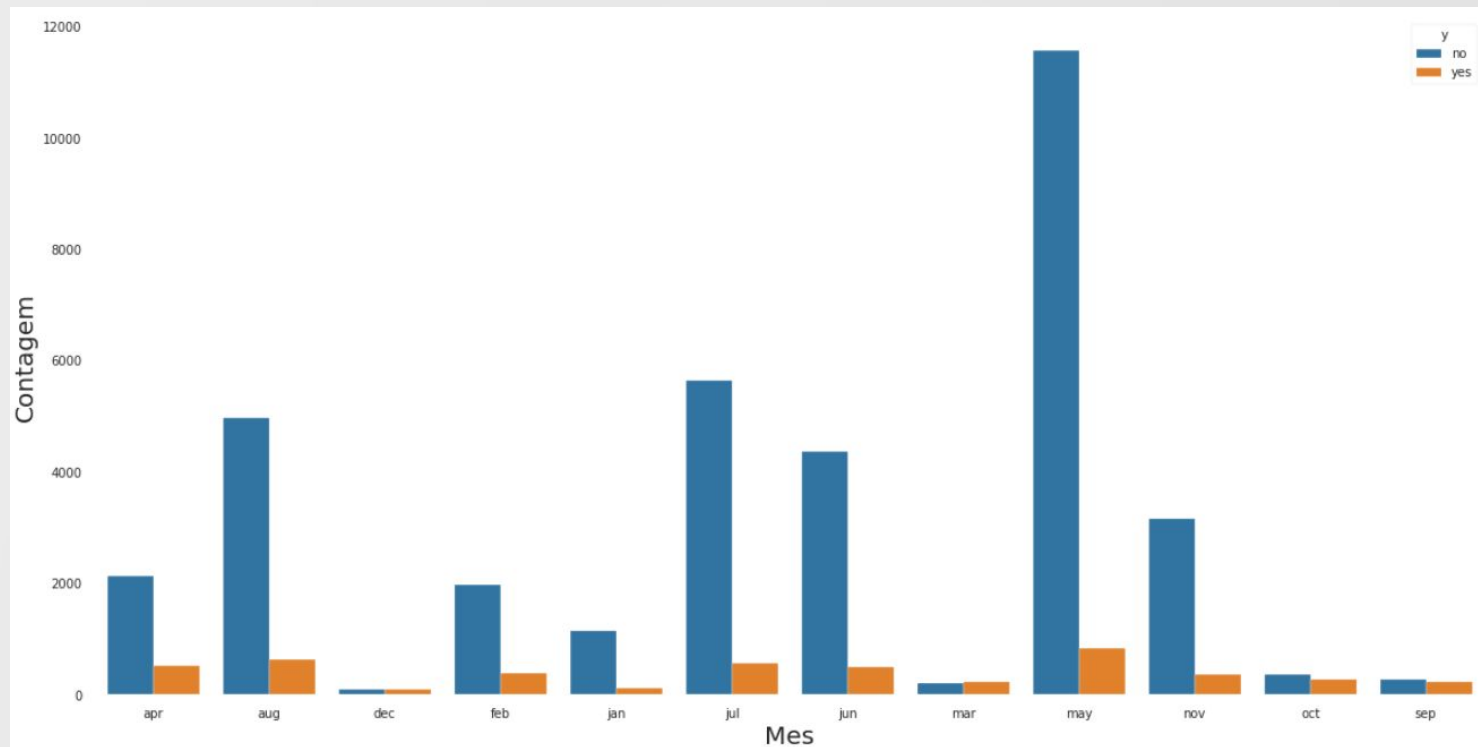
Dias Passados desde o Último Contato



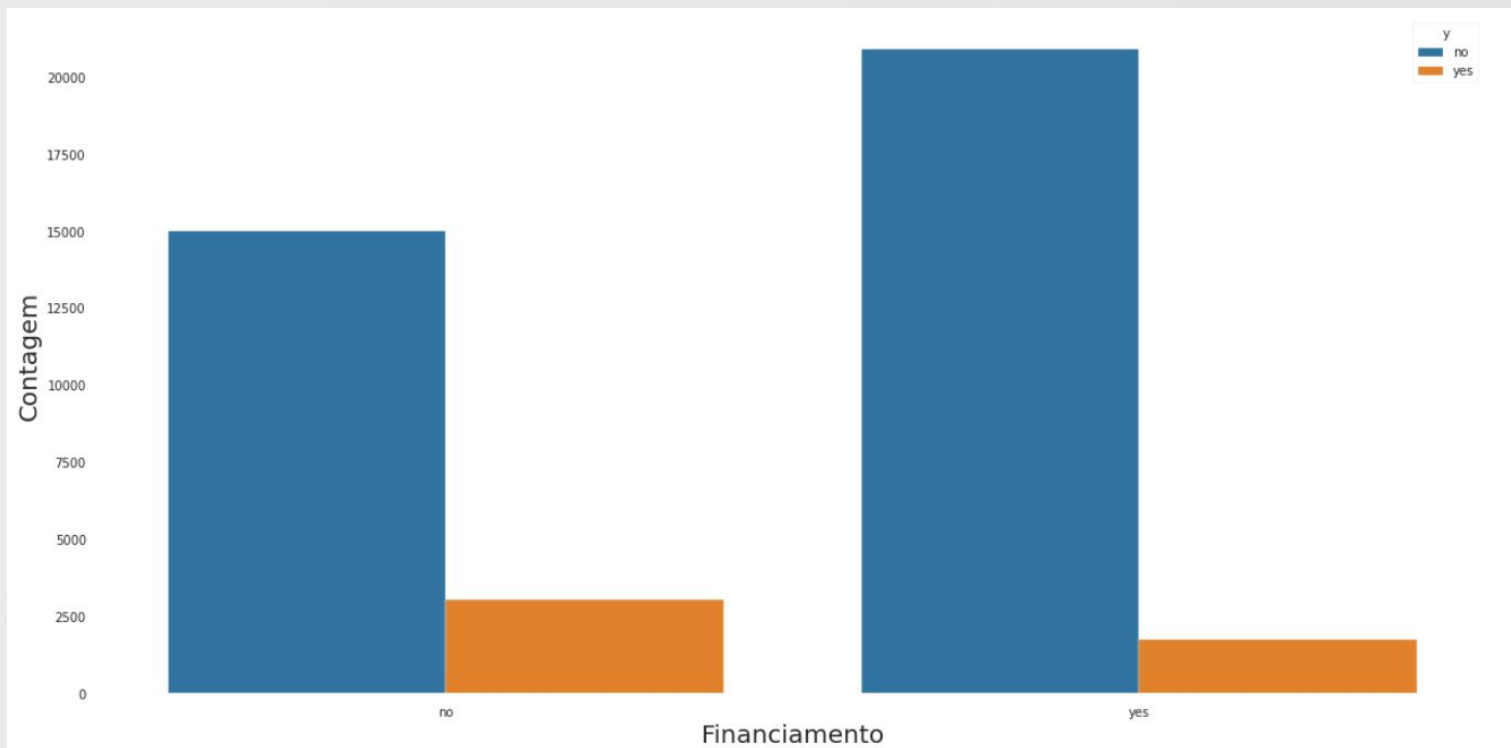
Resultado da Última Campanha



