

A CIENTIFIZAÇÃO DOS NEGÓCIOS: VISÃO GERAL E ESTUDO DE CASO DO SETOR FINANCEIRO¹

REINALDO G. GREGORI*
FERNANDO LINK**

Introdução

Uma das evoluções recentes mais notáveis nas práticas modernas de administração de empresas é a crescente utilização de métodos analíticos e de informações científicas na tomada de decisões estratégicas e táticas. Ainda nos dias de hoje, os “tomadores de decisões” se baseiam como sempre fizeram seus antecessores, no próprio conhecimento prático dos respectivos negócios, e até mesmo na própria “intuição”. Mas é inegável que estes profissionais vêm cada vez mais reconhecendo a necessidade e vantagens de combinar “inteligência científica” ao conhecimento prático na busca da assertividade. No mundo

¹ Os autores gostariam de agradecer ao Ricardo Galante Coimbra, analista de projetos na Cognatis, pelos inestimáveis comentários, sugestões e correções, Marco Aurélio Painelli Marsitch e Wagner Pacífico pela arte, ilustrações e diagramas utilizados .

* Diretor geral da Cognatis Estudos Mercadológicos, PhD em Demografia Econômica por UC Berkeley e Georgetown University.

** Gerente de Projetos da Cognatis Estudos Mercadológicos, economista pela USP

empresarial atual, caracterizado pelo alto grau de competitividade e dinamismo gerencial, por cenários de negócios crescentemente complexos, e por empresas administradas por profissionais especialmente treinados para tal, a prática de *cientificar* os processos de decisão estratégicos e táticos vem deixando de ser um diferencial competitivo para tornar-se uma prática básica de boa administração.

Neste trabalho apresentamos dois campos em que a aplicação de metodologias científicas, muitas delas até recentemente confinadas a setores acadêmicos ou de pesquisa aplicada, vem ganhando espaço e credibilidade por empresas de diversos segmentos e portes. São eles o geomarketing (em português muitas vezes traduzido como geomercadologia) e o CRM analítico. Será apresentada inicialmente uma descrição resumida da evolução do processo de *cientificação* na tomada de decisões em negócios, seguido da apresentação dos conceitos e aplicações do CRM analítico e geomarketing. Curiosamente, estes dois campos de inteligência mercadológica são freqüentemente tratados de modo independente pelas empresas, e até mesmo por empresas de consultoria especializadas, que tendem a se especializar em apenas um dos campos (dividindo-se entre empresas de geomarketing ou empresas de CRM analítico). Entretanto, não somente é possível integrar as abordagens e métodos utilizados por ambos os campos, como há ganhos significativos em tal prática, e este será o tema da seção seguinte. Para concluir, será apresentado um caso real de aplicação de métodos de geomarketing e CRM analítico integrados, com o objetivo de ilustrar e concretizar a discussão teórica apresentada nas seções anteriores.

O papel da ciência na tomada de decisões em negócios

A utilização de métodos científicos nas atividades de empresas é possivelmente tão antiga quanto a própria ciência formal. Inicialmente, tal utilização se concentrava em áreas diretamente ligadas à produção, seja no desenvolvimento de um novo produto, seja na melhoria no modo de produzir os itens existentes. Desde o início da segunda metade do século passado, entretanto, e mais acentuadamente nas últimas duas ou três décadas, a forma com que empresas vêm se apropriando de metodologias e conhecimentos científicos vem se transformando significativamente. Empresas vêm, desde então, e de modo crescente, utilizando estes elementos não apenas em seus processos de produção

e operação, mas também em suas decisões estratégicas, de planejamento, e até mesmo táticas (ou seja, de curto prazo). Neste capítulo discutimos dois campos de aplicações de métodos quantitativos em processos de decisões estratégicas e/ou táticas de empresas: o *geomarketing* e o *CRM analítico*. Estas aplicações vêm ganhando espaço no mercado internacional e brasileiro, e apresentam afinidade com o objeto e metodologias da demografia. Antes de prosseguir, entretanto, é importante esclarecer a terminologia adotada. Na prática, no mercado, há grande confusão sobre como denominar o *geomarketing* ou *CRM analítico*. São utilizados termos com “ferramentas”, “soluções”, “metodologias” etc. Procurando diminuir a confusão semântica, neste capítulo o *CRM analítico* e *geomarketing* serão referidos apenas como “campos” de aplicações ligadas à inteligência de negócios.

