

Resposta ao questionário de Semantix

Candidato: Gabriel Ernesto Jara

Questões

1. Qual profissão tem mais tendência a fazer um empréstimo? De qual tipo?
2. Fazendo uma relação entre número de contatos e sucesso da campanha quais são os pontos relevantes a serem observados?
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?
4. O resultado da campanha anterior tem relevância na campanha atual?
5. Qual o fator determinante para que o banco exija um seguro de crédito?
6. Quais são as características mais proeminentes de um cliente que possua empréstimo imobiliário?

Resposta às questões

Questão 1

Pergunta: Qual profissão tem mais tendência a fazer um empréstimo? De qual tipo?

Resolução: Para resolver a questão, filtrei os clientes que possuem um empréstimo, seguido da contagem dos clientes em cada profissão. Na tabela 1, se apresentam os resultados.

Tabela 1: Número de empréstimos por profissão. A coluna “**Total de empréstimos**” referece a o total de clientes com uma dada profissão que fizeram um só tipo de empréstimo ou os dois. A coluna “**Total de empréstimos imobiliário**” referece a o total de clientes com uma dada profissão que fizeram um empréstimo imobiliário independentemente de se também fizeram um empréstimo pessoal. A coluna “**Total de empréstimos pessoais**” referece a o total de clientes com uma dada profissão que fizeram um empréstimo pessoal independentemente de se também fizeram um empréstimo imobiliário. A coluna “**Somente empréstimos imobiliário**” referece a o total de clientes com uma dada profissão que somente fizeram um empréstimo imobiliário. A coluna “**Somente empréstimos de pessoal**” referece a o total de clientes com uma dada profissão que somente fizeram um empréstimo pessoal.

Profissão	Total de empréstimos	Total de empréstimos imobiliário	Total de empréstimos pessoais	Somente empréstimos imobiliário	Somente empréstimos pessoais
blue-collar	7599	7048	1684	5915	551
management	5225	4678	1253	3972	547
technician	4608	4115	1309	3299	493
admin.	3569	3182	991	2578	387
services	3063	2766	836	2227	297
entrepreneur	1016	869	356	660	147
self-employed	864	765	229	635	99
retired	699	491	309	390	208
unemployed	589	543	109	480	46
housemaid	490	398	152	338	92
student	257	249	12	245	8
unknown	28	26	4	24	2

Resposta: A tabela mostra que os trabalhadores braçais (blue-collar) são os que tem mais tendência para fazer ou pedir um empréstimo e o empréstimo imobiliário é o tipo de empréstimo que maior tendência tem os blu-collars , mas também todas as outras profissões.

Arquivos: Todos os arquivos estão na pasta Q1. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q1. O arquivo script.sh precisa do arquivos professions.dat, o qual contém a lista de profissões. A tabela 1 é gerada pelo output output-profession-with-loans.csv. O arquivo housing-loan-vs-profession.dat contém a coluna “**Total de empréstimos imobiliário**” O arquivo only-housing-loan-vs-profession.dat contém a coluna **Somente empréstimos imobiliário**”. O arquivo personal-loan-vs-profession.dat contém a coluna “**Total de empréstimos pessoais**”. O arquivo only-personal-loan-vs-profession.dat contém a coluna “**Somente empréstimos de pessoal**”.

Para a resposta das questões 2 e 3 a análise se fez separaram os clientes contatados em grupos por quantidade de número de contato realizado.

Questão 2

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

Resolução: Para responder à questão, primeiro filtrei os dados, separando em dos arquivos diferentes os resultados positivos (arquivo yes.dat - clientes que subscreveram ao investimento) do negativos (arquivo no.dat - clientes que não subscreveram ao investimento). Os arquivos estão na pasta Q2. Este arquivos serviram para fazer os histogramas de número de pessoas contatadas (que subscreveram e que não) versus o número de contatos realizados. Estes histogramas se mostram na Figura 1.

Comentário: Para considerar o sucesso de campanha é preciso considerar a relação entre quantidades de clientes que aceitaram o investimento (contato com sucesso) e de clientes que não aceitaram (contato sem sucesso ou falhado).