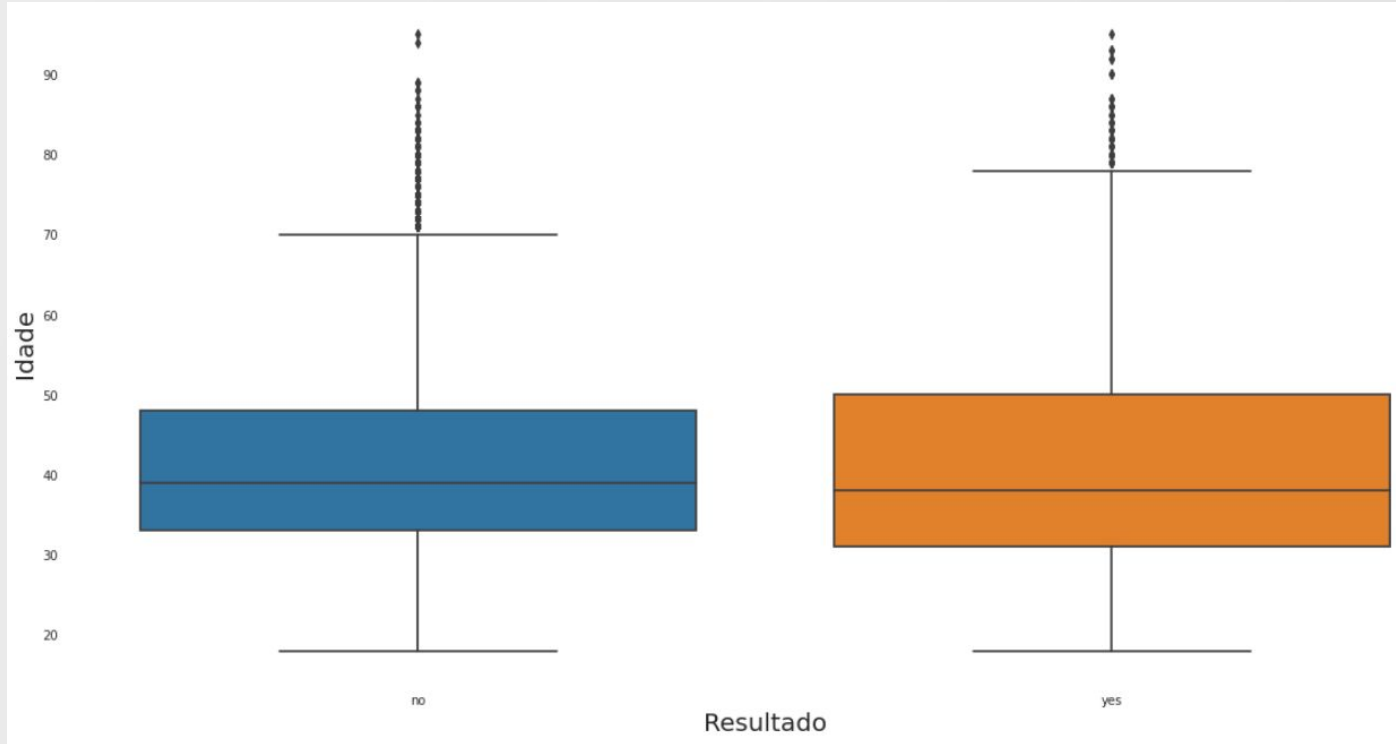
Mês do Último Contato



Financiamento de Casa




Faixa Etária



04

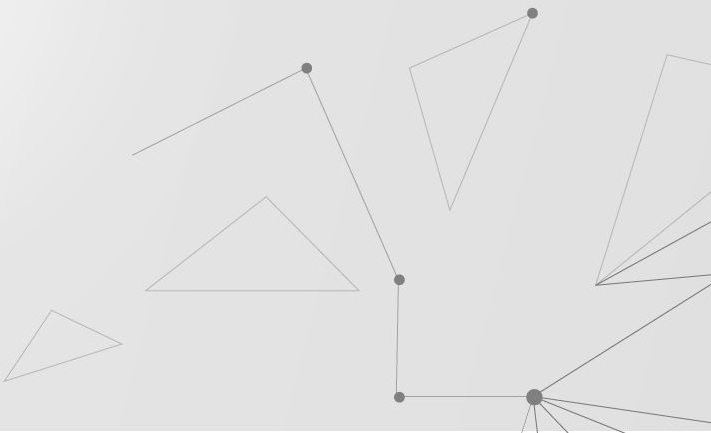
Análise Multivariada e Modelos

Predição de Clientes



Quais variáveis realmente mais impactam na predição?