A aceleração no desenvolvimento destes campos nas últimas décadas se deu graças, por um lado, à proliferação de informações pertinentes e estruturadas a respeito do comportamento do consumidor, e por outro ao desenvolvimento e barateamento de tecnologias de gestão de informações em grandes empresas. São utilizados, em ambos os casos, métodos de análise estatística de dados de clientes ou de populações de potencial mercadológico, sendo as principais fontes de dados as próprias organizações (dados internos), dados de listas (*mailings*) adquiridos no mercado, dados coletados em campo (pesquisa primária) ou bases de dados secundários, organizadas na forma de microdados ou dados tabulares. No caso do *geomarketing*, dados populacionais são freqüentemente comercializados com referências às áreas geográficas, como setores censitários (menor área geográfica utilizada e publicada pelo IBGE na coleta do censo demográfico).

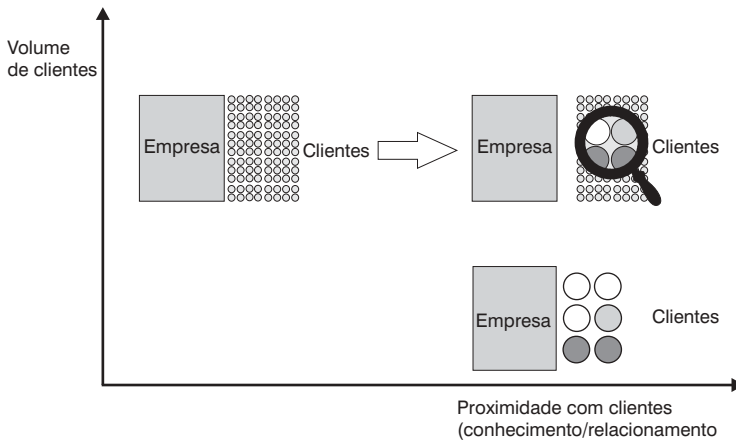
CRM e CRM Analítico

A sigla CRM significa literalmente “*Customer Relationship Management*”. No final da década passada (1990) e início desta, esta sigla foi exaustivamente utilizada por administradores modernos, “gurus” de negócios, e principalmente, vendedores de soluções de tecnologia que viram no *frenesi* do CRM uma grande oportunidade para alavancar os próprios negócios. Naqueles dias acreditava-se que as idéias e investimentos por detrás do CRM seriam a mais inteligente resposta para sobreviver e crescer em mercados cada vez mais competitivos e “*comoditizados*”,

e com margens de lucro cada vez menores.

O objetivo principal do CRM é criar e gerenciar um relacionamento lucrativo com cada um dos milhares ou milhões de clientes atendidos por uma organização. A premissa básica é que, com processos e tecnologia modernos seria possível desenvolver um relacionamento “personalizado” com cada cliente, possibilitando otimizar financeiramente todos os investimentos que a empresa faz para conquistar e manter seus clientes. A Figura 1 apresenta um diagrama estilizado da evolução do conceito do CRM.

Figura 1
Evolução do CRM



O quadrante inferior direito representa a situação existente em economias menos desenvolvidas, onde as empresas são menores, os setores mais pulverizados, e o volume médio de consumidores por empresa relativamente pequeno. Nesta situação, seria possível para os donos de empresas gerenciarem “naturalmente” o relacionamento com seus consumidores. Um dono de padaria ou loja de tecidos, por exemplo, seria capaz de se lembrar das preferências de cada cliente, além do respectivo comportamento creditício (quem paga em dia, quem não paga etc.), e implicitamente estimar o “valor” de cada cliente com base em tais informações. A partir deste conhecimento, poderia-se moldar o atendimento à sua clientela de modo a “tirar” o melhor de cada cliente.