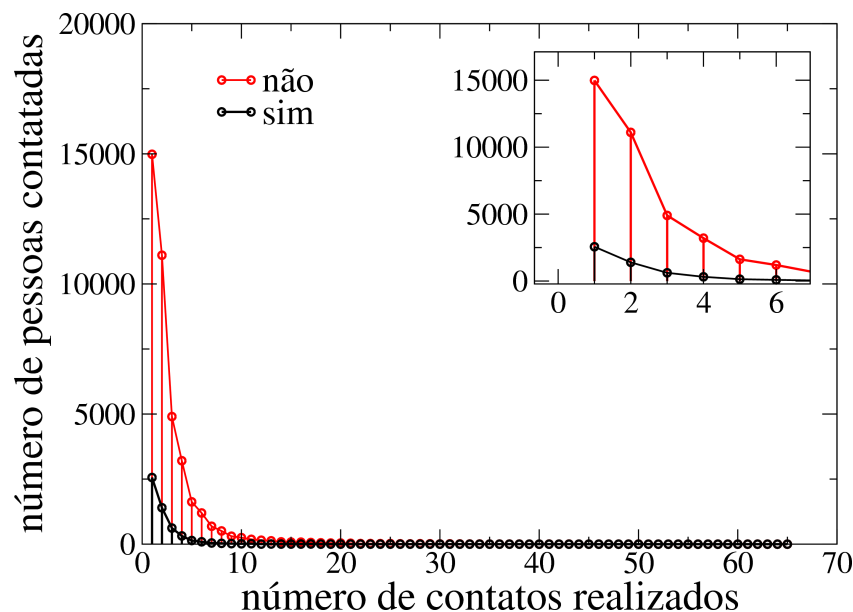


Figura 1: Histograma de número de pessoas contatadas que subscreveram (yes, preto) e que não subscreveram (não, vermelhos) versus o número de contatos realizados.

Resultado: A figura 1 indica que os pontos mais relevantes da relação o número de contatos e o sucesso da campanha são: a) do total de contatos a quantidade de contatos que foram bem-sucedidos é menor que aqueles que resultaram falhidos. b) menores número de contatos são melhor bem-sucedidos que número de contatos maiores.

Arquivos: Todos os arquivos estão na pasta Q2. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q2. O arquivos histo-yes.dat e histo-no.dat são os histogramas para as pessoas que subscreveram ao investimento e os que não se subscreveram, respetivamente. O arquivo histograma.agr é o gráfico apresentado na figura 1, está no formato arg para se leido pelo programa GRACE (usado para fazer os gráficos) livremente disponível para os Sistemas Operativos LINUX e WINDOWS.

Questão 3

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

Resolução: O histograma da figura 1 foi transformado a porcentagens relativas ao total de clientes contatados. Posteriormente, se calculou a porcentagem acumulativa ao longo do eixo X (número de contato). Ver Figura 2.

Gerou-se, a partir dos dados dos histograma na Figura 1, se calculou o total de contatos realizados na campanha. Este valor se usou para calcular a porcentagem de contatos realizados para cada grupo de clientes separados por “número de contato”. Com esse valor se calculou a porcentagem acumulativa a longo do eixo X (número de contato), que mede o esforço feito quando se aumentam o número de contato por cliente. Ver Figura 3.

