

Análise estatística de dados com objetivo de identificar clientes propensos a cancelar produtos de um banco

Daniel V. F. Falbel

16 de Novembro de 2014

Resumo

Para auxiliar um banco na criação de ações publicitárias para retenção de clientes que possuem um certo cartão de crédito, este trabalho apresenta uma análise estatística que permite a identificação dos clientes mais propensos a cancelar o produto.

1 Descrição do problema

Um banco deseja fazer ações de marketing para reter os seus clientes, evitar que eles cancelem seus produtos. Para poder fazer ações mais assertivas, o banco deseja saber qual é o perfil dos clientes com maior propensão a cancelar um certo cartão de crédito.

As variáveis que podem ajudar na identificação dos perfis estão listadas abaixo:

- sexo: M-masculino; F-feminino
- modulo: Segmentação de clientes; valores mais baixos representam clientes com menor renda ou investimento, valores mais altos representam clientes mais interessantes para a instituição
- cheque: Classificação da conta corrente; quanto maior o valor, mais "especial" o cliente
- evolcredor: Evolução do saldo credor médio trimestral (A: aumentou, D: diminuiu, M: manteve)
- evoldevedor: Evolução do saldo devedor médio trimestral (A: aumentou, D: diminuiu, M: manteve)
- evolpoup: do saldo da poupança trimestral (A: aumentou, D: diminuiu, M: manteve)
- idade: idade em anos
- salario: está categorizado em 10 categorias (quanto maior, maior o salário)
- cartaocancel: Cancelamento do cartão de crédito pelo banco 0 - não cancelou; 1 - cancelou
- bancsal: 0: não recebe salário pelo banco; 1: recebe salário pelo banco
- tempo: Tempo de permanência com um determinado produto, em meses
- status: 1: tempo refere-se ao tempo da contratação até o cancelamento do produto; 0: tempo da contratação até término de acompanhamento sem cancelamento (censura)

2 Análise Descritiva

Para verificar se existem possíveis inconsistências nos dados, foi feita uma análise exploratória. Além disso, podemos já ter uma ideia do que pode influenciar no tempo até o cancelamento do cartão de crédito.

Pelos gráficos da figura 1, podemos observar que 60% dos clientes são do sexo masculino. A maior parte dos clientes teve um aumento do saldo credor e do saldo devedor. Também podemos analisar que a maior parte dos clientes manteve estável o saldo da poupança. Observamos que a maior parte dos clientes estão no segmento 110 do módulo, e que cerca de 50% dos clientes estão na classificação 0 da conta corrente. Note que no gráfico (vi) existe a ocorrência de uma observação ausente, isto é, não temos a informação de que tipo de conta o cliente possui.

