

UNIVERSIDADE DO VALE DO RIO DOS SINOS – UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO ONLINE
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS

Dirceu Roden

APLICAÇÃO DE ANÁLISE ESTATÍSTICA NA PREVISÃO DE PERDA DE
CLIENTES: ESTUDO EM UMA EMPRESA ATACADISTA.

São Leopoldo

2019

UNIVERSIDADE DO VALE DO RIO DOS SINOS – UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO ONLINE
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS

Dirceu Roden

APLICAÇÃO DE ANÁLISE ESTATÍSTICA NA PREVISÃO DE PERDA DE
CLIENTES: ESTUDO EM UMA EMPRESA ATACADISTA.

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Especialista em *Big Data, Data Science e Data Analytics*, pelo curso de Pós-Graduação Lato Sensu em *Big Data, Data Science e Data Analytics* da Universidade do Vale do Rio dos Sinos – UNISINOS.
Orientador: Prof. Ms Claudia Angelita Fagundes Raupp

São Leopoldo

2019

Aplicação de análise estatística na previsão de perda de clientes: estudo em uma empresa atacadista.

Dirceu Roden

Universidade do Vale do Rio dos Sinos - Unisinos
Av. Unisinos, 950 - Cristo Rei – 93.022-750 – São Leopoldo – RS – Brasil
dirceu.roden@yahoo.com.br

Abstract. *Winning the customer represents a high cost for a company. Generally, it is required investments in marketing, relationships, discounts, working hours in negotiations and several other actions that demand financial resources. In order to recover those expenses, the company needs to hold the client for as long as possible. In this context, the aim of this work is to propose a statistical analysis model able to foresee which clients are about to leave and stop being clients, through a study in a wholesale company. It was used the client's registration data and relationship with the company to develop a predictive model of Logistic Regression. The proposed model achieved an accuracy of 94.6% and specificity of 48%, showing that it is effective for the prediction of loss of customers.*

Resumo. *Conquistar um cliente representa um custo elevado para uma empresa. De forma geral, são necessários investimentos em marketing, relacionamentos, descontos, horas de trabalho em negociações e diversas outras ações que demandam recursos financeiros. Para recuperar essas despesas, a empresa precisa manter o cliente pelo maior tempo possível. Nesse contexto, o objetivo desse trabalho é propor um modelo de análise estatística capaz de prever quais clientes estão prestes a deixar de ser clientes, através de um estudo em uma empresa atacadista. São usados dados cadastrais dos clientes e de relacionamento com a empresa para construir um modelo preditivo de Regressão Logística. O modelo proposto obteve uma acurácia de 94,6% e uma especificidade de 48%, mostrando ser eficaz na predição de perda de clientes.*

Palavras chave: Análise Estatística, Regressão Logística, Cliente.

1. Introdução

Na sociedade capitalista são as empresas que fazem a economia andar e também são responsáveis pela maior parte da geração e circulação de riquezas. Mas uma empresa precisa ter clientes, sem eles ela não existe. Tampouco existiriam seus funcionários, seus prestadores de serviços e seus parceiros.

Cientes disso, as organizações investem pesado para conquistar um cliente e, ao conquista-lo esperam manter ele pelo maior tempo possível. Kotler e Armstrong (2015) lembram que perder um cliente é muito diferente de perder uma venda, é perder as vendas de uma vida de relacionamento. Considerando que conquistar um cliente novo

custa cerca de cinco vezes mais que manter um cliente antigo, como também destacaram Kotler e Armstrong (2015), fica evidente a necessidade de as empresas terem projetos que visam a manutenção de sua carteira de clientes.

Apesar de ser mais barato manter um cliente a conquistar um novo, não é eficiente distribuir os recursos destinados à manutenção de clientes de forma igual entre eles. Lovelock e Wright (2001) constataram que a fidelidade do cliente está diretamente relacionada ao seu nível de satisfação. Há, portanto, os clientes que tem mais chances de parar e os que tem menos chances. Se uma empresa puder encontrar os clientes com maior probabilidade de parar, ela pode concentrar os esforços de manutenção nesse grupo.

Nesse contexto, a análise estatística surge como alternativa para auxiliar as empresas a identificar os clientes com maior risco de parar. Segundo Field (2009), fenômenos do mundo real podem ser explicados por modelos estatísticos construídos a partir de dados relacionados a tais fenômenos. Trazendo isso para a realidade de uma empresa, pode-se esperar que a perda de clientes possa ser explicada através da análise dos dados que a empresa possui sobre eles.

Dessa forma, esse estudo pretende realizar uma análise dos dados cadastrais dos clientes e de seu relacionamento com a empresa para construir um modelo estatístico e, com isso, responder a seguinte questão de pesquisa: Como prever que um cliente está prestes a deixar de ser cliente?

Para responder a esta pergunta, estabeleceu-se como objetivo principal propor um modelo estatístico capaz de prever a perda de clientes. Para atingir este objetivo, foram elencados os seguintes objetivos específicos: 1) coletar dados relativos ao cadastro de clientes e seu relacionamento com a empresa; 2) realizar uma análise estatística dos dados coletados a fim de encontrar correlação com a perda de clientes; 3) selecionar as variáveis mais importantes para o modelo; 4) desenvolver um modelo estatístico para prever a perda de clientes.

Este estudo é relevante pois visa possibilitar que uma empresa concentre os recursos destinados à manutenção de clientes naqueles com maior potencial de parar de comprar, tornando dessa forma, o uso dos recursos mais eficiente. Além disso, também deixa o processo de manutenção de clientes mais eficaz, pois permite que a empresa identifique aqueles que necessitam de mais atenção e possibilita que as ações sejam personalizadas, resultando em um índice de sucesso maior no processo de manutenção de clientes.

Inicialmente este trabalho abordará, em teoria, os principais conceitos envolvidos com a análise estatística relacionados ao objetivo deste estudo, seguida por uma pesquisa de trabalhos relacionados. Após, será explicada a metodologia usada no seu desenvolvimento e será apresentada a unidade de análise e os dados coletados. Na seção seguinte, apresentará um estudo estatístico das variáveis com o objetivo de encontrar possível correlação entre as variáveis independentes e a variável dependente. O trabalho segue com a seleção das variáveis e construção de diferentes modelos de regressão logística até encontrar o modelo com melhor assertividade. Por fim, serão discutidos os resultados e as conclusões obtidas com o estudo.

2. Fundamentação Teórica

2.1. A importância da retenção de clientes

Segundo Kotler e Keller (2012), as empresas que desejam expandir seus lucros e vendas devem investir na busca de novos clientes através de peças de propaganda, malas diretas, e-mails para clientes potenciais, participar de feiras, comprar cadastros de clientes, entre outras ações. Todas essas ações demandam tempo e recursos financeiros consideráveis. Kotler e Keller (2012) afirmam ainda que é excessivo o número de empresas que apresentam alta deserção de clientes e que, neste caso, conquistar novos clientes é como colocar água em um balde furado.

Evitar a perda de clientes também tem seu custo, é preciso manter relacionamentos que geram sua satisfação e investir em ações de fidelização. Porém, este custo é significativamente inferior se comparado ao custo de recuperar um cliente perdido. Kotler e Armstrong (2015) destacam que o custo de conquistar um cliente é cerca de cinco vezes maior que manter um cliente antigo.

Além disso, manter um cliente por um longo tempo possibilita um maior volume de vendas e assim, potencializa a possibilidade de a empresa obter lucro dessa relação. Uma empresa pode perder dinheiro em uma transação específica, mas se beneficiar de um relacionamento duradouro (KOTLER e ARMSTRONG, 2015). Isso justifica os investimentos na manutenção da carteira de clientes.

Os modelos estatísticos podem auxiliar as empresas a identificar os clientes com maior probabilidade de parar de comprar. Dessa forma, permitem que os investimentos sejam direcionados e personalizados, resultando em um melhor uso dos recursos. Nos tópicos a seguir serão apresentados alguns conceitos relacionados à análise estatística e criação de modelos estatísticos.

2.2. Análise estatística

A análise estatística é um componente da análise de dados. Normalmente a análise descritiva de um conjunto de dados é uma de suas primeiras etapas, justamente pelo fato de possibilitar ao pesquisador tomar conhecimento de informações importantes para as próximas etapas do processo. O objetivo da estatística descritiva é permitir ao pesquisador uma melhor compreensão sobre o comportamento dos dados através de tabelas, gráficos e medidas-resumo facilitando a identificação de tendências, variabilidade e valores atípicos (FÁVERO et al, 2009).

Além de conhecer o comportamento das variáveis em um conjunto de dados, é importante observar se existem relações entre essas variáveis e, se existirem, descobrir como elas ocorrem. Há diversas formas de mensurar as correlações entre as variáveis, entre elas podemos destacar a Correlação Bisserial por Ponto e o teste qui-quadrado.

A Correlação Bisserial por Ponto pode ser usada para encontrar a relação entre uma variável quantitativa e outra qualitativa dicotômica, ou seja, que tem apenas duas categorias. Para realizar o cálculo, deve-se identificar uma das categorias da variável qualitativa com o valor 0 e a outra com o valor 1, em seguida calcular a Correlação de Pearson (FIELD, 2009). O resultado da função será um valor entre -1 e 1 onde os valores abaixo de zero indicam correlação inversa e os valores positivos indicam

correlação direta. Quanto mais próximo de zero for o coeficiente, menor será a intensidade da correlação. Segundo Callegari-Jacques (2003), se o valor absoluto do coeficiente de correlação for menor que 0,30 há correlação linear fraca, se for entre 0,30 e menos de 0,60 há correlação linear moderada, se for entre 0,60 e menos de 0,90 há correlação linear forte e valores a partir de 0,90 representam uma correlação linear muito forte.

Para duas variáveis categóricas, pode-se verificar se existe associação através do teste qui-quadrado. Ele é um teste de hipóteses que tem como objetivo encontrar um valor de dispersão entre variáveis categóricas para determinar se existe associação. Segundo Field (2009), se o valor de significância é pequeno (menor que 0,05), rejeitamos a hipótese de que as variáveis são independentes e aceitamos a hipótese de que elas estão relacionadas.

Após conhecer o comportamento das variáveis e suas relações, pode-se seguir para a criação de um modelo estatístico. O tópico a seguir trata da Regressão Logística, uma das técnicas usadas na estatística para a criação de modelos preditivos.

2.3. Regressão logística

A análise de Regressão Logística é usada para prever o comportamento de uma variável dependente qualitativa através de um conjunto de variáveis independentes, podendo estas serem quantitativas ou qualitativas (FÁVERO et al., 2009). Pode-se prever a qual categoria uma observação pertence a partir de um conjunto de informações relacionadas. Ela pode ser usada, por exemplo, para prever se um crédito pode ou não ser liberado de acordo com o histórico de informações financeiras e outros dados do cliente.

A função logística assume valores entre 0 e 1 em uma curva “S”, indicando a probabilidade da ocorrência do evento (FÁVERO et al., 2009). O comportamento gráfico da função logística pode ser observado na Figura 1.