Com o desenvolvimento do setor de serviços, muitas empresas e setores econômicos se consolidam, surgindo grandes organizações voltadas para o varejo ou atendimento a clientes. Hoje, no Brasil, por exemplo, há bancos de varejo com mais de 6 milhões de clientes em carteira, e empresas de telecomunicações com bases de clientes maiores que muitos países europeus. A gestão do relacionamento individualizado, possível anteriormente em empresas de pequeno porte, se torna impossível no mundo das grandes empresas. Estas empresas passam a focar suas energias no *produto*, ao invés do cliente, que se torna invisível individualmente. Surge o chamado *marketing de massa*, cujo principal objetivo é a construção de uma estratégia de comunicação focada em grandes populações, independente da variabilidade das características de seus indivíduos. Este momento no ciclo de evolução do CRM está caracterizado pelo quadrante superior esquerdo, onde o relacionamento é baixo, e o volume de consumidores ou clientes é alto.

À medida que empresas passam a focar os próprios produtos ao invés de seus clientes, é natural que não desenvolvam uma relação eficaz de fidelidade com seus clientes. Torna-se para muitos setores e principalmente para empresas de serviços, cada vez mais difícil manter uma margem de lucro adequada. As inovações tecnológicas se tornam cada vez mais frequentes e previsíveis, e com difusões nos respectivos setores cada vez mais rápidas. Tudo conspira para que a guerra mercadológica gire em torno do preço, conseqüentemente diminuindo margens e favorecendo novas consolidações lideradas por grupos empresariais maiores.

É neste contexto de margens decrescentes, grandes volumes de transações e cliente, e rápido desenvolvimento tecnológico que surge o conceito do CRM. A idéia básica é simples: com o uso de novas tecnologias, que possibilitam por um lado organizar e administrar grandes bases de dados a custos continuamente decrescentes, e por outro registrar digitalmente a maioria dos contatos e transações entre empresas e consumidores, se torna possível gerar informações necessárias para que empresas passem a administrar os milhões de relacionamentos com clientes de modo mais inteligente e “individualizado”. A partir desta inteligência, podem ser criadas estratégias de relacionamento com seus clientes que maximizassem o potencial dos mesmos, dando importância para os clientes com alto potencial de geração de valor para a empresa, mesmo que isso implique

em negligenciar, propositalmente, clientes menos promissores. O CRM é um conceito poderoso e atraente para empresas no quadrante superior esquerdo, o que explica sua rápida popularização. Porém, a incorporação deste conceito não é tão simples como parece e setores compradores de CRM tiveram de passar por um grande (e por vezes bastante caro) ajuste de expectativas para descobrir isso.

O CRM Analítico

O *boom* do CRM gerou altos investimentos por grandes empresas que objetivavam aprender a gerenciar melhor os relacionamentos com seus clientes finais. Na maioria dos casos, foram feitos grandes investimentos em tecnologias como sistemas de gestão, contatos, sistemas de gestão de campanhas, sistemas para organização de dados (DW, Datamarts, DBM, ETLs etc.). A indústria de *software* (as chamadas soluções) também se beneficiou, devendo-se destacar as soluções analíticas, como SAS e SPSS, soluções de gestão de informações estratégicas (sistemas de *business intelligence*) como MicroStrategy, BRIO, Business Objects etc.), e soluções de gestão de contato com clientes (Sieble etc.).

Mas foram apenas os vendedores de tecnologia que se beneficiaram do *boom* do CRM. As empresas perceberam logo que de pouco adiantava disponibilizar grandes volumes de dados transacionais e cadastrais de seus clientes se não houvesse métodos para destes se extrair informações relevantes para ações de negócios. Surgiram novas oportunidades para profissionais treinados em análise de dados (estatísticos, economistas, antropólogos e demógrafos) e importantes silos de profissionais com este perfil foram criados em organizações até então populadas predominantemente por administradores, marqueteiros, vendedores, contadores, operários etc. Aos poucos, estes movimentos contribuíram para introduzir um novo perfil de profissional corporativo em certas indústrias e setores, com competências mais analíticas e técnicas que a média dos administradores. E contribuíram também para introduzir novos termos no já inflado vocabulário de jargões corporativos, como LTV (*lifetime value*, ou valor presente líquido do cliente), modelos de propensão e de retenção ou *anti-churn*, dentre outros.