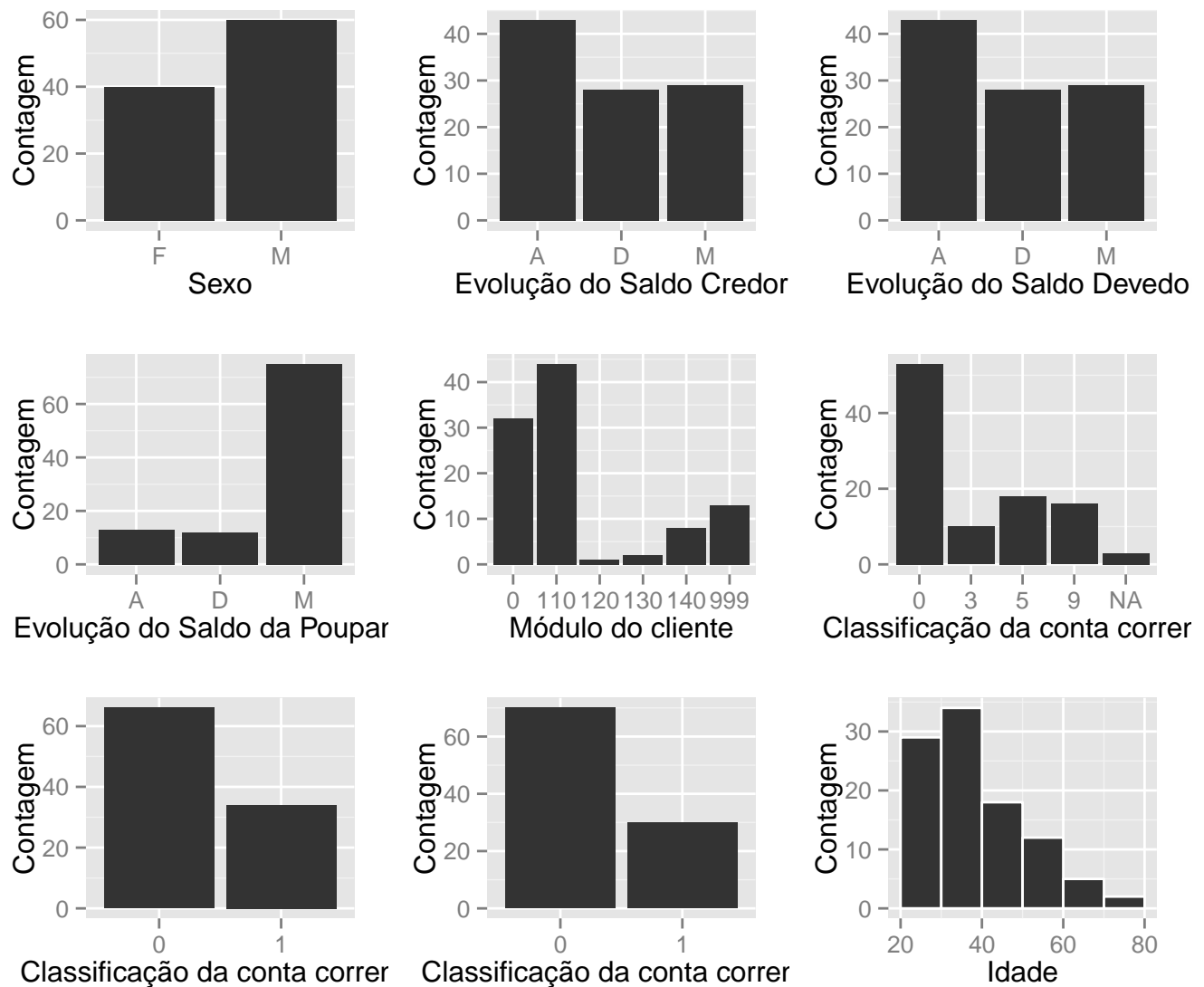


Figure 1: (i) quantidade de clientes em cada categoria da variável sexo, (ii) quantidade de clientes em cada categoria de evolução do saldo credor, (iii) quantidade de clientes em cada categoria de evolução do saldo devedor, (iv) quantidade de clientes em cada categoria de evolução do saldo da poupança, (v) quantidade de clientes em cada módulo (segmento) criado pelo banco, (vi) quantidade de clientes em cada classificação da conta corrente, (vii) quantidade de clientes que cancelaram o cartão (1), (viii) quantidade de clientes que recebem o salário por meio do banco, (ix) histograma da idade dos clientes

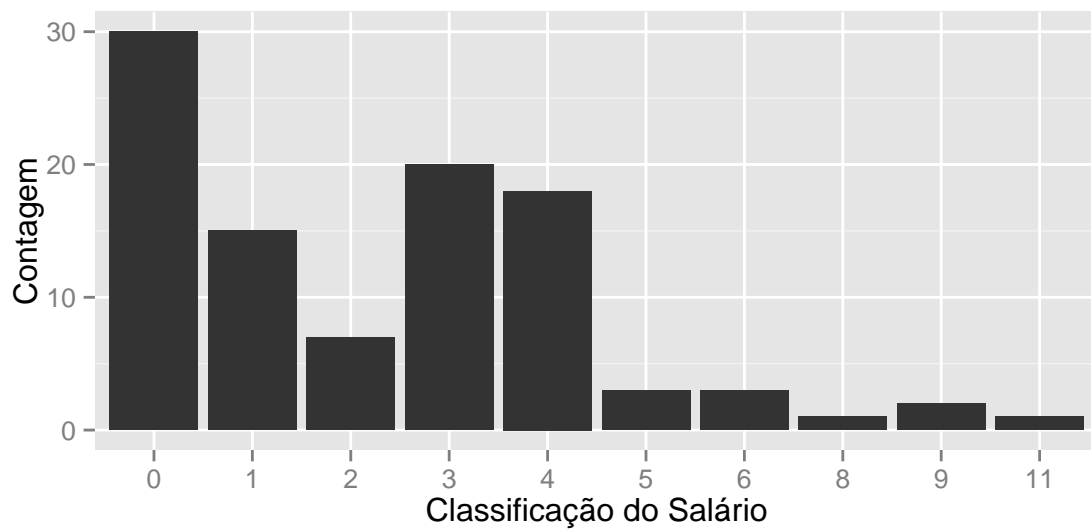


Figure 2: Quantidade de clientes em cada classificação do salário

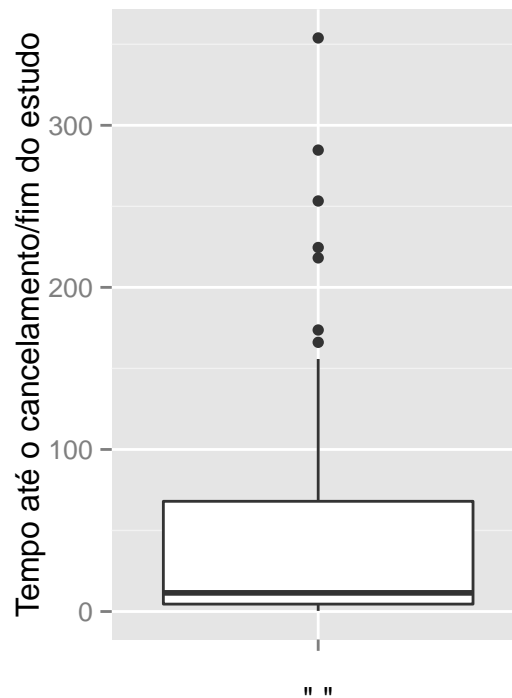


Figure 3: Boxplot do tempo até o cancelamento do cartão de crédito ou fim do acompanhamento

Pela figura 2 observamos que cerca de 30% dos clientes ganham até R\$300,00, em seguida as categorias com mais indivíduos são a 3 e a 4 que contém pessoas que ganham de R\$501,00 a R\$1500,00.