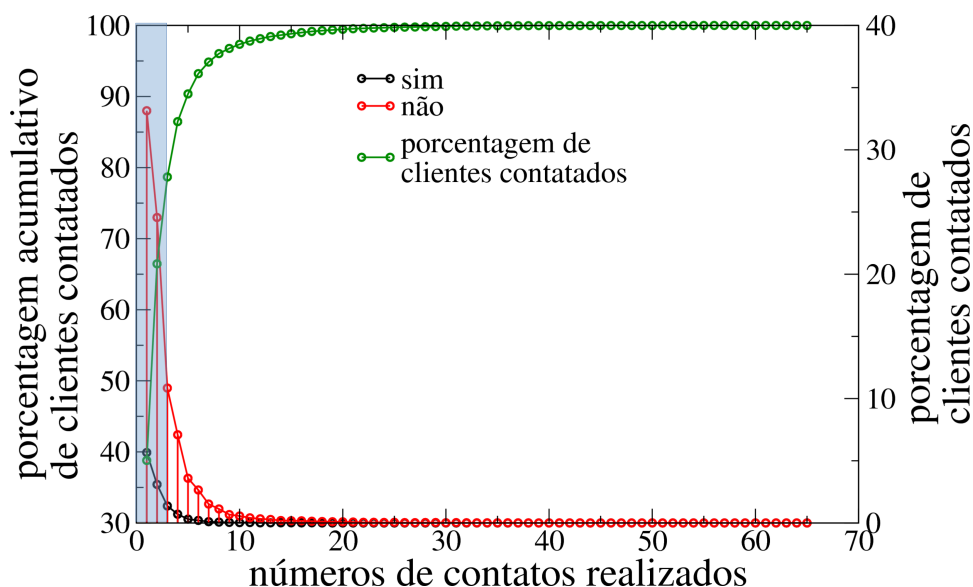


Figura 2: Histograma da porcentagem de pessoas contatadas que subscreveram (yes, preto) e que não subscreveram (não, vermelhos) versus o número de contatos realizados. A linea em verde é a porcentagem acumulativa a longo va variação na variável número de contato. A área em azul indica a região envolvida até um máximo de 3 contatos.

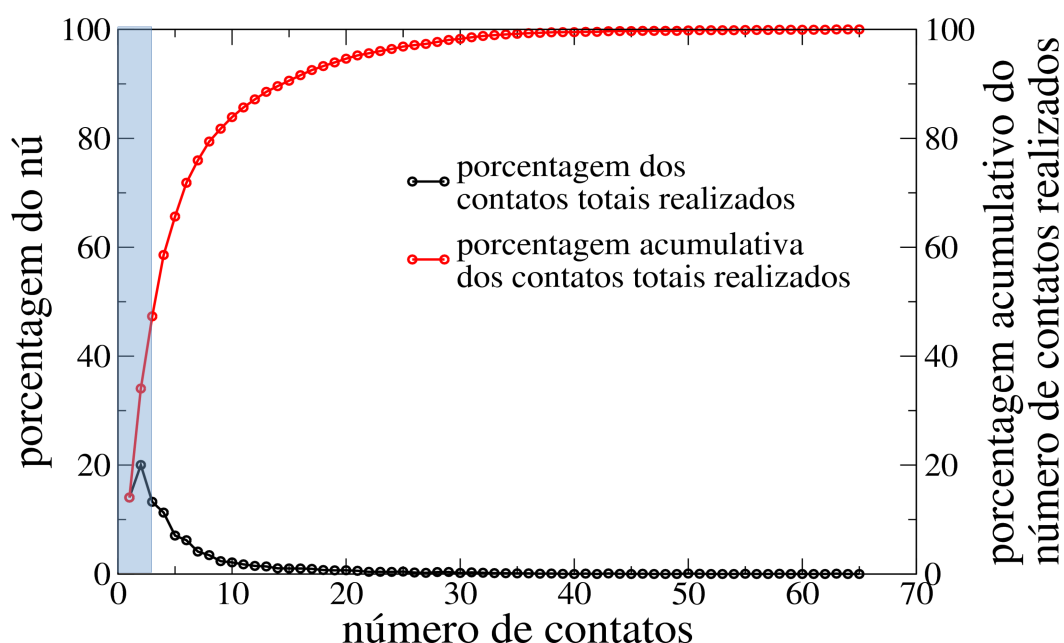


Figura 3: Porcentagem do contatos totais realizadas para cada grupo de clientes separados por “número de contatos” (linha persa). A linea vermelha é a porcentagem acumulativa dos contatos totais realizados ao longo va variação na variável “número de contato”. A área em azul indica a região envolvida até um máximo de 3 contatos.

Resultado: O número de ligações máximo para otimizar a adesão é 3 e o número médio é 2. Esta resposta se baseia na seguinte explicação A análise da figura 2 indica que com um número de contato máximo de 3 já se alcanço o 80% de todos los clientes contatos e, ademias, a partir a adesão de clientes é muito baixa. A figura 3 indica que com um máximo de “número de contato” de 3 se alcançou um esforço de 40% (é dizer, se fizeram um 40% de total de chamadas na campanha).