Identificação das variáveis que mais impactam na
predição de um modelo



Metodologia

Para realizar esta análise foram utilizadas 3 abordagens diferentes, gerando valores diferentes de importância de variáveis.

As 3 abordagens foram feitas utilizando uma validação cruzada de 5 folds

O conjunto de dados foi dividido em treino, validação e teste

As métricas utilizadas são Precisão e Recall;

As abordagens são:

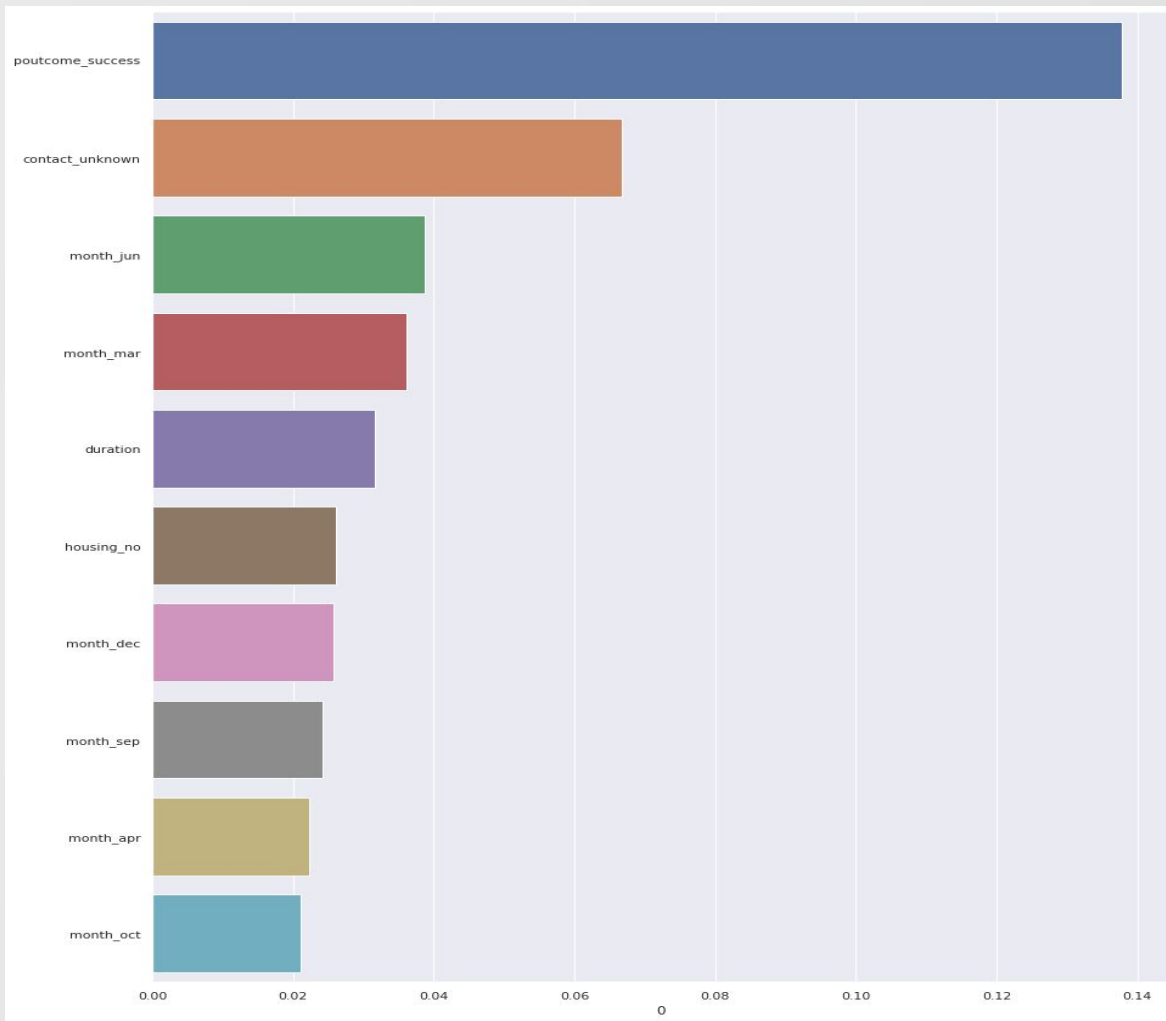
- Baseline model, apenas com padronização das entradas e modelo XGBoostClassifier;
- Oversampling do conjunto de dados minoritário ($y=1$) e modelo XGBoostClassifier;
- Oversampling do conjunto de dados minoritário ($y=1$), seleção de features e modelo XGBoostClassifier.