Este conjunto de técnicas e abordagens analíticas voltado para gerar conhecimento relevante para o negócio a partir da análise de dados comportamentais

e cadastrais dos clientes passou a ser conhecido como CRM Analítico (para diferenciar das soluções de CRM operacional e tecnológico). É neste sentido que o termo CRM analítico é usado neste capítulo, apesar do seu relativo desgaste e da freqüente confusão com relação ao seu significado.²

Superando a crise

Após o fim da bolha da internet, empreendedores no mundo inteiro passaram a “botar na ponta do lápis” suas expectativas de retorno com relação aos grandes investimentos feitos durante a década de 1990. Também os investimentos em CRM foram postos à prova, e em muitos casos concluiu-se que os resultados estavam bastante aquém do esperado, e que muitas das movimentações durante os anos de grandes investimentos foram feitas “por impulso”. Começou a proliferar uma percepção negativa em torno do *buzz word* CRM, o que levou muitas empresas vendedoras de tecnologias ou processos ligados ao CRM a adotar um novo vocabulário, mesmo continuando a vender as mesmas idéias e soluções. Na maioria dos casos, empresas que haviam investido grandes somas de recursos em projetos de CRM, concluíram que os resultados decepcionantes se deviam principalmente a não terem reorganizados seus processos e políticas para poderem se posicionar de modo diferente com relação aos seus clientes. Ou seja, o problema não teria sido a tecnologia, mas o fato de terem esquecido das demais transformações organizacionais que deveriam ter acompanhado o investimento tecnológico.

Mais recentemente, estas mesmas empresas vêm percebendo como obter valor sustentável das informações organizadas pelos grandes sistemas de CRM. Neste contexto, o CRM analítico, que é a utilização de métodos científicos na extração de valor informacional relevante dos grandes volumes de dados das empresas, passa a ser a “bola da vez”. Conseqüentemente, cresceram também de forma significativa as oportunidades para cientistas sociais com formação

² De fato, hoje se encontram muitos outros termos aparentemente capazes de substituir o CRM analítico como Data Mining, Modelos para Marketing, etc. Não cabe aqui uma discussão conceitual desta terminologia mas apenas a oportunidade de ressaltar que no vocabulário e mesmo literatura sobre estes temas há muita confusão e mistura conceitual, o que é natural considerando não se tratar este de um ambiente acadêmico e principalmente, considerando que o vocabulário corrente é fortemente influenciado por traduções mal feitas de produtos estrangeiros e/ou pelo discurso vendedor dos fornecedores de soluções e tecnologias.

quantitativa e experiência em análise de dados empíricos em empresas que fizeram este tipo de investimento, ou em empresas especializadas no fornecimento destes serviços para empresas que optam pela terceirização dos mesmos.

Aplicações

O leitor pouco familiarizado com as aplicações de CRM analítico deve estar se perguntando como exatamente, afinal, os métodos de análise de dados podem ajudar empresas a vender ou reter clientes. Nesta seção discutimos brevemente algumas aplicações práticas do CRM analítico e o papel da respectiva metodologia e abordagem científica. Uma descrição mais ampla sobre métodos estatísticos utilizados nestes ambientes, incluindo uma discussão interessante sobre a diferença entre os termos “*data mining*” e “modelos estatísticos” no contexto de aplicações empresariais pode ser encontrada em Berry, M. e Linoff G. (2000).

A seguir citamos alguns tipos de modelos com suas respectivas aplicações.

1. Modelos de Propensão – São popularmente chamados de **modelos de propensão** a aplicação de modelos preditivos (na maioria das vezes são regressões múltiplas lineares ou logísticas) com o objetivo de identificar a probabilidade (ou propensão) de um cliente comprar um determinado produto através de uma oferta direta (marketing direto ou telemarketing, por exemplo).
2. Modelos de Retenção (ou de atrito, ou *anti-churn*) – Estes modelos procuram identificar clientes que pretendem, num futuro próximo, desativar ou cancelar um ou vários produtos ou serviços oferecidos pela empresa. Tradicionalmente se aplicam a empresas prestadoras de serviços que exigem algum tipo de vínculo formal (bancos, telecomunicação, TV a cabo etc.), o que facilita a observação do evento “dependente”, a saber, o cancelamento ou desativação de um serviço. Dependendo de como é definido o evento a ser modelado, se a continuidade do cliente ou perda do mesmo, estes modelos podem ser chamados também de modelos de atrito (*attrition models*) ou modelos de *anti-churn*.³

³ A terminologia *Churn* é usada principalmente na indústria de operadoras de telecomunicação de longa distância ou celular, onde há hoje grande competitividade e poucas barreiras para mudança de serviço por parte do consumidor, que, a menos “retido” ou “blindado” pela operadora, corre o risco de trocar circuladamente de operadoras, eventualmente voltando à operadora inicial, e portanto “*churn*” como um motor que gira sobre si mesmo.