Arquivos: Todos os arquivos estão na pasta Q3. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q3. O arquivos histo-yes-porcentagem.dat e histo-no-porcentagem.dat são os histogramas que representam o porcentagem de das pessoas que subscreveram ao investimento e os que no se subcreveram, respetivamente, em função do número de contato realizado. O arquivo somatoria-acumulativa.dat é a somatória acumulativa das porcentagem a longo do eixo X (número de contato). O arquivo medicaao-esforco-porcentagem.dat é a porcentagem do contatos totais realizadas para cada grupo de clientes separados por “número de contatos”. O aquriovo somatoria-acumulativa-medicao-esforco-porcentagem.dat é a porcentage acumulativa dos contatos totais realizados ao longo va variação na variável “número de contato”. As figuras 2 e 3 são os arquivos figura2-prev.agr e figura3-prev.agr, estão no formato arg para se lido pelo programa GRACE livremente disponível para os Sistemas Operativos LINUX e WINDOWS. A figuras forão editadas para agregar a região marcada em azul com libreoffice e se gerou o arquivo figura2.ppt.

Questão 4

Pergunta: O resultado da campanha anterior tem relevância na campanha atual?

Resolução: contagem da diferentes possibilidades. Por exemplo: a) quantidade de clientes que com SUCESS na campanha prévia mas que disseram NO nesta campanha. b) Quantidade de clientes que com SUCCESS na campanha prévia mas que disseram SIM nesta campanha. Assim cruzando as alternativas de success failure unkown e other da campnha prévia com as possibilidades de YES or NO da presente campanha. Estes dados se usaram para construir a Tabela 2.

Tabela 2: Relação entre os dados da campanha prévia e a a adesão nesta campanha. A tabela apresenta a quantidades de cliente que adereram (yes ou não (no) nesta campanha e que na campanha prévia adereram (success), não adereram (failure), s/n (unknown) e otros (other).

This campaign	Previous Campaign			
	Sucesso	Failure	Unknown	Other
yes	978	618	2	306
no	533	4280	3	1531

Resultado: O resultado da campanha anterior teve alguma relevância na presente campanha porque o 64 % (978) dos cliente, que foram registrados como sucesso na campanha prévia, subscreveram nesta campanha.

Arquivos: Todos os arquivos estão na pasta Q4. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q4. O arquivo clients-were-contacted-prev-campaign.dat é separação dos cliente que foram previamente contatados. O arquivo previous-campaign-vs-this-campaign.dat é separaração dos campos de sucesso na campanha prévia e dos resultados desta campanha. Os arquivos previous-success-this-yes.dat; previous-failure-this-yes.dat; previous-unknown-this-yes.dat; previous-other-this-yes.dat são a quantidade de clientes que subscrevram-se na presente campanha e que na campanha prévia foram success, failure, unknoen ou other, respetivamente. Os arquivos previous-success-this-no.dat; previous-failure-this-no.dat; previous-unknown-this-no.dat; previous-other-this-no.dat são a quantidade de clientes que NÃO subscrevram-se na presente campanha e que na campanha prévia foram success, failure, unknoen ou other, respetivamente. O arquivo tabela-2.csv contém os dados na tabela 2. A tabela 2 está em formato xls no arquivo tabela-2.xls.

Questão 5

Pergunta: Qual o fator determinante para que o banco exija um seguro de crédito?

Resolução: Filtrou-se todos os clientes, que têm algum crédito em default. A justificativa é que se já possuem algum crédito em default são clientes apresentam risco para o banco. Para determinar quais características ou fatores são mais determinantes dentro deste grupo de clientes, se fez a contagem de clientes com as seguintes características: clientes com balance negativo (1), zero (2) ou positivo (3); clientes com empréstimo imobiliário (4), pessoal (5), os dois empréstimos (6), nenhum desses empréstimos (7), tendo algum empréstimo (8), e tendo algum empréstimo e balance negativo (9). Primeramente a análise se fez pensando que clientes em default poderiam apresentar balance negativo ou, além disso, ter já algum empréstimo. Contudo, essa análise se ampliou para balances zero e positivos.

Tabela 3: Tabela mostrando a quantidades de clientes algum crédito em default para diferentes fatores.

balance			empréstimo					algum empréstimo e balance negativo
negativo	zero	positivo	imobiliário	pessoal	algum empréstimo	não tendo empréstimo imobiliário ou pessoal	dois empréstimos: imobiliário e pessoal	
440	116	259	302	168	603	212	133	357

Resultado: Segundo a tabela e análise feito, o fator determinante para que o banco exija um seguro de crédito é que clientes possua algum empréstimo (seja imobiliário, pessoal ou os dois). Estes clientes representam o grupo de maior número segundo a tabela 3.