Baseline Model

Conjunto	Recall	Precisão
Validação	0.53	0.62
Teste	0.47	0.62

Baseline Model - Importância de Features

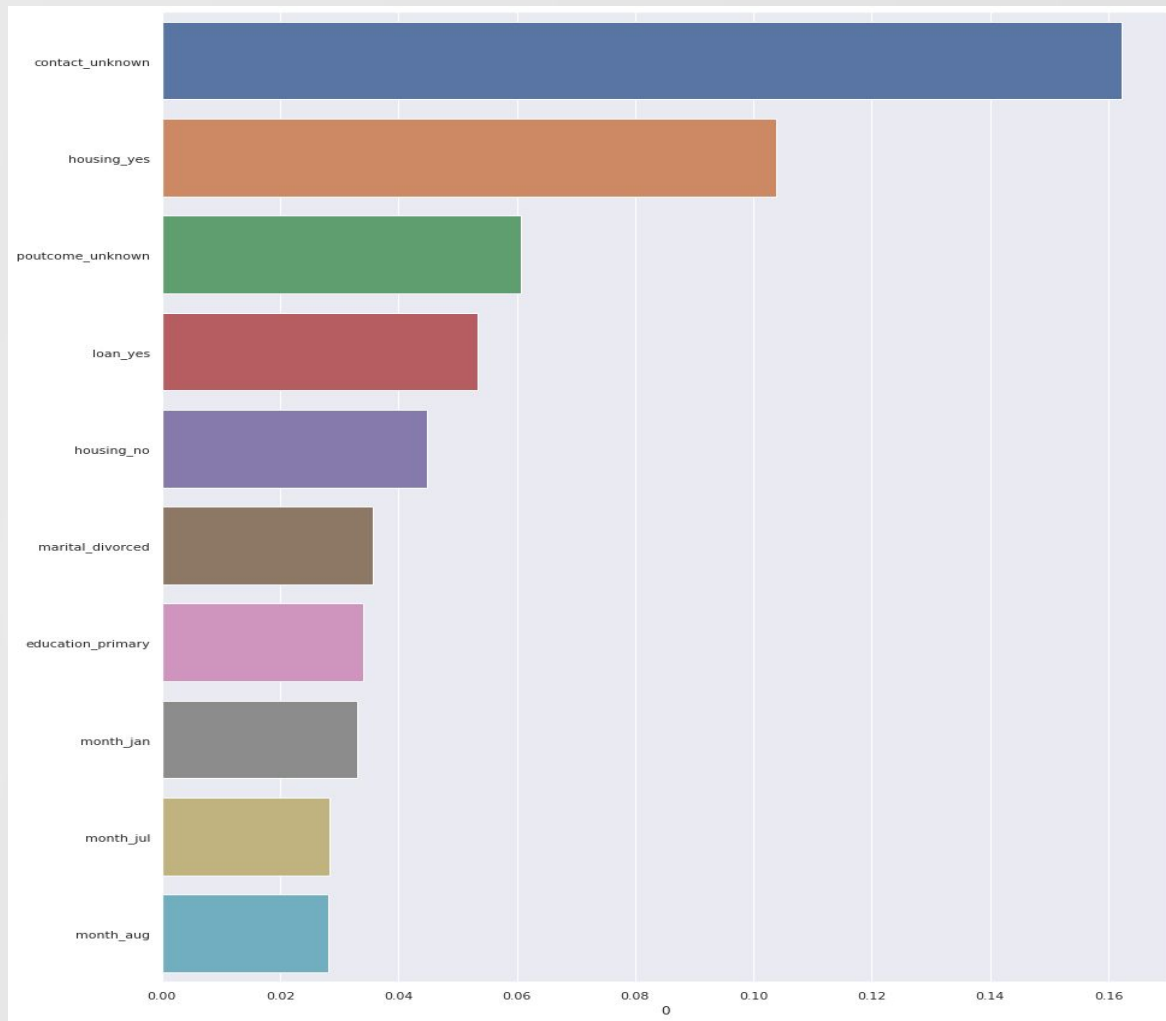