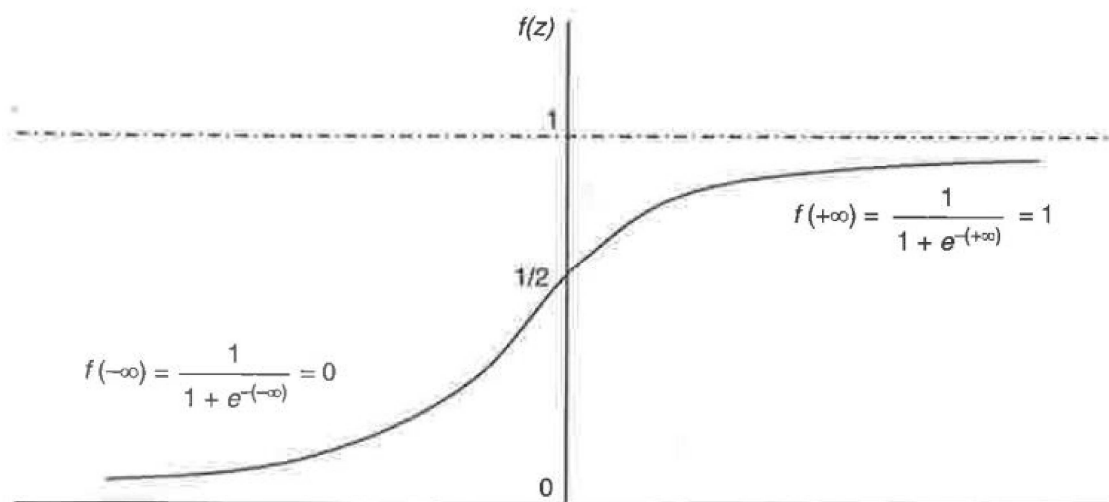


Figura 1. Função logística

De forma geral assume-se que: se a função logística retornar um valor abaixo de 0,5 consideramos o registro pertencente à classe 0 e, se retornar valor a partir de 0,5 consideramos pertencente à classe 1. Esse valor, que separa as duas classes é chamado

de valor de corte. Contudo, Fávero et al., (2009) afirmam que o ponto de corte pode assumir qualquer valor, de acordo com as necessidades do pesquisador. É comum ajustar esse valor quando as classes não estão balanceadas, ou seja, quando temos quantidades de registros diferentes em cada classe.

Após a criação do modelo preditivo, é necessário medir sua qualidade, ou seja, o quanto os resultados previstos se aproximam dos resultados observados. O tópico a seguir apresenta algumas métricas usadas para este fim.

2.4 Avaliação de modelos preditivos

Para avaliar um modelo preditivo existem algumas funções matemáticas que podem ser aplicadas. A escolha pela função e métricas adequadas dependem do tipo de modelo criado e também dos objetivos do estudo.

Para modelos de classificação é comum usar a matriz de confusão. Ela é uma matriz bidimensional que resume o desempenho da classificação. Numa das dimensões são indexados os valores observados e na outra os valores atribuídos pelo modelo de classificação (TING, 2017).

A acurácia é uma das métricas mais usadas para medir o desempenho geral de um modelo de classificação. Ela é uma medida que mostra a proporção dos resultados corretos em relação à população selecionada. No entanto, uma acurácia alta não implica em um bom modelo, a sensibilidade e a especificidade também precisam ser altas. A acurácia é suscetível a dados desbalanceados e, portanto, precisa ser interpretada com cautela (ZHU et al., 2010).

Para conjuntos de dados desbalanceados pode ser calculada a área sob a curva ROC. Ela representa a área sob uma curva formada pelo gráfico representando a sensibilidade e a especificidade (ZHU et al., 2010) A Figura 2 mostra um modelo do gráfico da curva ROC.

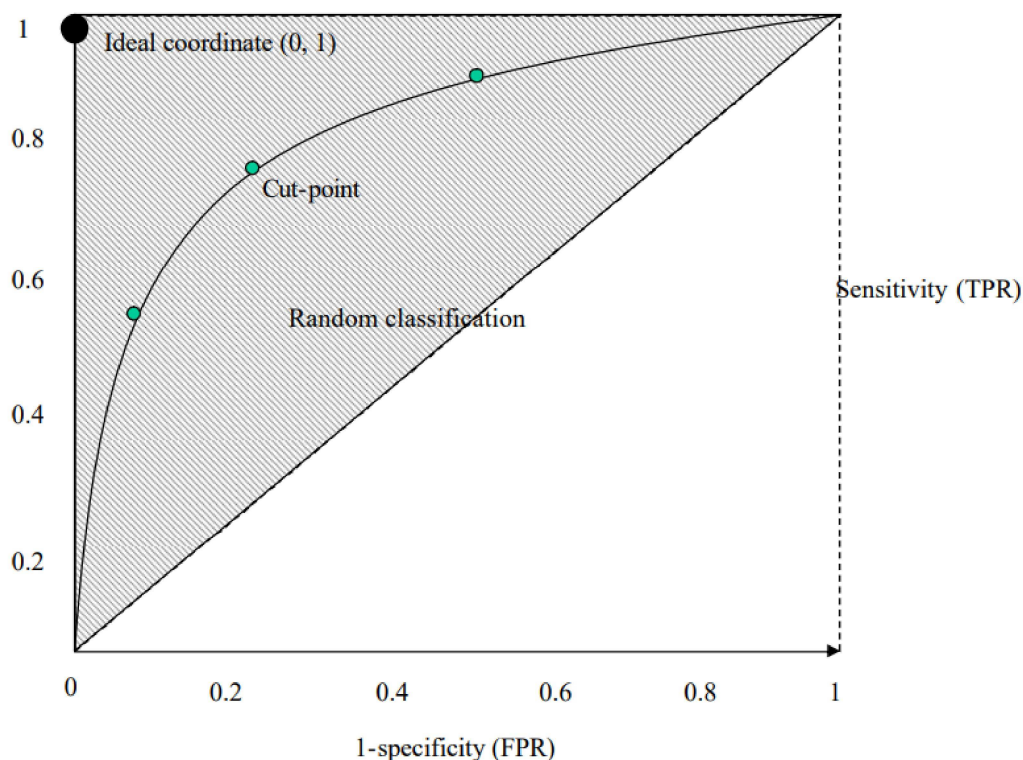


Figura 2. Gráfico da Curva ROC

Por fim, em modelos de regressão, pode-se medir o quanto a variância da variável dependente pode ser creditada às variáveis independentes, esse valor é representado pelo R^2 , também conhecido como coeficiente de determinação. Seu valor pode variar entre 0 e 1 indicando o percentual de explicação do modelo sobre os valores observados (FIELD, 2009).

2.5 Trabalhos relacionados

Na literatura, foram encontrados diversos trabalhos relacionados à predição de perda de clientes, na sua grande maioria aplicados à área de telecomunicações. Nesta seção, serão elencados os principais trabalhos encontrados, métodos usados e os resultados obtidos.

Uma das pesquisas mais importantes foi realizada por Schneider (2016), que fez uma revisão e classificação da literatura com ênfase em aprendizado de máquina para construção de modelos preditivos de perda de clientes. O autor analisou 82 artigos publicados entre os anos de 2000 e 2015. Ele constatou que 60% dos artigos publicados estão relacionados à área de telecomunicações, seguida de 16% nas áreas bancária e financeira e o restante dividido entre jornal, revisão, varejo, e-commerce, aviação, RH e marketing. Além de artigos que tratavam de mais de uma área de aplicação e outros sem nenhuma área específica. A pesquisa mostrou também que as técnicas de predição mais usadas foram Árvores de Decisão, Regressão Logística e Redes Neurais. O autor ainda constatou que a taxa de perda de clientes geralmente fica entre 2% e 3%, o que significa um alto nível de desbalanceamento das classes e representa um problema para os modelos preditivos.

Após o trabalho realizado por Scneider (2016), novas pesquisas foram publicadas, dentre as quais podemos destacar:

- Mishra e Reddy (2017), usaram Redes Neurais Convolucionais (CNN) para construir um modelo que obteve uma acurácia de 86,85%. A pesquisa foi realizada na área de telecomunicações.
- Li et al., (2016), propuseram um modelo de predição usando técnicas de agrupamento seguidas por uma classificação realizada usando o algoritmo Random Forest. Alcançaram uma acurácia de 80,42% e um recall de 52,43%. Esse estudo também foi realizado na área de telecomunicações.
- Wu e Meng (2016), usaram o algoritmo SMOTE para balancear as classes da base de dados e em seguida a submeteram ao algoritmo AdaBoost para realizar a classificação. O estudo foi aplicado na área de academias.
- Bertens et al., (2017), desenvolveram um modelo preditivo baseado em análise de sobrevivência na área de videogames.
- Pimentel e Goldschmidt (2018), usando uma abordagem diferenciada, propuseram um modelo preditivo que considera simultaneamente sentimentos presentes nas interações entre cliente e empresa com a ordem em que essas interações ocorrem ao longo do tempo. Foi usado um conjunto de dados com o histórico do conteúdo das interações que ocorreram por e-mail, chat, telefone, portal online, dentre outros. O trabalho comparou dois cenários, um com o atributo de sentimento e outro sem ele. No cenário com o atributo de sentimento os autores chegaram a uma acurácia 84,6% e no cenário sem o atributo de sentimento a acurácia ficou em 70,1%. Os autores concluíram, portanto, que a análise de sentimento colabora significativamente na melhoria da acurácia do modelo.

O presente trabalho se diferencia dos demais encontrados na literatura ao aplicar a técnica de Regressão Logística para previsão de perda de clientes em uma empresa da área atacadista. Não foram encontradas pesquisas anteriores nessa área. O público alvo das empresas atacadistas são os varejistas e não os consumidores finais, isso significa um contexto diferente de qualquer estudo já realizado.

3. Metodologia de Pesquisa

3.1. Enquadramento metodológico

A pesquisa se classifica como estudo documental de cunho quantitativo. Segundo Gil (1999), a pesquisa documental utiliza materiais primários, ou seja, que ainda não receberam tratamento analítico e que podem ser reelaborados de acordo com os objetivos da pesquisa. A pesquisa documental deve seguir as seguintes fases: a) determinação dos objetivos; b) elaboração do plano de trabalho; c) identificação das fontes; d) localização das fontes e obtenção do material; e) tratamento dos dados; f) confecção das fichas e redação do trabalho; g) construção lógica e redação do trabalho (GIL, 2007).

Por usar dados quantificáveis, esta pesquisa se classifica como quantitativa. Pesquisas com esse enfoque centram na objetividade, se caracterizam por apresentar

resultados quantificados e recorrem a linguagem matemática para descrever as causas de um fenômeno e as relações entre as variáveis (FONSECA, 2002).

Quanto à sua natureza, ela se classifica como pesquisa aplicada, pois visa gerar conhecimento para aplicação prática e resolução de um problema específico (GERHARDT e SILVEIRA, 2009).

Em relação aos objetivos, a pesquisa se classifica como descritiva. Esse tipo de pesquisa tem como finalidade a descrição das características de determinada população ou estabelecer relações entre as variáveis (GIL, 1999).

3.2. Unidade de análise

O setor atacadista no Brasil faturou 261,8 Bilhões de reais em 2018, o que representa 53,6% do mercado merceário brasileiro, que compreende produtos de uso comum das famílias, como alimentos, bebidas, higiene, limpeza e cuidados pessoais. Neste mesmo ano o setor empregou 384.374 pessoas e foram atendidos 1.070.967 pontos de venda (ABAD, 2019b).

Existem diversos modelos de atacado no Brasil, mas podemos classificar eles em quatro principais categorias (ABAD, 2019a):

- Atacadista generalista com entrega. Comercializa produtos da indústria sem vínculo de exclusividade nem território. Atende principalmente pequenos varejistas por meio de visitas de vendedores e entrega no estabelecimento.
- Atacadista de autosserviço. Compra e vende produtos da indústria sem vínculo de exclusividade. Nesta modalidade o cliente se desloca até a loja e realiza as compras. A entrega também fica por conta do próprio cliente.
- Distribuidor. Comercializa produtos da indústria com vínculo de exclusividade que pode ser por marca, empresa, região, categoria de produto ou canal. Atende os clientes no varejo por meio de visitas de vendedores e entrega as mercadorias no estabelecimento do cliente.
- Agente de serviços. Realiza algum serviço específico inerente à área atacadista, como representação comercial, operador logístico ou *merchandising*.

Este estudo foi realizado em uma empresa atacadista distribuidora localizada no Rio Grande do Sul. Ela atende cerca de 11 mil clientes, a maioria supermercados e restaurantes em mais de 300 cidade do seu estado.

3.3. Coleta dos dados

A amostra coletada para esta pesquisa se limita aos dados do mês de janeiro de 2019. São 11.149 registros de clientes que tiveram movimento naquele mês. Optou-se por esta amostragem por ser o mês mais recente com disponibilidade da informação sobre clientes que pararam.