Arquivos: Todos os arquivos estão na pasta Q5. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q5. Os arquivos balance-negative.dat balance-zero.dat balance-positive.dat contém a quantidade de clientes com crédito em default e balance negativo, zero e positivo, respectivamente. Os arquivos having-housing-loan.dat having-personal-loan.dat having-a-loan.dat having-none-loan.dat having-a-housing-and-personal-loans.dat são os arquivos que contém a quantidade de clientes com crédito em default e com empréstimo imobiliário, pessoal, algum destes dois empréstimos, nenhum empréstimo, possuindo esses dois tipos de empréstimos, respectivamente. O arquivo having-a-loan-and-balance-negative.dat contém os clientes com crédito em default possuindo empréstimos imobiliário ou pessoal e tendo um balance negativo. O arquivo tabela-3.csv contém a informação da tabela 3. A tabela 3 também está no formato xls.

Questão 6

Pergunta: Quais são as características mais proeminentes de um cliente que possua empréstimo imobiliário?

Resolução: Filtrei os clientes que possuem um empréstimo imobiliário. Trabalhando com este filtro, fiz uma contagem dos clientes com um empréstimo imobiliário e com características pessoais como tipo de emprego, educação e estado civil (tabela 4). Por outro lado, fiz uma histograma da idade e o balance dos clientes com empréstimo imobiliário (figura 4 e 5). Assim, esperava-se achar quais são as características mais frequentes dos clientes que possuem um empréstimo imobiliário.

Tabela 4: Quantidade de clientes com empréstimo imobiliário segundo seu trabalho, estado civil e educação.

job		marital		education	
unknown	20188	married	15321	unknown	20188
blue-collar	7048	single	6902	secondary	14038
management	4678	divorced	2907	tertiary	6378
technician	4115			primary	3894
admin.	3182				
services	2766				
entrepreneur	869				
self-employed	765				
unemployed	543				
retired	491				
housemaid	398				
student	249				

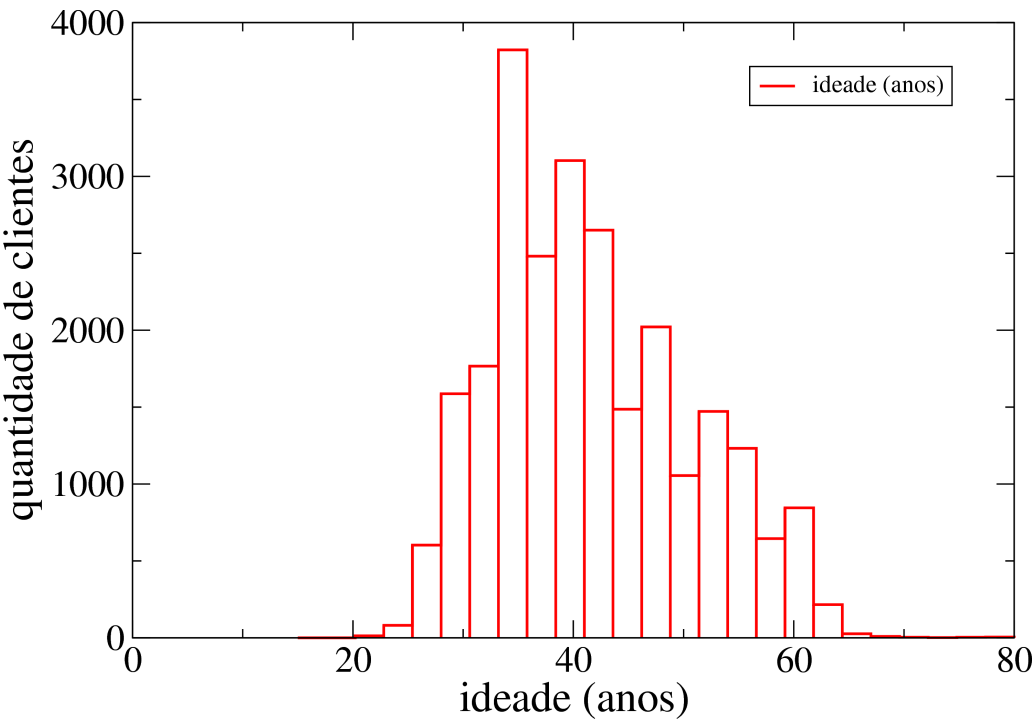


Figura 4: Histograma da idade dos clientes que possuem um empréstimo imobiliário.