Oversampling Model

Conjunto	Recall	Precisão
Validação	0.56	0.61
Teste	0.52	0.62

Oversampling Model

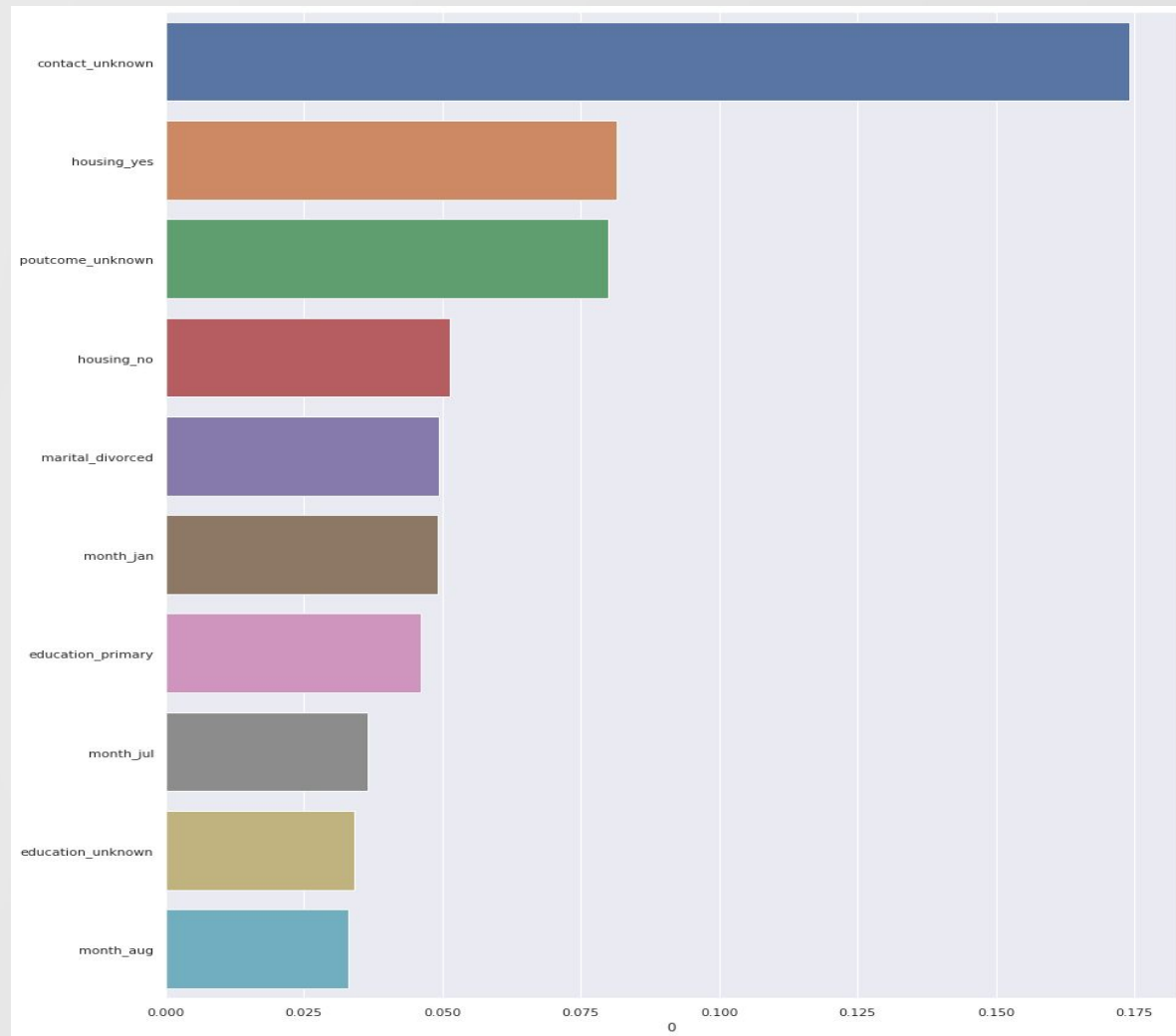