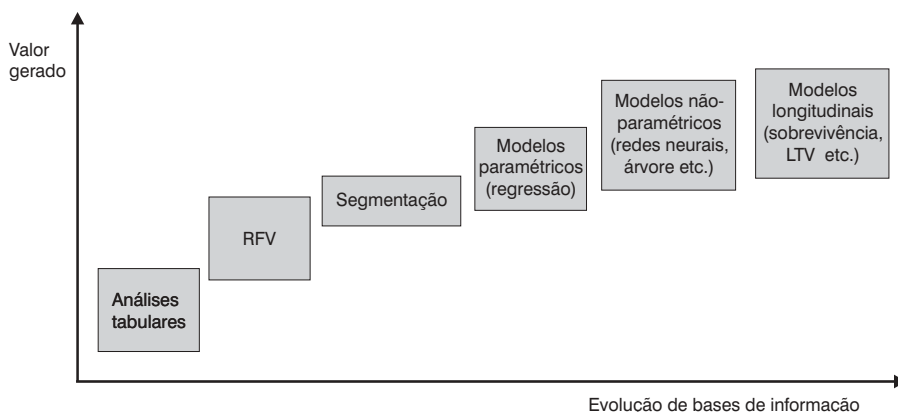
3. Modelos de *Life Time Value* (ou valor presente líquido) – Estes modelos procuram responder a seguinte pergunta: quanto vale cada cliente? O princípio é simples. O valor de um cliente é a soma descontada do fluxo futuro da receita líquida gerada pelo cliente até seu cancelamento ou rompimento com a empresa. O método utilizado pela Cognatis para estimar este valor é feito através de modelos de sobrevivência e/ou tábuas de vida (onde vida denota continuidade na prestação do serviço) para estimar o tempo médio (ou expectativa de vida) do cliente. À esta curva de sobrevivência é sobreposta uma curva de rentabilidade líquida esperada ao longo do tempo, e a partir dessa combinação resulta a curva de receita líquida esperada, já descontada a “mortalidade”.⁴ Este tipo de modelo é extremamente útil quando a empresa precisa saber como priorizar investimentos em marketing que afetam diferentemente seus clientes, ou quando está considerando adquirir outra empresa e precisa de uma metodologia para avaliar o valor dos clientes a serem assimilados.
4. Modelos de Segmentação – O grande volume de clientes e heterogeneidade entre eles torna quase impossível, para administradores, a tarefa de conhecer seus mercados e clientes. Os modelos de segmentação procuram amenizar esse problema classificando os milhares ou milhões de clientes em um número pequeno de segmentos, com base na similaridade dos mesmos. Apesar de não ser possível conhecer milhares ou milhões de clientes, é possível conhecer bem, digamos, dez “segmentos” de clientes. Assim, a segmentação proporciona aos administradores uma forma de gestão de relacionamento com clientes *quase* personalizada, baseada na identificação do segmento de pertencimento de cada cliente. Diferentes produtos, serviços, e políticas podem ser desenvolvidos para os distintos segmentos, aproveitando as diferenças entre os grupos para otimizar resultados. As técnicas estatísticas mais comumente utilizadas são aquelas normalmente aplicadas em modelos de *dusterização* de populações: modelos hierárquicos, não hierárquicos, *K-means* etc. Por vezes é necessário reduzir o número de fatores considerados no modelo de segmentação, o que pode ser feito através de análise fatorial ou similar. Mas a parte mais desafiante no

⁴ Este método lembra os cálculos demográficos em que a mortalidade feminina é descontada ao serem estimados o número de nascidos vivos de uma população a partir de um perfil de fecundidade por idade.

desenvolvimento de modelos de segmentação não é a aplicação de métodos estatísticos, mas sim entender profundamente os objetivos e expectativas de negócio para garantir que os segmentos resultantes serão úteis e aplicáveis para a empresa, do contrário, há um alto risco da segmentação não ser usada nas ações estratégicas da empresa.

