

Nome do candidato: Diogo Bueno Moreira

Questões

1. Qual profissão tem mais tendência a fazer um empréstimo? De qual tipo?

Optei por realizar um código em que a informação pudesse ser plotada, para facilitar a visualização dos dados. Criei 2 novos conjuntos de dados, sendo um só com linhas de “loan = yes”, referente ao caso em que o cliente tinha um empréstimo pessoal, e outro somente com linhas de “housing = yes”, caso em que o cliente tinha um empréstimo imobiliário. Fiz a contagem para cada profissão e plotei os gráficos.

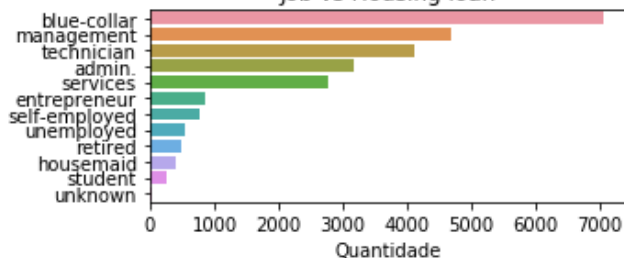
Housing Loan

blue-collar	7048
management	4678
technician	4115
admin.	3182
services	2766
entrepreneur	869
self-employed	765
unemployed	543
retired	491
housemaid	398
student	249
unknown	26

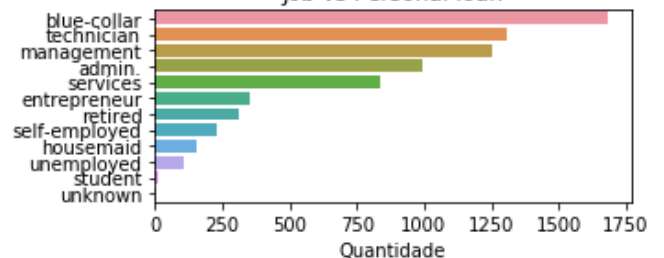
Personal Loan

blue-collar	1684
technician	1309
management	1253
admin.	991
services	836
entrepreneur	356
retired	309
self-employed	229
housemaid	152
unemployed	109
student	12
unknown	4

Job vs Housing loan



Job vs Personal loan



Dessa forma, ao analisar-se as imagens e a contagem dos dados, infere-se que a profissão “blue-collar” tem mais tendência a realizar um empréstimo, principalmente para imóvel, ou seja, do tipo “housing loan”. Mesmo para o caso

do empréstimo pessoal (“personal loan”) essa profissão também é a que mais adquire.

2. Fazendo uma relação entre número de contatos e sucesso da campanha quais são os pontos relevantes a serem observados?

Nessa questão fiquei em dúvida sobre qual campanha a pergunta se refere (atual ou anterior). Fiz a análise com relação à campanha atual.

Criei duas novas bases de dados, sendo uma contendo apenas as linhas com a variável de saída $y = \text{yes}$ e outra com $y = \text{no}$, pois representavam, respectivamente, se o cliente adquiriu ou não o serviço oferecido pelo banco (depósito a prazo).

Do total de 45.211 clientes que foram contactados, apenas 5.289 (11,7%) inscreveram-se para um termo de depósito, enquanto que 39.922 (88,3%) não. Ao analisar o número de contatos realizado para cada cliente – tanto os que adquiriram um depósito a prazo, como os que não – e levando em consideração apenas os 10 primeiros números de contatos que mais se repetiram, obteve-se a seguinte tabela:

y = yes			y = no		
Nº de contatos	Repetições	Percentual	Nº de contatos	Repetições	Percentual
1	2.561	48,42%	1	14.983	37,53%
2	1.401	26,49%	2	11.104	27,81%
3	618	11,68%	3	4.903	12,28%
4	317	5,99%	4	3.205	8,03%
5	139	2,63%	5	1.625	4,07%
6	92	1,74%	6	1.199	3,00%
7	47	0,89%	7	688	1,72%
8	32	0,61%	8	508	1,27%
9	21	0,40%	9	306	0,77%
10	14	0,26%	10	252	0,63%

A partir da análise da tabela acima, observa-se que o peso do primeiro contato é muito relevante para o êxito ou não da campanha. Isso demonstra que deve ser dado um cuidado especial ao 1º contato com o cliente, haja vista que praticamente a metade dos clientes que responderam positivamente à campanha, o fizeram com apenas 1 contato.

Por outro lado, dos clientes que não aderiram à campanha, boa parte foram contactados apenas 1 vez. Nesse contexto, isso demonstra que faltou trabalhar um pouco mais com esses clientes (fazer novos contatos) na tentativa de convencê-los. Ao analisar-se o valor médio de contatos para os dois casos, obteve-se: 2.846350 para clientes que não aderiram à campanha e 2.141047 para clientes que aderiram à campanha. Isso corrobora com o fato de que não se contactou muito aos clientes que não aderiram ao serviço ofertado pelo banco, pois em média, foram realizados menos de 3 contatos. Entretanto, é necessário um certo cuidado com a quantidade máxima de contatos, pois conforme esse número aumenta, também há diminuição da adesão.

3. Baseando-se nos resultados de adesão desta campanha qual o número médio e o máximo de ligações que você indica para otimizar a adesão?