- Importância de Features



Oversampling + Feature Selection Model

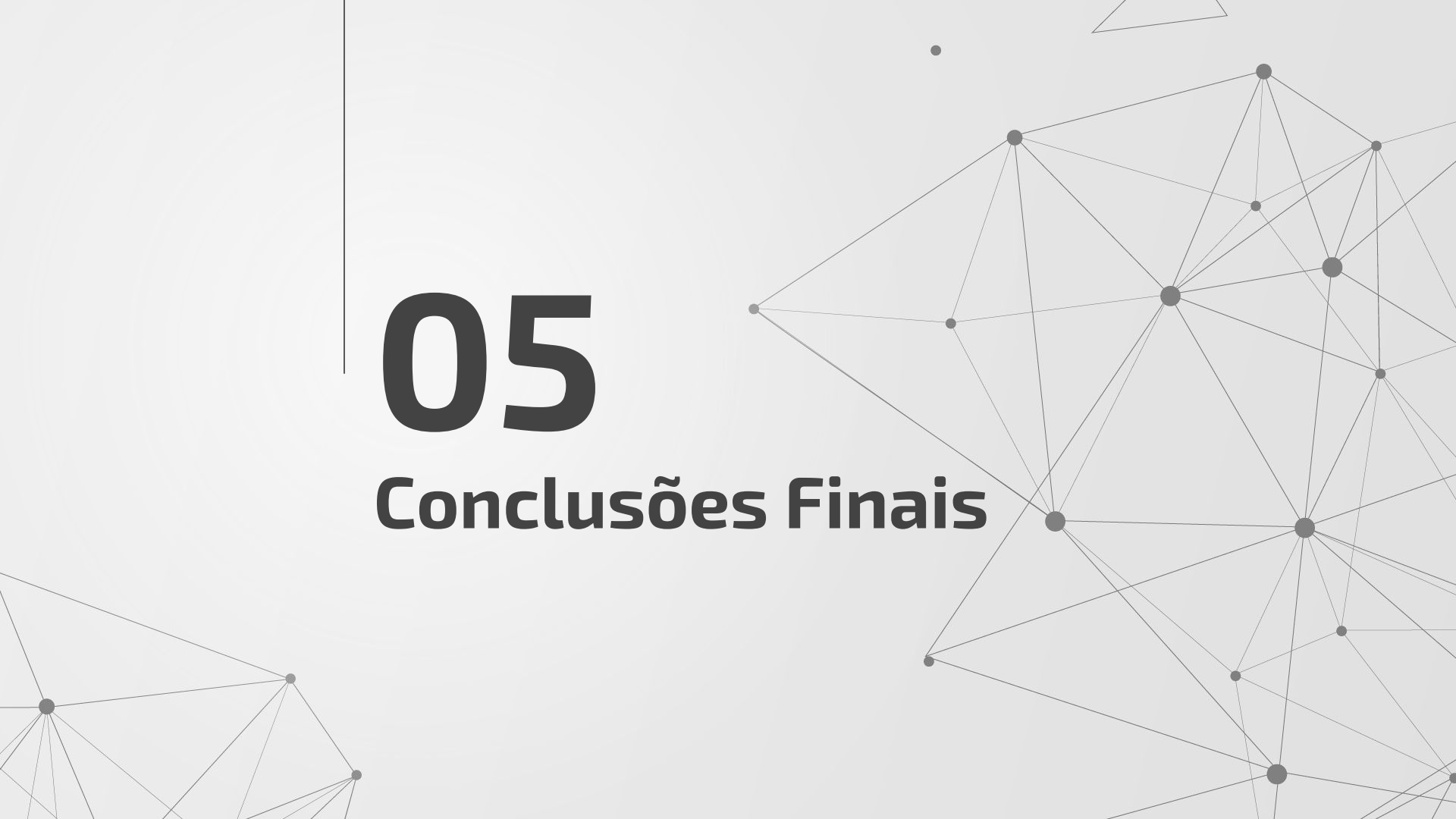
Conjunto	Recall	Precisão
Validação	0.54	0.59
Teste	0.48	0.59