O tipo e complexidade de modelos estatísticos utilizado no contexto do CRM analítico, bem como o valor informacional gerado por eles vêm evoluindo ao longo do tempo, conforme ilustrado na Figura 2. Pode-se dizer que hoje algumas empresas utilizam metodologias estatísticas de estimativas ou previsões (*forecasts*) tão ou mais sofisticadas que aquelas usadas em pesquisas sociais de cunho acadêmico, evidenciando a existência de um processo evolutivo na “cientificação” dos negócios, conforme sugerido anteriormente neste capítulo.

Figura 2
Ilustração da evolução na adoção de técnicas no CRM Analítico



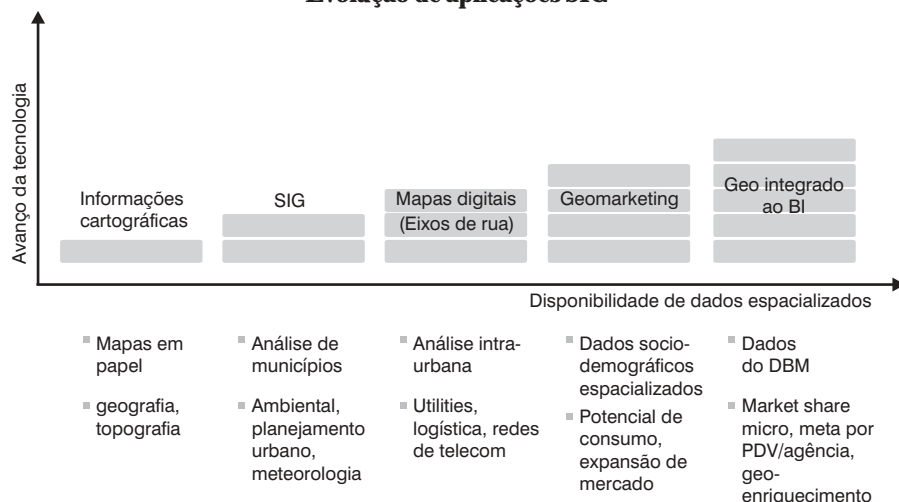
O geomarketing e a espacialização da inteligência mercadológica

Com a evolução do SIG (sistemas de informação geográfica) e proliferação de dados sociodemográficos passíveis de serem “espacializados”, análises geodemográficas começaram a ser aplicadas para estimar o potencial de demanda por serviços ou produtos de populações geograficamente circunscritas. Setores de serviços tradicionalmente oferecidos pelo setor público, como energia, água, saneamento, educação e telecomunicações foram pioneiros na aplicação

de SIG a dados geodemográficos em decisões de planejamento estratégico e até mesmo em decisões táticas com relação à oferta do serviço no curto prazo.

Não demorou para que empresas privadas, fornecedoras de serviços e produtos, e cujos mercados consumidores estão naturalmente relacionados ao espaço geográfico percebessem as possibilidades criadas por este tipo de análise e começassem a internalizar ou contratar serviços de SIG para tomada de decisões estratégicas. De modo geral, as aplicações de SIG à decisões mercadológicas em empresas privadas passaram a ser denominadas por *geomarketing*,⁵ e é esta definição ampla da expressão que utilizaremos neste texto. Na Figura 3, mostramos de modo genérico a evolução na aplicação da lógica geográfica e do SIG em negócios e em setores públicos.

Figura 3
Evolução de aplicações SIG



As aplicações de geomarketing surgiram ainda na década de 1980, principalmente nos EUA e Inglaterra, mas se popularizaram durante os anos 1990 com o barateamento significativo do SIG e de informações vetoriais (principalmente mapas) e de informações geodemográficas, ver, por exemplo, em Longley e Batty (eds) (2003), Longley e Clarke (eds) (1995) e Pick, J. (2005). Nos EUA e Inglaterra, as informações vetoriais e sociodemográficas associadas

⁵ Alguns analistas e autores traduziram geomarketing como geomercadologia, mas o termo é mais comumente usado em inglês no Brasil e internacionalmente.