Utilizei o mesmo conjunto de dados da questão 2, ou seja, contendo somente as linhas com resultado de $y = \text{yes}$. Conforme já mencionado, com essa filtragem restaram 5.289 linhas do conjunto de dados inicial que, originalmente, continha 45.211. Após isso, fiz a análise estatística da coluna “campaign”, contendo o número de contatos realizado para cada cliente. O resultado obtido foi o seguinte:

count	5289.000000
mean	2.141047
std	1.921826
min	1.000000
25%	1.000000
50%	2.000000
75%	3.000000
max	32.000000

Baseado nesses dados, o número médio indicado seria de 2 ligações. Para verificar o valor máximo, averigui o número de ligações e quantas vezes o mesmo se repetia, baseado na tabela gerada na questão 2.

Diante disso, o número máximo de 3 ligações seria uma boa alternativa para otimizar a adesão, pois se levarmos em consideração o grupo de clientes que aderiram à campanha, o total deles que receberam 3 chamadas está em torno de 11,7%, sendo o 3º maior número da lista. Além disso, conforme verificou-se anteriormente na questão 2, o primeiro contato é o mais decisivo e ao aumentar-se o número de ligações, a tendência é diminuir a quantidade de clientes que aderem à campanha.

4. O resultado da campanha anterior tem relevância na campanha atual?

Para a resolução dessa questão, chequei a correlação entre as variáveis “poutcome”, referente o resultado da campanha anterior, e “y”, correspondente a adesão do serviço oferecido na campanha atual. Com isso, verifica-se que o fato da campanha anterior haver sido bem sucedida (“sucess”) apresenta uma boa correlação com o sucesso da campanha atual (“y_yes”). Isso leva a concluir que o resultado da campanha anterior tem relevância na campanha atual. Além disso, é possível inferir que a duração (em segundos), referente ao último contato com o cliente, também possui relevância na campanha atual.

FEATURES	CORRELAÇÃO COM Y_YES
Duration	0.394521
poutcome_success	0.306788
month_mar	0.129456
month_oct	0.128531
month_sep	0.123185
Pdays	0.103621
Previous	0.093236
job_retired	0.079245
job_student	0.076897
month_dec	0.075164
education_tertiary	0.066448
marital_single	0.063526
Balance	0.052838
month_feb	0.038417
job_management	0.032919
poutcome_other	0.031955
Age	0.025155
job_unemployed	0.020390
contact_telephone	0.014042
education_unknown	0.012053
job_self-employed	0.000855
job_unknown	0.000267
month_aug	-0.008536
month_jan	-0.008783
job_technician	-0.008970
month_nov	-0.014937
job_housemaid	-0.015195
month_jun	-0.016805
job_entrepreneur	-0.019662
default_yes	-0.022419
job_services	-0.027864
Day	-0.028348
month_jul	-0.034382
education_secondary	-0.036388
marital_married	-0.060260
loan_yes	-0.068185
job_blue-collar	-0.072083
Campaign	-0.073172
month_may	-0.102500
housing_yes	-0.139173
contact_unknown	-0.150935
poutcome_unknown	-0.167051

5. Qual o fator determinante para que o banco exija um seguro de crédito?

Para essa análise, usei a mesma base de dados da questão 3, ou seja, contendo somente as linhas de dados em que os clientes aderiram à campanha (y=yes). Nesse tipo de análise, as colunas relevantes são “default”, variável que informa se o cliente tem inadimplência ou não, e “balance”, a qual informa o saldo médio anual do cliente. Fazendo a contagem dos clientes inadimplentes, que realizaram a aquisição do serviço oferecido pelo banco, o total foi de apenas 52. Por outro lado, ao checar a renda média anual desses clientes, o total de clientes com renda média anual negativa foi de 502. A renda média desses clientes é um valor baixo (em torno de € 1.804,00 – vide figura abaixo). Além disso, outro fator relevante, foi fazer a contagem de quantos clientes receberam acima e abaixo da média anual. Os totais encontrados foram, respectivamente, 1.529 e 3.760. Isso demonstra que a *feature* “balance” é um fator determinante para que o banco exija um seguro de crédito.

Index	balance
count	5289.000000
mean	1804.267915
std	3501.104777
min	-3058.000000
25%	210.000000
50%	733.000000
75%	2159.000000
max	81204.000000

6. Quais são as características mais proeminentes de um cliente que possua empréstimo imobiliário?

A fim de realizar essa análise, reduzi o conjunto de dados para apenas as linhas em que “housing = yes”, referente ao fato do cliente possuir empréstimo imobiliário. O conjunto de dados ficou com 25.130 linhas. Analisando os dados estatísticos das variáveis numéricas, percebe-se que a média de idade é de 39 anos e a renda anual média é de € 1.175,00, conforme figura abaixo:

Index	age	balance	day	duration	campaign	pdays	previous
count	25130.000000	25130.000000	25130.000000	25130.000000	25130.000000	25130.000000	25130.000000
mean	39.175288	1175.103064	15.598249	259.331476	2.698488	51.312495	0.656665
std	8.926807	2483.285761	8.026836	258.321907	3.140979	113.862848	2.707428
min	20.000000	-4057.000000	1.000000	0.000000	1.000000	-1.000000	0.000000
25%	32.000000	62.000000	8.000000	103.000000	1.000000	-1.000000	0.000000
50%	38.000000	412.000000	15.000000	182.000000	2.000000	-1.000000	0.000000
75%	45.000000	1257.750000	21.000000	318.000000	3.000000	-1.000000	0.000000
max	78.000000	58544.000000	31.000000	4918.000000	63.000000	854.000000	275.000000

Já para o caso das variáveis categóricas, plotei os gráficos para uma melhor análise visual (vide figuras abaixo). É possível perceber que a maioria dos clientes que possuem empréstimo imobiliário, apresentam as seguintes características proeminentes: trabalham com atividades que envolvem força física e habilidade manual (“blue-collar”), são casados, possuem educação secundária e não são inadimplentes.