Na figura 3, vemos que os clientes em mediana têm o produto por 11 meses. Aproximadamente 75% dos clientes tem tempo até 68 meses. O indivíduo que possui o produto a mais tempo, o tem a 353 meses. No estudo observamos que 18 clientes foram censurados, ou seja, ainda não tinham cancelado o cartão de crédito até a data de fim do acompanhamento.

Como não encontramos nenhuma irregularidade nas variáveis, passamos para uma análise descritiva bivariada. Aqui cruzamos as variáveis do banco de dados com o tempo até o cancelamento do produto, para assim termos dicas de como as variáveis estão relacionadas ao tempo que o cliente permanece com o cartão.

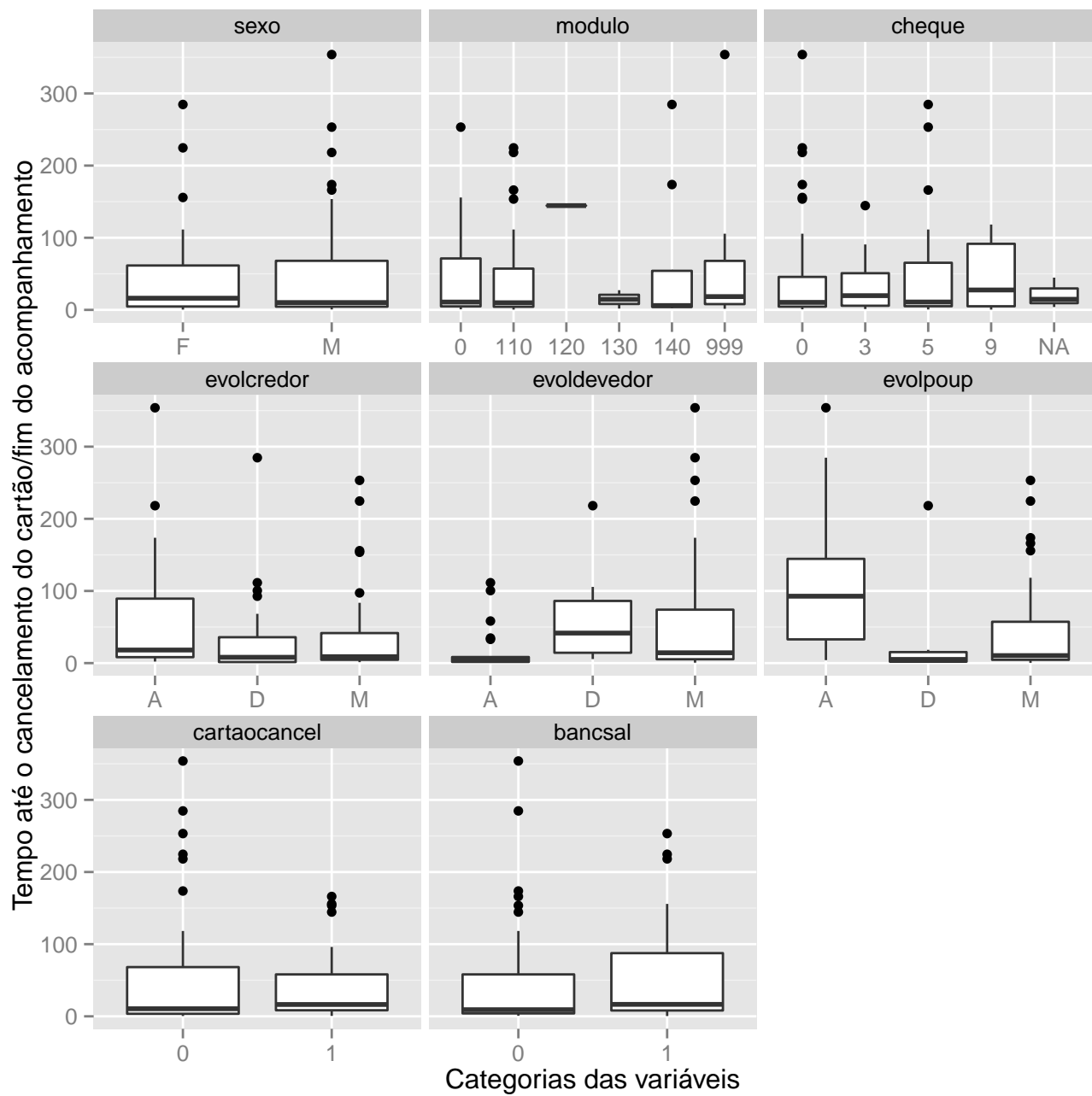


Figure 4: Boxplots do tempo até o cancelamento do cartão/fim do acompanhamento para cada variável do banco de dados