Os dados foram coletados do banco de dados da empresa através de consulta SQL. Os dados incluem informações cadastrais dos clientes e seu relacionamento com a empresa. São 29 variáveis, sendo 28 independentes e uma dependente. As variáveis são apresentadas na Tabela 1.

Tabela 1. Apresentação das variáveis

Variável	Descrição
CIDADE_EBCT	Código de identificação da cidade do cliente.
TPESSOA	Tipo de pessoa (0 = Jurídica; 1 = Física).
RAMO	Ramo de atividade, conforme classificação da empresa.
TEMPO_EMPRESA	Tempo decorrido desde o cadastro do cliente.
VENDA_MARCAS	Valor da venda de produtos de marcas próprias no mês (em mil Reais).
VENDA_TERCEIROS	Valor da venda de produtos de terceiros no mês (em mil Reais).
VENDA_TOTAL	Valor da venda total no mês (em mil Reais).
BAIXOU_LIM_CRED	Se o limite de crédito do cliente foi baixado (0 = Não foi Baixado; 1 = Foi Baixado).
LIMITE_CREDITO	Valor do último limite de crédito (em mil Reais).
TIPO_CREDITO	Categoria de crédito conforme classificação da empresa.
QTD_PEDIDOS	Quantidade de pedidos realizados no mês.
VENDA_MARCAS_U3M	Valor médio da venda de produtos de marcas próprias nos últimos três meses (em mil Reais).
VENDA_TERCEIROS_U3M	Valor médio da venda de produtos de terceiros nos últimos três meses (em mil Reais).
VENDA_TOTAL_U3M	Valor médio da venda nos últimos três meses (em mil Reais).
MIX_MARCAS	Quantidade de produtos positivados das marcas próprias.
MIX_TERCEIROS	Quantidade de produtos de terceiros positivados.
TEM_CONTRATO	Se o cliente possui contrato de compra com a empresa (0 = Não Possui; 1 = Possui).
MESES_COM_VENDA	Quantidade de meses que o cliente comprou da empresa, considerando desde a data de seu cadastro.
USOU_CREDITO	Se o cliente fez algum pedido usando o crédito (0 = Não usou; 1 = Usou).
TEVE_RECLAMACAO	Se o cliente registrou alguma reclamação (0 = Não Registrou; 1 = Registrou).
TEVE_BLOQ_PED	Se o cliente teve algum pedido bloqueado (0 = Não Teve; 1 Teve Pedido Bloqueado).
TEVE_DESC_FIDEL	Se o cliente recebeu algum desconto referente ao programa de fidelização (0 = Não Recebeu; 1 = Recebeu).
TEVE_DESC_V155	Se o cliente recebeu algum desconto referente a verba de negociação do vendedor (0 = Não Recebeu; 1 = Recebeu).
TEVE_TAB_ESP	Se o cliente possui tabela de preços diferenciada (0 = Não Possui; 1 = Possui).
DIAS_SV	Quantidade de dias transcorridos desde a última compra.
IE_BAIXADA	Se a inscrição estadual do cliente foi baixada (0 = Não Foi; 1 = Foi Baixada).

PERIODO_MED	Intervalo médio entre as compras, considerando os últimos 12 meses.
PERIODO_STD	Desvio padrão do intervalo médio entre as compras, considerando os últimos 12 meses.
PAROU	Se o cliente parou de comprar (0 = Não Parou; 1 = Parou).

4. Análise e Transformação dos Dados

Do total de registros da base de dados, 10.755 são de clientes que não pararam de comprar e os outros 394 são de clientes que pararam. O que representa um percentual de perda de clientes de 3,53. Tem-se dessa forma, um conjunto de dados desbalanceado.

Nas seções seguintes serão apresentadas e analisadas as principais medidas das variáveis independentes e sua correlação com a variável dependente.

4.1. Variáveis quantitativas

A Tabela 2 apresenta algumas medidas descritivas das variáveis quantitativas.

Tabela 2. Medidas descritivas

Variável	Mín.	Máx.	Méd.	Desvio Padrão	Assimetria	K-S (valor-p)
TEMPO_EMPRESA	0,0	29,0	9,5	7,5	0,533	0,000
VENDA_MARCAS	0,0	611,6	1,2	7,5	55,435	0,000
VENDA_TERCEIROS	0,0	63,5	1,0	1,6	10,504	0,000
VENDA_TOTAL	0,0	611,6	2,2	7,8	49,199	0,000
LIMITE_CREDITO	0,0	800,0	3,6	12,9	34,725	0,000
QTD_PEDIDOS	0,0	22,0	2,4	1,6	1,685	0,000
VENDA_MARCAS_U3M	0,0	616,5	1,3	7,5	54,880	0,000
VENDA_TERCEIROS_U3M	0,0	28,2	0,9	1,3	5,983	0,000
VENDA_TOTAL_U3M	0,0	616,5	2,3	7,8	49,416	0,000
MIX_MARCAS	0,0	159,0	15,7	16,6	2,264	0,000
MIX_TERCEIROS	0,0	272,0	15,7	15,9	2,887	0,000
MESES_COM_VENDA	1,0	265,0	93,4	80,2	0,769	0,000
DIAS_SV	0,0	29,0	8,1	7,6	0,915	0,000
PERIODO_MED	0,0	320,0	18,7	16,7	5,206	0,000
PERIODO_STD	0,0	226,3	8,7	13,1	6,753	0,000

Percebe-se que todas as variáveis possuem o valor médio mais próximo do valor mínimo do que do valor máximo. Pode-se observar também que todas possuem distribuição assimétrica positiva. Além disso, o teste de Kolmogorov-Smirnov apresentou valor-p < 0,05 para todas as variáveis. Segundo Field (2009), valores de significância abaixo de 0,05 no teste de Kolmogorov-Smirnov indicam um desvio da normalidade. Com isso, pode-se concluir que há distribuição não-normal para todo conjunto de variáveis quantitativas.

Uma distribuição não-normal indica que há valores atípicos, que podem introduzir tendenciosidades no modelo. Nestes casos, pode-se realizar uma transformação nos dados de forma a reduzir o impacto desses valores (FIELD, 2009).

Para este estudo, foi realizada uma transformação logarítmica nas variáveis quantitativas. Essa transformação é adequada quando há assimetria positiva, pois o logaritmo achata a cauda direita da distribuição, reduzindo a assimetria (FIELD, 2009).

Após a transformação, foi realizado o teste de Correlação Bisserial por Ponto entre as variáveis independentes e a variável dependente. A Tabela 3 mostra os valores encontrados.

Tabela 3. Correlação das variáveis independentes com a dependente

Variável	Correlação com PAROU	Valor-p
TEMPO_EMPRESA	-0,124	0,000
VENDA_MARCAS	-0,113	0,000
VENDA_TERCEIROSS	-0,094	0,000
VENDA_TOTAL	-0,129	0,000
LIMITE_CREDITO	-0,126	0,000
QTD_PEDIDOS	-0,146	0,000
VENDA_MARCAS_U3M	-0,117	0,000
VENDA_TERCEIROS_U3M	-0,107	0,000
VENDA_TOTAL_U3M	-0,143	0,000
MIX_MARCAS	-0,159	0,000
MIX_TERCEIROS	-0,088	0,000
MESES_COM_VENDA	-0,174	0,000
DIAS_SV	0,114	0,000
PERIODO_MED	-0,002	0,833
PERIODO_STD	0,010	0,293

O resultado do teste mostra que as variáveis PERIODO_MED e PERIODO_STD não possuem correlação significativa com a variável dependente, pois possuem valor-p $> 0,05$ e, dessa forma, são candidatas a não entrarem no modelo. Entre as demais, a variável DIAS_SV é a única que possui correlação positiva, o que indica que conforme aumenta a quantidade de dias que o cliente está sem comprar aumentam as chances de que não compre mais. Todas as demais variáveis possuem correlação negativa, o que significa que quando seu valor aumenta, a probabilidade de o cliente parar diminui.

Também pode-se observar que as variáveis MESES_COM_VENDA e MIX_MARCAS são as que possuem maior correlação com a variável dependente. Isso significa que os clientes que compram há mais tempo e que compram um número maior de produtos distintos das marcas próprias da empresa tem menor chance de parar de comprar.

4.2. Variáveis qualitativas

Para a análise das variáveis qualitativas foi feita uma tabulação cruzada (ou tabela de contingência) para detalhar a distribuição de cada variável qualitativa independente entre as categorias da variável dependente. Também foi realizado o teste qui-quadrado para verificar a existência de alguma relação entre as variáveis independentes e a dependente. A Tabela 4 mostra os valores da tabulação cruzada.

Tabela 4. Tabulação cruzada

Variável	Categoria	PAROU		Qty. de Registros
		% Não	% Sim	
CIDADE_EBCT *	71986	95,6	4,4	1.321
	71242	97,7	2,3	432
	68020	98,3	1,7	417
	67733	96,3	3,7	296
	72842	96,5	3,5	259
TPESSOA	0	96,8	3,2	10.443
	1	92,2	7,8	706
RAMO *	3	97,1	2,9	3.692
	21	96,4	3,6	1.520
	1	98,6	1,4	1.367
	20	92,2	7,8	708
	2	97,1	2,9	652
BAIXOU_LIM_CRED	0	96,8	3,2	10.996
	1	71,9	28,1	153
TIPO_CREDITO *	1	97,8	2,2	6.900
	33	91,3	8,7	841
	12	98,8	1,2	602
	39	91,2	8,8	532
	41	92,8	7,2	528
TEM_CONTRATO	0	96,3	3,7	10.564
	1	99,8	0,2	585
USOU_CREDITO	0	93,2	6,8	3.391
	1	97,9	2,1	7.758
TEVE_RECLAMACAO	0	96,6	3,4	10.759
	1	91,8	8,2	390
TEVE_BLOQ_PED	0	96,5	3,5	11.005
	1	94,4	5,6	144
TEVE_DESC_FIDEL	0	96,3	3,7	10.526
	1	99,5	0,5	623
TEVE_DESC_V155	0	95,6	4,4	6.488
	1	97,7	2,3	4.661
TEVE_TAB_ESP	0	96,3	3,7	10.627
	1	100	0	522

IE_BAIXADA	0	96,5	3,5	11.088
	1	86,9	13,1	61

* Para facilitar a visualização foram exibidos somente as cinco categorias mais frequentes.

A Tabela 5 apresenta o resultado do teste qui-quadrado com as variáveis categóricas da base de dados da pesquisa.

Tabela 5. Resultado do teste qui-quadrado

Variável	Valor	Valor-p
CIDADE_EBCT	326,182	0,216
TPESSOA	40,056	0,000
RAMO	276,377	0,000
BAIXOU_LIM_CRED	274,72	0,000
TIPO_CREDITO	374,202	0,000
TEM_CONTRATO	20,483	0,000
USOU_CREDITO	148,143	0,000
TEVE_RECLAMACAO	25,867	0,000
TEVE_BLOQ_PED	1,749	0,186
TEVE_DESC_FIDEL	18,053	0,000
TEVE_DESC_V155	36,027	0,000
TEVE_TAB_ESP	20,062	0,000
IE_BAXADA	15,515	0,000

Ao analisar os dados da tabela cruzada (Tabela 4) e o resultado do teste qui-quadrado (Tabela 5) podemos concluir o seguinte:

- Há uma certa diferença no percentual de clientes que pararam entre uma cidade e outra. Por exemplo, na cidade código 71986, 4,4% dos clientes pararam de comprar, já na cidade código 68020 foram apenas 1,7%. Porém, o teste qui-quadrado mostra que essa variação não é significativa, pois o valor-p ficou acima de 0,05, o que indica que a diferença ocorre ao acaso e não há associação.
- Há maior probabilidade de perder um cliente pessoa física em relação aos clientes pessoa jurídica. O qui-quadrado significativo mostra que a possibilidade de o cliente parar de comprar está associada ao seu tipo de pessoa.
- Há diferenças significativas de perda de clientes entre diferentes ramos de atividade. No ramo de atividade código 1 houve perda de apenas 1,4% dos clientes, já no ramo de atividade código 20 houve perda de 7,8% dos clientes. O teste qui-quadrado mostrou um valor-p menor que 0,05, o que indica que há correlação entre o ramo de atividade e a perda de clientes.
- A perda de clientes está fortemente relacionada à redução do seu limite de crédito. Entre os clientes que tiveram redução do limite a perda foi de 28,1%, já entre os que não tiveram redução a perda foi de apenas 3,2%. O teste qui-quadrado mostra que a relação é significativa, pois o valor-p ficou abaixo de 0,05.