Oversampling + Feature Selection Model - Importância de Features



05

Conclusões Finais



Conclusões - Variáveis

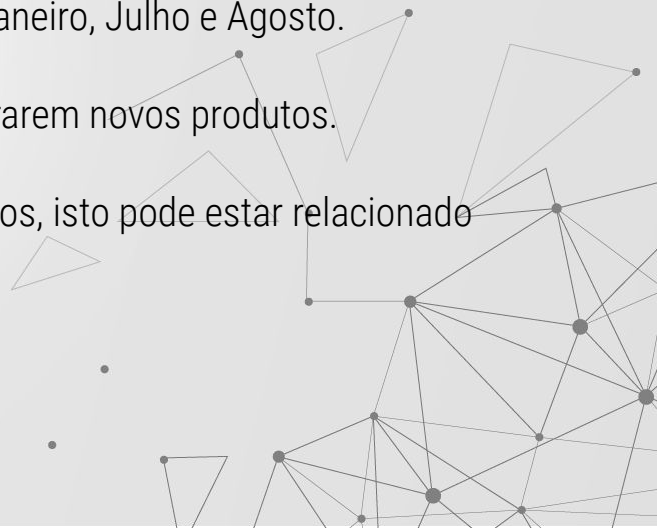
Quando o cliente contactado aderiu a campanhas anteriores, sua chance de aderência à um novo produto é muito grande.

A duração do contato também é um fator muito importante, clientes que aderiram à nova campanha tiveram uma duração de contato 2x maior do que clientes que não aderiram.

O mês de contato também interessa muito, sendo os mais propícios: Janeiro, Julho e Agosto.

Clientes que são contactados menos vezes têm mais chance de comprarem novos produtos.

Clientes que não têm sua casa financiada podem aderir a novos produtos, isto pode estar relacionado a estabilidade financeira do mesmo.



Conclusões - Modelos

O modelo selecionado para produção é o Oversampling Model, já que apresenta um bom trade off entre precisão e recall.

Esse trade off se mantém estável tanto no conjunto de validação quanto no conjunto de teste.



Conclusões - Estratégias para Vendas

Para maiores chances de sucesso em suas campanhas, a instituição bancária deve priorizar clientes com histórico de adquirir novos produtos em campanhas anteriores.

Estes clientes devem ser contactados de forma mais esporádica, porém, deve-se utilizar estratégias para engajar os clientes em conversas mais longas.

A análise da situação financeira de um cliente também um fator determinante, sendo necessário focar em clientes estáveis financeiramente.