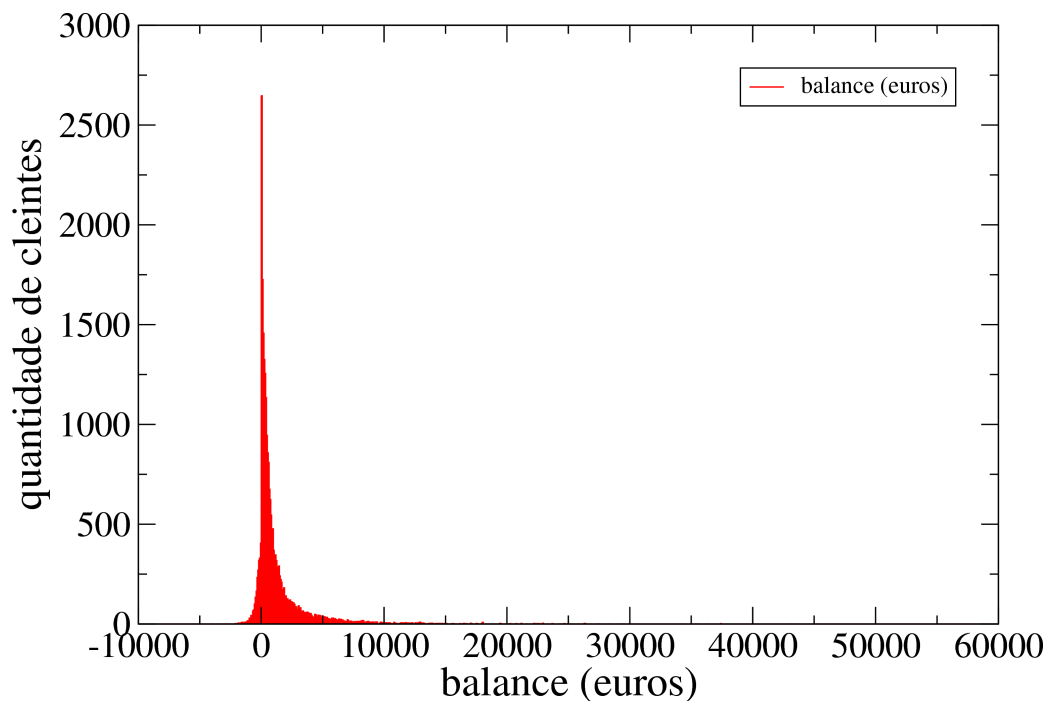


Figura 5: Histograma do balance dos clientes que possuem um empréstimo imobiliário.

Resultado: Segundo a tabela 4 as características proeminentes de um cliente que possui um empréstimo imobiliário são: trabalho “blue-collar”, estado civil casado e educação de ensino médio. No caso da idade, a figura 4 indica que os clientes com idade entre 35 e 45 anos são os que mais frequentemente possuem o empréstimo imobiliário. A análise do balance indica que os clientes com balance zero até positivo são os que mais possuem o empréstimo.

Arquivos: Todos os arquivos estão na pasta Q6. O arquivo script.sh é o executável para a geração dos outputs. Os outputs são gerados na pasta outputs dentro da pasta Q6. Os arquivos job.dat, education.dat e marital.dat contêm os valores que as variáveis job, education e marital podem assumir. Os arquivos job-*.dat, marital-*.dat e education-* contêm quantidade de clientes com empréstimo imobiliário para cada variável em job, marital, e education, respectivamente. Os arquivos age.dat e balance.dat possuem a idade e o balance de todos os clientes com empréstimo imobiliário. Os arquivos histograma-idade.agr e histograma-balance.agr são arquivos para serem lidos pelo programa GRACE (usado para fazer os gráficos) livremente disponível para os Sistemas Operativos LINUX e WINDOWS. O arquivo tabela-4.xls é a tabela 4.