- O tipo de crédito tem associação com a perda de clientes. No tipo de crédito código 12 houve perda de apenas 1,2% dos clientes, já no tipo de crédito código 39 houve perda de 8,8%. O valor-p menor que 0,05 no teste qui-quadrado confirma que essa associação não ocorre por acaso e que há relação entre as variáveis.
- Os clientes que possuem contrato de compra com a empresa têm baixa probabilidade de parar de comprar, apenas 0,5% pararam. Entre os que não possuem contrato foram 3,7% que pararam. O teste qui-quadrado confirma que essa associação é significativa com valor-p menor que 0,05.
- Entre os clientes que usaram crédito, 2,1% pararam de comprar, já entre os que não usaram crédito, 6,8% pararam. O teste qui-quadrado mostra que há associação com a perda de clientes, pois o valor-p ficou abaixo de 0,05.
- Para o grupo de clientes que registraram alguma reclamação, houve perda de 8,2%, por outro lado, entre os que não registraram reclamações houve perda de 3,4%. O teste qui-quadrado mostra que há associação significativa entre o registro de reclamações e a perda do cliente.
- 5,6% dos clientes que tiveram algum pedido bloqueado pararam de comprar e 3,5% dos que não tiveram pedido bloqueado pararam. Apesar da diferença, o teste qui-quadrado com valor-p maior que 0,05 mostra que não há associação entre as variáveis e que a variação ocorre ao acaso.
- Entre os clientes que receberam desconto do programa de fidelização apenas 0,5% pararam de comprar. Já entre os que não receberam esse desconto houve perda de 3,7%. A associação é significativa e comprovada com o valor-p menor que 0,05 no teste qui-quadrado.
- Conceder um desconto de negociação pode diminuir a probabilidade de perder o cliente. Entre os que receberam um desconto, 2,3% pararam de comprar. Já entre os que não receberam, 4,4% pararam. O teste qui-quadrado mostra que essa variação não ocorre ao acaso e que há associação entre as variáveis.
- Nenhum cliente com tabela de preços especial parou de comprar, já entre os que usam tabela de preços normal houve 3,7% de perda. Essa variação é significativa, pois o valor-p do teste qui-quadrado ficou abaixo de 0,05.
- Entre os clientes que tiveram a Inscrição Estadual baixada, 13,1% pararam de comprar. Entre os demais clientes apenas 3,5% pararam. O teste qui-quadrado mostra que há associação entre as variáveis e que a variação não ocorre ao acaso. O valor-p ficou abaixo de 0,05.

5. Descrição e Análise dos Resultados

Após a análise e ajustes nos dados foi aplicada a técnica de Regressão Logística com o objetivo de criar um modelo matemático capaz de prever se um cliente vai parar de comprar. A Regressão Logística é uma técnica estatística utilizada para prever o valor de uma variável dependente categórica através de variáveis independentes métricas ou não métricas (FÁVERO et al., 2009). Essa técnica é adequada para este estudo, pois o

objetivo é prever se um cliente vai parar de comprar. Existe, portanto, uma variável dependente categórica que pode ser classificada em duas categorias: sim ou não.

Inicialmente foi realizada uma Regressão Logística com todas as variáveis independentes para, em seguida, realizar a análise dos resultados obtidos e fazer os ajustes necessários no modelo.

Foi usado como ponto de corte o valor 0,15. Segundo Fávero et al. (2009), o ideal é usar amostra equilibrada, ou seja, 50% dos valores da amostra devem estar na classe 0 e 50% na classe 1 e um ponto de corte de 0,5. Entretanto, quando se possui uma amostra não balanceada pode-se usar o ponto de corte ajustado de acordo com a proporção das classes na amostra. Fávero et al., (2009) ainda afirmam que nada impede que o analista/pesquisador use outro ponto de corte, conforme o risco que queira assumir em relação aos erros na classificação.

O método de Regressão Logística usado foi o de entrada forçada. Nesse método, todas as variáveis independentes são colocadas no modelo em um único bloco e as estimativas dos parâmetros são calculadas para o bloco (FIELD, 2009). A Tabela 6 mostra o resultado da classificação.

Tabela 6. Tabela de classificação do Modelo 1

		Previsto		
		PAROU		Porcentagem correta
		0	1	
PAROU	0	10.317	427	96,0
	1	176	218	55,3
Porcentagem global				94,6

O modelo inicial obteve um percentual de acerto global de 94,6%, sendo 55,3% de acerto entre os clientes que pararam e 96% de acerto entre os clientes que não pararam. O R quadrado de Cox & Snell ficou em 0,099 e o R quadrado de Nagelkerke ficou em 0,375. Já a Curva ROC (*Receiver Operating Characteristic*) apresentou uma área de 0,757.

Os valores de R quadrado ficaram baixos, o que indica que as variáveis independentes têm pouco poder de explicação sobre a variável dependente. Por outro lado, o valor da área sob a Curva ROC indica que o modelo possui um poder discriminatório bom. Segundo Fávero et al. (2009), valores menores ou iguais a 0,5 de área abaixo da Curva ROC indicam que o modelo não possui poder de discriminação entre as classes, valores entre 0,7 e 0,8 indicam que o poder de discriminação é aceitável e valores maiores que 0,8 indicam um poder de discriminação excelente.

A Tabela 7 mostra os coeficientes das variáveis independentes e sua significância. As variáveis CIDADE_EBCT, RAMO E TIPO_CREDITO foram removidas da tabela para reduzir seu tamanho. Elas possuem centenas de categorias e deixariam a tabela muito extensa. Em análise preliminar observou-se que nenhuma delas é significativa para o modelo, tendo valor-p acima de 0,05 para todas as categorias.

Tabela 7. Variáveis no Modelo 1

Variável	B	S.E.	Wald	df	Sig.	Exp(B)
TPESSOA	-18,894	25493,342	0,000	1	0,999	0,000
BAIXOU_LIM_CRED	-1,366	0,280	23,709	1	0,000	0,255
TEM_CONTRATO	-0,073	1,211	0,004	1	0,952	0,930
USOU_CREDITO	0,161	0,330	0,237	1	0,627	1,174
TEVE_RECLAMACAO	-1,301	0,260	25,045	1	0,000	0,272
TEVE_BLOQ_PED	0,249	0,506	0,242	1	0,623	1,283
TEVE_DESC_FIDEL	-0,007	0,801	0,000	1	0,993	0,993
TEVE_DESC_V155	0,006	0,143	0,002	1	0,964	1,006
TEVE_TAB_ESP	16,528	1487,753	0,000	1	0,991	15070517
IE_BAIXADA	-1,157	0,488	5,634	1	0,018	0,314
TEMPO_EMPRESA	1,178	0,304	14,988	1	0,000	3,247
VENDA_MARCAS	0,377	2,348	0,026	1	0,873	1,458
VENDA_TERCEIROS	0,625	2,472	0,064	1	0,800	1,869
VENDA_TOTAL	-1,574	2,981	0,279	1	0,597	0,207
LIMITE_CREDITO	0,223	0,519	0,184	1	0,668	1,250
QTD_PEDIDOS	-3,624	0,755	23,029	1	0,000	0,027
VENDA_MARCAS_U3M	4,390	2,235	3,857	1	0,050	80,622
VENDA_TERCEIROS_U3M	5,364	2,303	5,426	1	0,020	213,650
VENDA_TOTAL_U3M	-7,568	2,800	7,304	1	0,007	0,001
MIX_MARCAS	-0,468	0,241	3,753	1	0,053	0,627
MIX_TERCEIROS	-0,147	0,247	0,353	1	0,552	0,864
MESES_COM_VENDA	-1,681	0,242	48,137	1	0,000	0,186
DIAS_SV	0,590	0,183	10,460	1	0,001	1,805
PERIODO_MED	0,112	0,184	0,369	1	0,544	1,118
PERIODO_STD	0,098	0,188	0,270	1	0,603	1,103
CONSTANTE	-38,940	52541,374	0,000	1	0,999	0,000

Pode-se observar na Tabela 7 que as variáveis CIDADE_EBCT, TEVE_BLOQ_PED, PERIODO_MED e PERIODO_STD (CIDADE_EBCT conforme comentado no parágrafo anterior) possuem valor-p acima de 0,05, o que significa que não são significativos para o modelo. Isso já era esperado para estas quatro variáveis porque durante a análise preliminar já foi visto que não havia correlação significativa delas com a variável dependente.

A Tabela 7 mostra ainda que as variáveis TPESSOA, TEM_CONTRATO, USOU_CREDITO, TEVE_DESC_FIDEL, TEVE_DESC_V155, TEVE_TAB_ESP, VENDA_MARCAS, VENDA_TERCEIROS, VENDA_TOTAL, LIMITE_CREDITO, MIX_MARCAS e MIX_TERCEIROS também possuem valor-p acima de 0,05 indicando que não são significativas para o modelo.

Após essa verificação, foi realizada uma nova Regressão Logística excluindo as variáveis independentes não significativas. O segundo modelo foi construído somente com as variáveis BAIXOU_LIM_CRED, TEVE_RECLAMACAO, IE_BAIXADA, TEMPO_EMPRESA, QTD_PEDIDOS, VENDA_MARCAS_U3M,

VENDA_TERCEIROS_U3M, VENDA_TOTAL_U3M, MESES_COM_VENDA e DIAS_SV. Foi mantido o parâmetro de ponto de corte com valor em 0,15 e o método de entrada forçada. A Tabela 8 mostra o resultado da classificação por este novo modelo.

Tabela 8. Tabela de classificação do Modelo 2

		Previsto		
		PAROU		Porcentagem correta
		0	1	
PAROU	0	10.454	301	97,2
	1	272	122	31,0
Porcentagem global				94,9

Nota-se que, apesar das variáveis terem sido consideradas não significativas no primeiro modelo, a remoção delas causou uma piora considerável na acurácia da predição. A assertividade global subiu para 94,9%, mas a diferença mais significativa está na assertividade entre os clientes que pararam. Ela caiu de 55,3% para 31%. Essa é uma questão importante, pois o objetivo é justamente encontrar os clientes que vão parar de comprar. A área abaixo da Curva ROC também teve resultado pior no segundo modelo. Seu valor ficou em 0,641 ante 0,757 no primeiro modelo.

Dessa forma, foi realizada uma terceira Regressão Logística usando o método *Backward Wald*, onde inicialmente todas as variáveis são incluídas no modelo e em seguida removidas uma a uma de acordo com o menor valor no teste *Wald* (FIELD, 2009).

Neste modelo permaneceram as variáveis CIDADE_EBCT, BAIXOU_LIM_CRED, TIPO_CREDITO, TEM_CONTRATO, TEVE_RECLAMACAO, TEMPO_EMPRESA, QTD_PEDIDOS, VENDA_MARCAS_U3M, VENDA_TERCEIROS_U3M, VENDA_TOTAL_U3M, MIX_MARCAS, MESES_COM_VENDA E DIAS_SV.

A Tabela 9 mostra a tabela de classificação desse modelo.

Tabela 9. Tabela de classificação do Modelo 3

		Previsto		
		PAROU		Porcentagem correta
		0	1	
PAROU	0	10.344	400	96,3
	1	205	189	48,0
Porcentagem global				94,6

O percentual de acerto global de 94,6%, sendo 48% de acerto entre os clientes que pararam e 96,3% de acerto entre os clientes que não pararam. O R quadrado de Cox & Snell ficou em 0,092 e o R quadrado de Nagelkerke ficou em 0,348. Já a Curva apresentou uma área de 0,651.

A Tabela 10 mostra um resumo das medidas de avaliação dos três modelos testados.

Tabela 10. Medidas dos modelos

Medida	Modelo 1	Modelo 2	Modelo 3
R quadrado de Cox & Snell	0,099	0,058	0,092
R quadrado de Nagelkerke	0,375	0,222	0,348
Área abaixo da curva ROC	0,757	0,641	0,651
Percentual de acerto entre os que pararam (especificidade)	55,3	31,0	48,0
Percentual de acerto entre os que não pararam (sensibilidade)	96,0	97,2	96,3
Percentual de acerto global	94,6	94,9	94,6

Analisando os resultados obtidos, sugere-se o uso do Modelo 3 de Regressão Logística por usar apenas as variáveis consideradas significativas e por apresentar uma assertividade boa. Dessa forma, o estudo segue com a avaliação do comportamento das variáveis usadas no Modelo 3. O Apêndice A mostra as estatísticas das variáveis que permaneceram nesse modelo.

Observando os resultados do Modelo 3, pode-se apontar que:

- Há grande diferença na probabilidade de um cliente parar de comprar de acordo com sua cidade. Por exemplo, os clientes da cidade código EBCT 66346 ficaram com $\text{Exp}(B)$ em 5,739. O que quer dizer que tem 5,74 vezes mais chances de parar se comparado aos clientes da cidade base (código EBCT 108376). Já os clientes da cidade código EBCT 66320 ficaram com $\text{Exp}(B)$ em 0,175. O que significa que tem 82,5% menos chances de parar em relação aos clientes da cidade base, pois $1 - 0,175 = 0,825$.
- A redução do limite de crédito está associada a um aumento nas chances de o cliente parar de comprar. Não reduzir o limite de crédito representa uma redução de 72% de chances de o cliente parar, pois $1 - 0,28 = 0,72$.
- O tipo de crédito está associado a uma grande diferença na probabilidade de o cliente parar de comprar. Por exemplo, os clientes com tipo de crédito código 1 tem 348000 vezes mais chances de parar se comparados aos clientes do tipo de crédito base, código 55. Já os clientes com tipo de crédito código 9 tem 97% menos chances de parar se comparados aos clientes do tipo de crédito base, pois $1 - 0,03 = 0,97$.
- Assinar contrato de venda com o cliente está associado com uma redução significativa de chances de o cliente parar de comprar. Clientes sem contrato tem 6,38 vezes mais chances de parar.
- O registro de reclamação pelo cliente com a empresa está associado a um aumento nas chances de ele parar de comprar. Clientes sem reclamação tem uma probabilidade de parar 72% menor, pois $1 - 0,28 = 0,72$.
- O aumento do tempo em que o cliente compra da empresa, do volume de vendas de produtos de marcas próprias e de terceiros nos últimos 3 meses e a quantidade

de dias que está sem comprar estão associados a um aumento nas chances de o cliente parar de comprar.

- O aumento da quantidade de pedidos efetuados, do valor total de venda nos últimos 3 meses, da quantidade de produtos distintos de marcas próprias e da quantidade de meses que o cliente já comprou da empresa está associada a uma redução nas chances de o cliente parar de comprar.

O cálculo da predição da possibilidade de o cliente parar de comprar é realizado da seguinte forma:

$$Y = 1 / (1 + \exp(-30,5 - 1,273 * \text{BAIXOU_LIM_CRED}(0) + 1,853 * \text{TEM_CONTRATO}(0) - 1,285 * \text{TEVE_RECLAMAÇÃO}(0) + 1,038 * \text{TEMPO_EMPRESA} - 4,186 * \text{QTD_PEDIDOS} + 4,086 * \text{VENDA_MARCAS_U3M} + 4,535 * \text{VENDA_TERCEIROS_U3M} - 6,912 * \text{VENDA_TOTAL_U3M} - 0,639 * \text{MIX_MARCAS} - 1,458 * \text{MESES_COM_VENDA} + 0,606 * \text{DIAS_SV} + 0,405 * \text{CIDADE_EBCT}(66303) - 1,739 * \text{CIDADE_EBCT}(66320) + \dots K * \text{CIDADE_EBCT}(K) + 12,76 * \text{TIPO_CREDITO}(1) - 7,051 * \text{TIPO_CREDITO}(2) + \dots N * \text{TIPO_CREDITO}(N)))$$

Onde K representa as cidades e seus respectivos valor B conforme listado na tabela do Apêndice A, assim como N representa os tipos de crédito e seus respectivos valores B conforme também listado na tabela do Apêndice A. O Y representa a probabilidade de o cliente parar de comprar, considerando o ponto de corte em 0,15. Ou seja, valores de Y abaixo de 0,15 significa que o cliente não deve parar de comprar e valores a partir de 0,15 significa que o cliente deve parar.

6. Discussão

Os dados observados e analisados nas seções anteriores mostram que é possível prever com boa assertividade quais clientes estão prestes a parar de comprar, mas a pesquisa também mostrou que há alguns detalhes importantes que precisam ser considerados pela empresa objeto deste estudo caso implante este processo. Nessa seção serão discutidos os achados importantes e os conhecimentos obtidos por esta pesquisa.

A primeira questão que chama a atenção é uma divergência em duas métricas usadas para avaliar o resultado da regressão logística realizada. Enquanto o R quadrado apresentou valores baixos, a área sob a curva ROC apresentou um valor bom. O valor do R quadrado baixo significa que as variáveis independentes tem baixo poder de explicação sobre a variável dependente. Em decorrência disso, poderia se imaginar que a assertividade do modelo ficaria baixa. Mas, analisado a acurácia geral e o valor da área sob a curva ROC, percebe-se que o resultado foi satisfatório.

Outra questão importante está relacionada à escolha do modelo. É preciso tomar cuidado para não ser enganado ao avaliar o índice de assertividade geral que, apesar de ter ficado alto, não necessariamente significa que o modelo ficou bom. Isso acontece porque a base de dados da pesquisa está desbalanceada. Apenas 3,53% dos clientes pararam de comprar, o que significa que, se o modelo prevesse que todos os clientes continuariam comprando teria uma assertividade geral de 96,47%. Um valor importante a ser observado é a especificidade, ou seja, o percentual de acertos entre os clientes que pararam de comprar, pois o objetivo é exatamente descobrir quem são esses clientes. O

modelo recomendado (Modelo 3) teve uma assertividade de 48% nesse quesito. Isso quer dizer que a empresa poderia conhecer com antecedência quase metade dos clientes que pararam de comprar e realizar ações para tentar evitar essas perdas. No primeiro modelo esse índice foi ligeiramente melhor, chegando a 55,3%, porém esse modelo usou todas as 28 variáveis independentes, o que o deixa mais complexo. Já o terceiro modelo é mais enxuto, foram usadas apenas 13 variáveis independentes. A escolha pelo modelo a ser adotado dependerá da política e objetivos da empresa.

Uma implicação importante ao colocar em prática um processo de previsão de perda de clientes como o sugerido neste trabalho está no fato de que a assertividade esperada pelo modelo não poderá mais ser medida, isso porque o índice de perda de clientes poderá ser afetado pelas ações de manutenção realizadas pela empresa após obter os resultados da predição pela regressão logística. Podemos tomar como exemplo os dados usado para esta pesquisa. A regressão foi realizada com dados do mês de janeiro de 2019 para prever quais clientes não comprariam mais no mês de fevereiro. Na prática, essa regressão seria realizada no início do mês de fevereiro, mas os dados reais de perda de clientes para verificar a assertividade do modelo só estariam disponíveis no final de fevereiro. No entanto, ao longo do mês a empresa realiza ações no sentido de manter os clientes e, se estas ações forem eficientes, o número de clientes que vão parar de comprar será menor que o previsto pela regressão. Dessa forma, fica comprometida a verificação da assertividade do modelo. Sugere-se que a empresa use outra forma de acompanhar a eficiência da previsão de perda de clientes, como por exemplo, a evolução da taxa de perda de clientes. Se no mês de janeiro foi de 3,53%, como vimos anteriormente, espera-se que nos meses seguintes esse valor vá gradativamente sendo reduzido.

Em comparação aos trabalhos relacionados, o modelo proposto por este estudo obteve uma assertividade superior. Mishra e Reddy (2017), obtiveram acurácia de 86,85%, Li et al., (2016), alcançaram acurácia de 80,42% e Pimentel e Goldschmidt (2018), conseguiram 84,6% de acurácia. Porém, em muitos estudos foram usadas métricas de avaliação do modelo diferentes, o que pode inviabilizar a comparação dos resultados. É preciso considerar ainda, que não foram encontrados estudos anteriores aplicados a empresas atacadistas, o que também prejudica a comparação dos resultados.

7. Conclusão

Este estudo teve como objetivo criar um modelo estatístico para prever a perda de clientes, visto que o custo para recuperar um cliente é consideravelmente mais alto que manter ele. Ao tomar conhecimento sobre quais clientes estão prestes a parar de comprar os executivos das empresas podem concentrar os esforços de manutenção nesse grupo, evitando assim o desperdício de recursos com clientes que não correm risco de parar.

Foi realizado um estudo em uma empresa do ramo atacadista distribuidor, da qual se obteve uma base de dados de 11.149 registros com 28 variáveis independentes e uma variável dependente. Os dados envolveram informações cadastrais dos clientes e informações sobre seu relacionamento com a empresa. Após a análise estatística e aplicação de três diferentes modelos de regressão logística, foi sugerido um modelo com 13 variáveis independentes que obteve um índice de acertividade geral de 94,6% e uma

assertividade entre os clientes que pararam (especificidade) de 48%. O modelo proposto, portanto, atinge o objetivo desse estudo.

Os resultados alcançados aplicam-se especificamente à empresa alvo deste estudo, pois os dados são específicos dela. Mas o trabalho foi conduzido de forma que o modelo possa ser facilmente adaptado em uma grande quantidade de empresas. Em cada empresa poderão haver variáveis diferentes e conseqüentemente os resultados também poderão variar. Cabe a cada empresa selecionar suas variáveis e realizar os ajustes no seu próprio modelo.

Para trabalhos futuros sugere-se a aplicação de diferentes técnicas de classificação além da Regressão Logística. Podem ser usados algoritmos de aprendizado de máquina como redes neurais, árvores de decisão, naive bayes entre outros. Cada algoritmo possui suas próprias características e podem apresentar resultados distintos dos encontrados nesta pesquisa.

Referências

- ABAD, Associação Brasileira de Atacadistas Distribuidores. **Setor no Brasil**. 2019a. Disponível em: <https://abad.com.br/servicos/dados-do-setor/modelos-de-negocio>. Acessado em: 30/09/2019
- _____. **RANKING ABAD/NIELSEN 2019 (ANO BASE 2018)**. 2019b. Disponível em: <https://abad.com.br/servicos/dados-do-setor/ranking>. Acessado em: 30/09/2019.
- BERTENS, Paul; GUITART, Anna; PERIÁÑEZ, África. **Games and big data: A scalable multi-dimensional churn prediction model**. New York, USA. IEEE, 2017.
- CALLEGARI-JACQUES, Sidia M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artemed, 2003.
- FABBRO, Verónica; DEROCHE, Ariel; BASSO, Diego; POLLO-CATTANEO, Florencia. **Analysis of algorithms applied to Churn Analysis**. Curitiba, Braz. J. of Develop., 2019.
- FÁVERO, Luiz Paulo; BELFIORE, Patrícia; SILVA, Fabiana Lopes da; CHAN, Betty Lilian. **Análise de dados – Modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.
- FIELD, Andy. **Descobrimo a estatística usando o SPSS**. Porto Alegre: Artemed, 2009.
- FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002.
- GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999.
- _____. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 2007.
- KOTLER, Philip; ARMSTRONG, Gary. **Princípios de Marketing**. São Paulo: Pearson, 2015.

- KOTLER, Philip; KELLER, Kevin Lane. **Administração de Marketing**. São Paulo: Pearson, 2012.
- LI, Hui; YANG, Deliang; YANG, Lingling; YAOLU; LIN Xiaola. **Supervised Massive Data Analysis for Telecommunication Customer Churn Prediction**. Atanta, USA. IEEE, 2016.
- LOVELOCK, Christopher; WRIGHT, Lauren. **Serviços: Marketing e Gestão**. São Paulo: Saraiva, 2001.
- MISHRA, Abinash; REDDY, U. Srinivasulu. **A novel approach for churn prediction using deep learning**. Coimbatore, India. IEEE, 2017.
- PIMENTEL, Thiago P.; GOLDSCHMIDT, Ronaldo R.. **A Sequential Pattern Detection and Sentiment Analysis Combined Approach to the Churn Prediction Problem in Client Relationship Management Environments**. Porto Alegre, Sociedade Brasileira de Computação, 2018.
- SCHNEIDER, Pedro Henrique. **Análise preditiva de Churn com ênfase em técnicas de Machine Learning: uma revisão**. FGV Repositório Digital, 2016.
- TING, Kai Ming. **Confusion Matrix**. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA, 2017.
- WU, Xiaojun; MENG, Sufang. **E-commerce customer churn prediction based on improved SMOTE and AdaBoost**. Kunming, China. IEEE, 2016.
- ZHU, Wen; ZENG, Nancy; WANG Ning. **Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations**. Northeast SAS User Group proceedings, Section of Health Care and Life Sciences, Baltimore, Maryland, 2010.

APÊNDICES

A. Dados estatísticos das variáveis do Modelo 3

Variável	B	S.E.	Wald	df	Sig.	Exp(B)
CIDADE_EBCT(108376)			136,491	307	1,000	
CIDADE_EBCT(66303)	,405	42459,532	,000	1	1,000	1,499E+00
CIDADE_EBCT(66320)	-1,739	44038,677	,000	1	1,000	1,757E-01
CIDADE_EBCT(66346)	1,747	41193,619	,000	1	1,000	5,739E+00
CIDADE_EBCT(66354)	-,008	43914,991	,000	1	1,000	9,923E-01
CIDADE_EBCT(66389)	3,353	43568,752	,000	1	1,000	2,859E+01
CIDADE_EBCT(66419)	,129	44724,266	,000	1	1,000	1,138E+00
CIDADE_EBCT(66435)	1,415	48847,206	,000	1	1,000	4,118E+00
CIDADE_EBCT(66486)	19,184	40179,915	,000	1	1,000	2,146E+08
CIDADE_EBCT(66494)	,190	42083,303	,000	1	1,000	1,209E+00
CIDADE_EBCT(66508)	19,535	40179,915	,000	1	1,000	3,046E+08
CIDADE_EBCT(66516)	1,868	41742,353	,000	1	1,000	6,477E+00
CIDADE_EBCT(66540)	17,972	40179,915	,000	1	1,000	6,387E+07
CIDADE_EBCT(66559)	,711	41845,416	,000	1	1,000	2,037E+00
CIDADE_EBCT(66605)	,275	42411,536	,000	1	1,000	1,317E+00
CIDADE_EBCT(66613)	17,713	40179,915	,000	1	1,000	4,930E+07
CIDADE_EBCT(66621)	,187	42414,554	,000	1	1,000	1,206E+00
CIDADE_EBCT(66630)	20,313	40179,915	,000	1	1,000	6,633E+08
CIDADE_EBCT(66656)	1,330	42837,483	,000	1	1,000	3,783E+00
CIDADE_EBCT(66664)	,888	40952,747	,000	1	1,000	2,430E+00
CIDADE_EBCT(66710)	20,497	40179,915	,000	1	1,000	7,974E+08
CIDADE_EBCT(66737)	,658	41972,451	,000	1	1,000	1,930E+00
CIDADE_EBCT(66770)	18,943	40179,915	,000	1	1,000	1,686E+08
CIDADE_EBCT(66800)	,339	41028,326	,000	1	1,000	1,404E+00
CIDADE_EBCT(66826)	1,062	42442,750	,000	1	1,000	2,891E+00
CIDADE_EBCT(66869)	19,528	40179,915	,000	1	1,000	3,026E+08
CIDADE_EBCT(66885)	,217	49079,503	,000	1	1,000	1,242E+00
CIDADE_EBCT(66966)	,037	41281,882	,000	1	1,000	1,038E+00
CIDADE_EBCT(67040)	19,273	40179,915	,000	1	1,000	2,346E+08
CIDADE_EBCT(67113)	1,777	56832,213	,000	1	1,000	5,914E+00
CIDADE_EBCT(67121)	-,613	44822,405	,000	1	1,000	5,418E-01
CIDADE_EBCT(67199)	19,858	40179,915	,000	1	1,000	4,208E+08
CIDADE_EBCT(67245)	,958	40861,882	,000	1	1,000	2,606E+00
CIDADE_EBCT(67288)	,149	41262,329	,000	1	1,000	1,160E+00
CIDADE_EBCT(67326)	-1,876	56832,213	,000	1	1,000	1,532E-01
CIDADE_EBCT(67350)	19,017	40179,915	,000	1	1,000	1,815E+08
CIDADE_EBCT(67377)	18,321	40179,915	,000	1	1,000	9,048E+07
CIDADE_EBCT(67385)	-,428	56832,213	,000	1	1,000	6,517E-01

CIDADE_EBCT(67393)	18,249	40179,915	,000	1	1,000	8,422E+07
CIDADE_EBCT(67407)	18,355	40179,915	,000	1	1,000	9,364E+07
CIDADE_EBCT(67423)	,990	56832,213	,000	1	1,000	2,691E+00
CIDADE_EBCT(67466)	17,707	40179,915	,000	1	1,000	4,898E+07
CIDADE_EBCT(67474)	2,314	42794,969	,000	1	1,000	1,012E+01
CIDADE_EBCT(67482)	-,104	41372,127	,000	1	1,000	9,016E-01
CIDADE_EBCT(67555)	18,824	40179,915	,000	1	1,000	1,497E+08
CIDADE_EBCT(67636)	,989	44587,649	,000	1	1,000	2,689E+00
CIDADE_EBCT(67644)	,926	41359,377	,000	1	1,000	2,525E+00
CIDADE_EBCT(67695)	,109	43147,770	,000	1	1,000	1,115E+00
CIDADE_EBCT(67709)	18,969	40179,915	,000	1	1,000	1,730E+08
CIDADE_EBCT(67717)	19,227	40179,915	,000	1	1,000	2,239E+08
CIDADE_EBCT(67733)	18,595	40179,915	,000	1	1,000	1,191E+08
CIDADE_EBCT(67784)	19,284	40179,915	,000	1	1,000	2,371E+08
CIDADE_EBCT(67806)	19,085	40179,915	,000	1	1,000	1,944E+08
CIDADE_EBCT(67814)	1,076	40810,474	,000	1	1,000	2,932E+00
CIDADE_EBCT(67849)	19,620	40179,915	,000	1	1,000	3,318E+08
CIDADE_EBCT(67857)	1,326	41795,875	,000	1	1,000	3,768E+00
CIDADE_EBCT(67890)	2,572	42424,093	,000	1	1,000	1,309E+01
CIDADE_EBCT(67911)	,440	40534,408	,000	1	1,000	1,552E+00
CIDADE_EBCT(67920)	1,062	40618,892	,000	1	1,000	2,891E+00
CIDADE_EBCT(67954)	1,140	41452,377	,000	1	1,000	3,126E+00
CIDADE_EBCT(67989)	1,563	44749,925	,000	1	1,000	4,775E+00
CIDADE_EBCT(68004)	-,410	44537,870	,000	1	1,000	6,638E-01
CIDADE_EBCT(68020)	17,646	40179,915	,000	1	1,000	4,608E+07
CIDADE_EBCT(68110)	20,133	40179,915	,000	1	1,000	5,544E+08
CIDADE_EBCT(68160)	1,826	41287,450	,000	1	1,000	6,209E+00
CIDADE_EBCT(68187)	18,052	40179,915	,000	1	1,000	6,915E+07
CIDADE_EBCT(68217)	17,351	40179,915	,000	1	1,000	3,432E+07
CIDADE_EBCT(68276)	19,579	40179,915	,000	1	1,000	3,185E+08
CIDADE_EBCT(68292)	,297	41580,001	,000	1	1,000	1,346E+00
CIDADE_EBCT(68462)	1,344	43299,642	,000	1	1,000	3,834E+00
CIDADE_EBCT(68500)	,272	46219,153	,000	1	1,000	1,312E+00
CIDADE_EBCT(68535)	2,357	45737,120	,000	1	1,000	1,056E+01
CIDADE_EBCT(68560)	,621	42211,999	,000	1	1,000	1,861E+00
CIDADE_EBCT(68608)	18,341	40179,915	,000	1	1,000	9,234E+07
CIDADE_EBCT(68632)	1,598	41211,823	,000	1	1,000	4,944E+00
CIDADE_EBCT(68659)	,388	40638,598	,000	1	1,000	1,473E+00
CIDADE_EBCT(68675)	1,055	40866,159	,000	1	1,000	2,872E+00
CIDADE_EBCT(68721)	17,916	40179,915	,000	1	1,000	6,036E+07
CIDADE_EBCT(68780)	18,223	40179,915	,000	1	1,000	8,209E+07
CIDADE_EBCT(68802)	2,117	43056,518	,000	1	1,000	8,306E+00
CIDADE_EBCT(68810)	20,264	40179,915	,000	1	1,000	6,316E+08

CIDADE_EBCT(68837)	1,977	49196,458	,000	1	1,000	7,224E+00
CIDADE_EBCT(68845)	,568	44675,768	,000	1	1,000	1,764E+00
CIDADE_EBCT(68918)	18,204	40179,915	,000	1	1,000	8,051E+07
CIDADE_EBCT(68934)	17,734	40179,915	,000	1	1,000	5,031E+07
CIDADE_EBCT(68950)	1,509	40477,070	,000	1	1,000	4,522E+00
CIDADE_EBCT(68985)	,252	49214,294	,000	1	1,000	1,286E+00
CIDADE_EBCT(69000)	1,341	44374,439	,000	1	1,000	3,824E+00
CIDADE_EBCT(69027)	,599	43785,921	,000	1	1,000	1,820E+00
CIDADE_EBCT(69132)	19,982	40179,915	,000	1	1,000	4,764E+08
CIDADE_EBCT(69191)	18,462	40179,915	,000	1	1,000	1,043E+08
CIDADE_EBCT(69205)	17,938	40179,915	,000	1	1,000	6,169E+07
CIDADE_EBCT(69230)	19,103	40179,915	,000	1	1,000	1,979E+08
CIDADE_EBCT(69256)	-,594	49105,399	,000	1	1,000	5,523E-01
CIDADE_EBCT(69280)	1,483	48573,259	,000	1	1,000	4,405E+00
CIDADE_EBCT(69337)	20,250	40179,915	,000	1	1,000	6,227E+08
CIDADE_EBCT(69353)	1,099	42208,936	,000	1	1,000	3,000E+00
CIDADE_EBCT(69396)	,496	41619,291	,000	1	1,000	1,643E+00
CIDADE_EBCT(69400)	19,177	40179,915	,000	1	1,000	2,130E+08
CIDADE_EBCT(69434)	,704	40690,511	,000	1	1,000	2,023E+00
CIDADE_EBCT(69477)	19,302	40179,915	,000	1	1,000	2,415E+08
CIDADE_EBCT(69485)	3,671	46221,405	,000	1	1,000	3,931E+01
CIDADE_EBCT(69540)	2,598	42711,286	,000	1	1,000	1,344E+01
CIDADE_EBCT(69566)	1,348	40562,053	,000	1	1,000	3,850E+00
CIDADE_EBCT(69604)	,528	44800,686	,000	1	1,000	1,695E+00
CIDADE_EBCT(69612)	19,485	40179,915	,000	1	1,000	2,899E+08
CIDADE_EBCT(69620)	19,635	40179,915	,000	1	1,000	3,369E+08
CIDADE_EBCT(69647)	1,293	42543,032	,000	1	1,000	3,642E+00
CIDADE_EBCT(69663)	18,171	40179,915	,000	1	1,000	7,792E+07
CIDADE_EBCT(69698)	1,199	48707,835	,000	1	1,000	3,317E+00
CIDADE_EBCT(69701)	19,166	40179,915	,000	1	1,000	2,107E+08
CIDADE_EBCT(69710)	1,509	48615,534	,000	1	1,000	4,524E+00
CIDADE_EBCT(69728)	18,325	40179,915	,000	1	1,000	9,088E+07
CIDADE_EBCT(69736)	19,462	40179,915	,000	1	1,000	2,833E+08
CIDADE_EBCT(69760)	,983	41573,501	,000	1	1,000	2,672E+00
CIDADE_EBCT(69809)	,563	43092,247	,000	1	1,000	1,756E+00
CIDADE_EBCT(69825)	,393	43387,600	,000	1	1,000	1,481E+00
CIDADE_EBCT(69841)	20,002	40179,915	,000	1	1,000	4,862E+08
CIDADE_EBCT(69850)	1,124	46194,053	,000	1	1,000	3,077E+00
CIDADE_EBCT(69868)	,494	42347,670	,000	1	1,000	1,638E+00
CIDADE_EBCT(69876)	1,156	41098,194	,000	1	1,000	3,179E+00
CIDADE_EBCT(69884)	1,016	40463,641	,000	1	1,000	2,761E+00
CIDADE_EBCT(69914)	18,427	40179,915	,000	1	1,000	1,007E+08
CIDADE_EBCT(69930)	,945	42370,296	,000	1	1,000	2,574E+00

CIDADE_EBCT(70092)	2,089	48525,211	,000	1	1,000	8,074E+00
CIDADE_EBCT(70114)	,742	42830,498	,000	1	1,000	2,101E+00
CIDADE_EBCT(70165)	1,021	43179,253	,000	1	1,000	2,777E+00
CIDADE_EBCT(70173)	17,867	40179,915	,000	1	1,000	5,748E+07
CIDADE_EBCT(70343)	-,345	49124,001	,000	1	1,000	7,079E-01
CIDADE_EBCT(70378)	1,394	44671,346	,000	1	1,000	4,030E+00
CIDADE_EBCT(70386)	18,890	40179,915	,000	1	1,000	1,600E+08
CIDADE_EBCT(70408)	17,660	40179,915	,000	1	1,000	4,674E+07
CIDADE_EBCT(70610)	1,232	41676,935	,000	1	1,000	3,429E+00
CIDADE_EBCT(70629)	18,512	40179,915	,000	1	1,000	1,096E+08
CIDADE_EBCT(70637)	18,100	40179,915	,000	1	1,000	7,260E+07
CIDADE_EBCT(70661)	1,395	41986,377	,000	1	1,000	4,034E+00
CIDADE_EBCT(70670)	1,286	42599,830	,000	1	1,000	3,619E+00
CIDADE_EBCT(70718)	,955	42604,233	,000	1	1,000	2,599E+00
CIDADE_EBCT(70769)	18,961	40179,915	,000	1	1,000	1,716E+08
CIDADE_EBCT(70807)	-,432	48439,790	,000	1	1,000	6,494E-01
CIDADE_EBCT(70831)	,583	43278,652	,000	1	1,000	1,791E+00
CIDADE_EBCT(70858)	,992	41064,720	,000	1	1,000	2,696E+00
CIDADE_EBCT(70971)	17,534	40179,915	,000	1	1,000	4,122E+07
CIDADE_EBCT(70980)	1,855	42668,237	,000	1	1,000	6,389E+00
CIDADE_EBCT(71013)	1,026	41673,649	,000	1	1,000	2,789E+00
CIDADE_EBCT(71021)	,164	40792,906	,000	1	1,000	1,178E+00
CIDADE_EBCT(71048)	18,819	40179,915	,000	1	1,000	1,489E+08
CIDADE_EBCT(71056)	20,641	40179,915	,000	1	1,000	9,214E+08
CIDADE_EBCT(71072)	1,973	43393,088	,000	1	1,000	7,192E+00
CIDADE_EBCT(71080)	1,386	40909,886	,000	1	1,000	3,998E+00
CIDADE_EBCT(71099)	,583	49176,644	,000	1	1,000	1,791E+00
CIDADE_EBCT(71129)	2,222	41509,264	,000	1	1,000	9,222E+00
CIDADE_EBCT(71137)	1,101	41346,467	,000	1	1,000	3,008E+00
CIDADE_EBCT(71145)	19,091	40179,915	,000	1	1,000	1,955E+08
CIDADE_EBCT(71153)	2,562	44538,855	,000	1	1,000	1,297E+01
CIDADE_EBCT(71188)	4,038	56832,213	,000	1	1,000	5,668E+01
CIDADE_EBCT(71196)	,626	43815,291	,000	1	1,000	1,870E+00
CIDADE_EBCT(71200)	19,202	40179,915	,000	1	1,000	2,185E+08
CIDADE_EBCT(71218)	19,299	40179,915	,000	1	1,000	2,406E+08
CIDADE_EBCT(71242)	17,983	40179,915	,000	1	1,000	6,458E+07
CIDADE_EBCT(71293)	17,604	40179,915	,000	1	1,000	4,419E+07
CIDADE_EBCT(71390)	19,320	40179,915	,000	1	1,000	2,459E+08
CIDADE_EBCT(71412)	17,475	40179,915	,000	1	1,000	3,885E+07
CIDADE_EBCT(71447)	,978	40756,232	,000	1	1,000	2,659E+00
CIDADE_EBCT(71455)	19,875	40179,915	,000	1	1,000	4,283E+08
CIDADE_EBCT(71463)	19,925	40179,915	,000	1	1,000	4,500E+08
CIDADE_EBCT(71471)	1,463	42201,164	,000	1	1,000	4,317E+00

CIDADE_EBCT(71480)	1,831	41588,465	,000	1	1,000	6,242E+00
CIDADE_EBCT(71498)	19,061	40179,915	,000	1	1,000	1,898E+08
CIDADE_EBCT(71501)	1,651	56832,213	,000	1	1,000	5,211E+00
CIDADE_EBCT(71579)	,559	43143,959	,000	1	1,000	1,749E+00
CIDADE_EBCT(71587)	18,423	40179,915	,000	1	1,000	1,002E+08
CIDADE_EBCT(71633)	,655	40812,663	,000	1	1,000	1,924E+00
CIDADE_EBCT(71706)	17,971	40179,915	,000	1	1,000	6,381E+07
CIDADE_EBCT(71714)	19,356	40179,915	,000	1	1,000	2,549E+08
CIDADE_EBCT(71722)	18,321	40179,915	,000	1	1,000	9,048E+07
CIDADE_EBCT(71773)	2,050	49169,968	,000	1	1,000	7,765E+00
CIDADE_EBCT(71811)	-1,054	43081,730	,000	1	1,000	3,484E-01
CIDADE_EBCT(71820)	1,169	48791,146	,000	1	1,000	3,220E+00
CIDADE_EBCT(71846)	,774	44559,425	,000	1	1,000	2,169E+00
CIDADE_EBCT(71870)	,108	41116,503	,000	1	1,000	1,115E+00
CIDADE_EBCT(71927)	22,666	40179,915	,000	1	1,000	6,979E+09
CIDADE_EBCT(71978)	19,501	40179,915	,000	1	1,000	2,946E+08
CIDADE_EBCT(71986)	18,838	40179,915	,000	1	1,000	1,518E+08
CIDADE_EBCT(72044)	1,996	46294,867	,000	1	1,000	7,360E+00
CIDADE_EBCT(72109)	18,792	40179,915	,000	1	1,000	1,449E+08
CIDADE_EBCT(72133)	19,894	40179,915	,000	1	1,000	4,365E+08
CIDADE_EBCT(72168)	-,096	41794,409	,000	1	1,000	9,087E-01
CIDADE_EBCT(72230)	1,792	44557,860	,000	1	1,000	6,000E+00
CIDADE_EBCT(72281)	1,619	44727,955	,000	1	1,000	5,050E+00
CIDADE_EBCT(72290)	1,357	41166,633	,000	1	1,000	3,884E+00
CIDADE_EBCT(72311)	,654	42694,281	,000	1	1,000	1,924E+00
CIDADE_EBCT(72451)	18,343	40179,915	,000	1	1,000	9,256E+07
CIDADE_EBCT(72478)	17,908	40179,915	,000	1	1,000	5,987E+07
CIDADE_EBCT(72508)	1,261	42134,007	,000	1	1,000	3,528E+00
CIDADE_EBCT(72524)	18,679	40179,915	,000	1	1,000	1,295E+08
CIDADE_EBCT(72559)	18,031	40179,915	,000	1	1,000	6,775E+07
CIDADE_EBCT(72613)	,549	44705,725	,000	1	1,000	1,732E+00
CIDADE_EBCT(72621)	,561	48351,003	,000	1	1,000	1,753E+00
CIDADE_EBCT(72656)	20,688	40179,915	,000	1	1,000	9,658E+08
CIDADE_EBCT(72702)	,419	41923,627	,000	1	1,000	1,520E+00
CIDADE_EBCT(72729)	,841	41128,234	,000	1	1,000	2,318E+00
CIDADE_EBCT(72737)	2,503	48884,681	,000	1	1,000	1,222E+01
CIDADE_EBCT(72770)	,298	42797,944	,000	1	1,000	1,347E+00
CIDADE_EBCT(72800)	,407	41256,803	,000	1	1,000	1,502E+00
CIDADE_EBCT(72842)	18,714	40179,915	,000	1	1,000	1,341E+08
CIDADE_EBCT(72907)	18,063	40179,915	,000	1	1,000	6,996E+07
CIDADE_EBCT(72915)	1,163	40903,864	,000	1	1,000	3,201E+00
CIDADE_EBCT(72923)	-,343	40672,596	,000	1	1,000	7,099E-01
CIDADE_EBCT(72940)	17,881	40179,915	,000	1	1,000	5,829E+07

CIDADE_EBCT(72966)	2,020	44512,576	,000	1	1,000	7,538E+00
CIDADE_EBCT(72982)	1,936	56832,213	,000	1	1,000	6,933E+00
CIDADE_EBCT(73040)	1,174	42551,360	,000	1	1,000	3,234E+00
CIDADE_EBCT(73059)	2,434	44878,122	,000	1	1,000	1,140E+01
CIDADE_EBCT(73075)	,048	41241,461	,000	1	1,000	1,049E+00
CIDADE_EBCT(73083)	1,049	49162,218	,000	1	1,000	2,854E+00
CIDADE_EBCT(73091)	,689	42555,979	,000	1	1,000	1,991E+00
CIDADE_EBCT(73130)	18,869	40179,915	,000	1	1,000	1,565E+08
CIDADE_EBCT(73237)	2,104	42744,366	,000	1	1,000	8,198E+00
CIDADE_EBCT(73245)	19,417	40179,915	,000	1	1,000	2,708E+08
CIDADE_EBCT(73253)	,108	56832,213	,000	1	1,000	1,114E+00
CIDADE_EBCT(73261)	1,205	40649,236	,000	1	1,000	3,336E+00
CIDADE_EBCT(73270)	,254	40544,486	,000	1	1,000	1,289E+00
CIDADE_EBCT(73296)	19,006	40179,915	,000	1	1,000	1,796E+08
CIDADE_EBCT(73342)	,985	43894,284	,000	1	1,000	2,677E+00
CIDADE_EBCT(73350)	1,054	44829,004	,000	1	1,000	2,870E+00
CIDADE_EBCT(73407)	1,850	44750,115	,000	1	1,000	6,363E+00
CIDADE_EBCT(73415)	1,641	41369,023	,000	1	1,000	5,163E+00
CIDADE_EBCT(73423)	1,033	56832,213	,000	1	1,000	2,809E+00
CIDADE_EBCT(73466)	18,238	40179,915	,000	1	1,000	8,334E+07
CIDADE_EBCT(73482)	1,105	40458,096	,000	1	1,000	3,018E+00
CIDADE_EBCT(73520)	,834	40991,523	,000	1	1,000	2,304E+00
CIDADE_EBCT(73539)	3,109	46285,975	,000	1	1,000	2,240E+01
CIDADE_EBCT(73636)	1,341	42584,906	,000	1	1,000	3,822E+00
CIDADE_EBCT(73695)	18,436	40179,915	,000	1	1,000	1,015E+08
CIDADE_EBCT(73709)	18,764	40179,915	,000	1	1,000	1,409E+08
CIDADE_EBCT(73741)	1,493	44491,347	,000	1	1,000	4,449E+00
CIDADE_EBCT(73776)	1,603	41862,146	,000	1	1,000	4,970E+00
CIDADE_EBCT(73784)	-22,01	69608,730	,000	1	1,000	2,755E-10
CIDADE_EBCT(73792)	18,260	40179,915	,000	1	1,000	8,516E+07
CIDADE_EBCT(73806)	18,732	40179,915	,000	1	1,000	1,366E+08
CIDADE_EBCT(73814)	19,381	40179,915	,000	1	1,000	2,612E+08
CIDADE_EBCT(73865)	,881	44474,367	,000	1	1,000	2,414E+00
CIDADE_EBCT(73873)	4,285	56832,213	,000	1	1,000	7,258E+01
CIDADE_EBCT(73890)	19,203	40179,915	,000	1	1,000	2,187E+08
CIDADE_EBCT(73911)	,979	40725,023	,000	1	1,000	2,662E+00
CIDADE_EBCT(73954)	1,349	42381,611	,000	1	1,000	3,853E+00
CIDADE_EBCT(73962)	1,480	41165,546	,000	1	1,000	4,394E+00
CIDADE_EBCT(74047)	1,646	46243,691	,000	1	1,000	5,185E+00
CIDADE_EBCT(74055)	19,096	40179,915	,000	1	1,000	1,965E+08
CIDADE_EBCT(74080)	,552	41668,193	,000	1	1,000	1,736E+00
CIDADE_EBCT(74098)	19,112	40179,915	,000	1	1,000	1,996E+08
CIDADE_EBCT(74128)	1,429	41065,366	,000	1	1,000	4,175E+00

CIDADE_EBCT(74209)	19,238	40179,915	,000	1	1,000	2,265E+08
CIDADE_EBCT(74217)	,965	42517,849	,000	1	1,000	2,625E+00
CIDADE_EBCT(74233)	20,038	40179,915	,000	1	1,000	5,042E+08
CIDADE_EBCT(74241)	18,654	40179,915	,000	1	1,000	1,262E+08
CIDADE_EBCT(74268)	19,237	40179,915	,000	1	1,000	2,262E+08
CIDADE_EBCT(74284)	1,038	41288,496	,000	1	1,000	2,823E+00
CIDADE_EBCT(74306)	18,832	40179,915	,000	1	1,000	1,508E+08
CIDADE_EBCT(74322)	17,692	40179,915	,000	1	1,000	4,824E+07
CIDADE_EBCT(74403)	19,990	40179,915	,000	1	1,000	4,801E+08
CIDADE_EBCT(74438)	19,450	40179,915	,000	1	1,000	2,800E+08
CIDADE_EBCT(74446)	,289	42100,516	,000	1	1,000	1,335E+00
CIDADE_EBCT(74489)	1,556	41561,929	,000	1	1,000	4,739E+00
CIDADE_EBCT(74497)	19,638	40179,915	,000	1	1,000	3,379E+08
CIDADE_EBCT(74500)	,073	42037,483	,000	1	1,000	1,076E+00
CIDADE_EBCT(74519)	,595	45892,962	,000	1	1,000	1,814E+00
CIDADE_EBCT(74560)	18,679	40179,915	,000	1	1,000	1,295E+08
CIDADE_EBCT(74578)	18,061	40179,915	,000	1	1,000	6,979E+07
CIDADE_EBCT(74659)	19,175	40179,915	,000	1	1,000	2,125E+08
CIDADE_EBCT(74675)	-,001	48905,850	,000	1	1,000	9,985E-01
CIDADE_EBCT(74764)	-2,005	56832,213	,000	1	1,000	1,347E-01
CIDADE_EBCT(74802)	1,476	42527,601	,000	1	1,000	4,375E+00
CIDADE_EBCT(74810)	20,069	40179,915	,000	1	1,000	5,198E+08
CIDADE_EBCT(74829)	17,782	40179,915	,000	1	1,000	5,282E+07
CIDADE_EBCT(74837)	,945	40564,646	,000	1	1,000	2,572E+00
CIDADE_EBCT(74853)	17,699	40179,915	,000	1	1,000	4,861E+07
CIDADE_EBCT(74861)	-,521	45334,638	,000	1	1,000	5,937E-01
CIDADE_EBCT(74888)	18,419	40179,915	,000	1	1,000	9,984E+07
CIDADE_EBCT(74900)	1,762	44583,757	,000	1	1,000	5,826E+00
CIDADE_EBCT(74942)	-,371	44003,994	,000	1	1,000	6,903E-01
CIDADE_EBCT(74977)	,132	46043,902	,000	1	1,000	1,141E+00
CIDADE_EBCT(75051)	,996	46103,484	,000	1	1,000	2,709E+00
CIDADE_EBCT(89605)	2,016	48864,999	,000	1	1,000	7,506E+00
CIDADE_EBCT(92835)	19,650	40179,915	,000	1	1,000	3,420E+08
CIDADE_EBCT(92843)	1,010	49142,049	,000	1	1,000	2,744E+00
CIDADE_EBCT(93289)	,530	42647,236	,000	1	1,000	1,698E+00
CIDADE_EBCT(93394)	1,218	48111,991	,000	1	1,000	3,380E+00
CIDADE_EBCT(97225)	19,123	40179,915	,000	1	1,000	2,018E+08
CIDADE_EBCT(97233)	19,243	40179,915	,000	1	1,000	2,277E+08
CIDADE_EBCT(97241)	1,609	43062,157	,000	1	1,000	5,000E+00
CIDADE_EBCT(97250)	19,561	40179,915	,000	1	1,000	3,128E+08
CIDADE_EBCT(97268)	,928	41081,787	,000	1	1,000	2,530E+00
CIDADE_EBCT(97276)	19,903	40179,915	,000	1	1,000	4,404E+08
CIDADE_EBCT(97284)	1,596	41553,431	,000	1	1,000	4,933E+00

CIDADE_EBCT(97292)	1,252	43900,309	,000	1	1,000	3,497E+00
CIDADE_EBCT(97306)	,423	42529,513	,000	1	1,000	1,526E+00
CIDADE_EBCT(105406)	18,679	40179,915	,000	1	1,000	1,295E+08
CIDADE_EBCT(105813)	1,052	49063,809	,000	1	1,000	2,864E+00
CIDADE_EBCT(105821)	1,769	44716,084	,000	1	1,000	5,866E+00
CIDADE_EBCT(105830)	,769	43848,702	,000	1	1,000	2,157E+00
CIDADE_EBCT(108341)	3,029	45288,307	,000	1	1,000	2,067E+01
CIDADE_EBCT(108342)	1,460	43224,827	,000	1	1,000	4,305E+00
CIDADE_EBCT(108343)	,505	42873,095	,000	1	1,000	1,656E+00
CIDADE_EBCT(108344)	-,062	45867,388	,000	1	1,000	9,398E-01
CIDADE_EBCT(108372)	43,237	49213,765	,000	1	,999	5,993E+18
CIDADE_EBCT(108373)	1,789	48817,299	,000	1	1,000	5,984E+00
BAIXOU_LIM_CRED(0)	-1,273	,261	23,729	1	,000	2,801E-01
TIPO_CREDITO(55)			73,115	27	,000	
TIPO_CREDITO(1)	12,760	14005,751	,000	1	,999	3,480E+05
TIPO_CREDITO(2)	-7,051	42563,316	,000	1	1,000	8,667E-04
TIPO_CREDITO(3)	13,309	14005,751	,000	1	,999	6,029E+05
TIPO_CREDITO(4)	14,874	14005,751	,000	1	,999	2,882E+06
TIPO_CREDITO(5)	-4,654	20682,476	,000	1	1,000	9,520E-03
TIPO_CREDITO(6)	15,209	14005,751	,000	1	,999	4,029E+06
TIPO_CREDITO(7)	13,881	14005,751	,000	1	,999	1,068E+06
TIPO_CREDITO(9)	-3,439	42563,316	,000	1	1,000	3,209E-02
TIPO_CREDITO(10)	14,216	14005,751	,000	1	,999	1,493E+06
TIPO_CREDITO(12)	11,960	14005,751	,000	1	,999	1,563E+05
TIPO_CREDITO(14)	-1,861	23512,034	,000	1	1,000	1,555E-01
TIPO_CREDITO(24)	-3,989	42563,316	,000	1	1,000	1,852E-02
TIPO_CREDITO(32)	-4,539	26323,736	,000	1	1,000	1,068E-02
TIPO_CREDITO(33)	13,298	14005,751	,000	1	,999	5,961E+05
TIPO_CREDITO(36)	11,898	14005,751	,000	1	,999	1,469E+05
TIPO_CREDITO(38)	14,355	14005,751	,000	1	,999	1,716E+06
TIPO_CREDITO(39)	13,137	14005,751	,000	1	,999	5,072E+05
TIPO_CREDITO(41)	13,352	14005,751	,000	1	,999	6,292E+05
TIPO_CREDITO(43)	11,471	14005,751	,000	1	,999	9,585E+04
TIPO_CREDITO(44)	11,362	14005,751	,000	1	,999	8,603E+04
TIPO_CREDITO(46)	-,311	42563,316	,000	1	1,000	7,328E-01
TIPO_CREDITO(47)	34,317	42563,317	,000	1	,999	8,007E+14
TIPO_CREDITO(49)	-3,635	42563,316	,000	1	1,000	2,639E-02
TIPO_CREDITO(50)	-3,147	14543,853	,000	1	1,000	4,300E-02
TIPO_CREDITO(51)	12,858	14005,751	,000	1	,999	3,837E+05
TIPO_CREDITO(52)	-1,921	42563,316	,000	1	1,000	1,465E-01
TIPO_CREDITO(53)	-4,711	31232,953	,000	1	1,000	8,999E-03
TEM_CONTRATO(0)	1,853	1,044	3,151	1	,076	6,379E+00
TEVE_RECLAMACAO(0)	-1,285	,251	26,174	1	,000	2,767E-01

TEMPO_EMPRESA	1,038	,293	12,561	1	,000	2,824E+00
QTD_PEDIDOS	-4,186	,647	41,870	1	,000	1,520E-02
VENDA_MARCAS_U3M	4,086	1,925	4,504	1	,034	5,950E+01
VENDA_TERCEIROS_U3M	4,535	1,989	5,197	1	,023	9,322E+01
VENDA_TOTAL_U3M	-6,912	2,505	7,615	1	,006	9,962E-04
MIX_MARCAS	-,639	,179	12,736	1	,000	5,278E-01
MESES_COM_VENDA	-1,458	,220	43,921	1	,000	2,327E-01
DIAS_SV	,606	,178	11,571	1	,001	1,832E+00
CONSTANTE	-30,50	42550,026	,000	1	,999	5,637E-14