a setores geográficos se tornaram acessíveis após os departamentos de estatística destes países passarem a comercializar informações demográficas agregadas por setores censitários e outras unidades geográficas, ver, por exemplo, em Martin e Zongley (1995), Waters (1995) e Martin (1995). No caso americano, o *Census Bureau* disponibilizou também, a baixo custo (e baixa qualidade) mapas de ruas (*street layers*) para a maioria das grandes cidades do país, impulsionando a utilização de SIG em negócios e no próprio setor público. No Brasil, o IBGE facilitou significativamente a utilização de informações agregadas por setores censitários a partir da publicação do Censo de 2000, impulsionando o uso de informações geodemográficas por empresas e pelo setor público.

Até o final da década de 1990, estudos e serviços de *geomarketing* no Brasil esbarravam na dificuldade de obtenção das informações necessárias. A partir da publicação do Censo de 2000, a informação geodemográfica⁶ bruta se tornou bem mais acessível, resultando no aumento da complexidade e amplitude dos estudos de *geomarketing* desenvolvidos por empresas especializadas no mercado que até então se limitavam, de modo geral, ao mapeamento de dados brutos. O desafio destes estudos deixa de ser a “obtenção” de informações brutas, como era antes do Censo 2000, passando a ser “como” usar estes dados para que resultem em informações mais assertivas e completas. Começam a surgir também maiores oportunidades para analistas e empresas de consultoria com outras especializações, principalmente para aquelas com maior tradição e competência em análises de dados, que até então não haviam percebido como utilizar suas capacitações em estudos de *geomarketing*. Hoje, os melhores estudos do gênero utilizam análises e modelos estatísticos sofisticados e integram, através destas metodologias, bases de informações de diversas fontes e em diversos níveis de agregação.⁷

O campo do *geomarketing* é extremamente fértil para o demógrafo com interesse em aplicações para negócios. O principal objetivo da maioria destes estudos é estimar o potencial mercadológico (ou potencial de demanda) de regiões, ou

⁶ Denominamos de dado geodemográfico toda informação socioeconômica ou demográfica de populações circunscritas geograficamente, passível de ser mapeada (georreferenciada) e disponibilizada em unidades geográficas de baixa agregação como setores censitários ou distritos.

⁷ O próprio IBGE disponibiliza dados sociodemográficos em diversos níveis de agregação: setor censitário, área de ponderação, região metropolitana ou município etc. Estas bases de informação podem ser combinadas através de interpolações baseadas em métodos estatísticos comumente utilizados por analistas com competências em *CRM analítico*, como *segmentação métodos interativos* ou *microssimulação demográfica*.

áreas intra-urbanas de interesse. Naturalmente, o potencial mercadológico em qualquer área geográfica está relacionado às características demográficas e econômicas da população que reside e/ou trabalha nestas regiões, além de fatores puramente mercadológicos como existência e proximidade da concorrência, *canibalização* pela concorrência da própria rede, infra-estrutura urbana etc.

Uma das principais dificuldades encontradas nestes estudos é estimar volumes populacionais em pequenas áreas em anos intercensitários. O problema é conhecido por demógrafos e discutido com grande propriedade por Paulo Jannuzzi nesta edição e em trabalhos anteriores. Resumidamente, projeções populacionais para pequenas áreas são particularmente difíceis devido à grande imprevisibilidade dos movimentos de mobilidade intra-urbana, tornando muitas vezes obsoletas as ferramentas tradicionais utilizadas por demógrafos em projeções de populações municipais ou maiores, como projeções por componentes ou baseadas em séries temporais. Infelizmente, no Brasil muitos analistas e empresas privadas de *geomarketing* parecem ignorar o problema, fazendo uso indiscriminado de métodos que interpolam linearmente o dado de crescimento estimado ou projetado para o município (normalmente oriundo do IBGE ou outra fonte externa) entre as pequenas áreas intra-urbanas, como setores censitários ou distritos. Esta é uma área de pesquisa em que demógrafos podem fazer e já fazem grandes contribuições, e que precisa ser mais bem divulgada entre analistas e empresas especializadas no mercado privado.

Aplicações

No Brasil, os primeiros setores a adotar aplicações de *geomarketing* em tomada de decisões foram o varejo (grande e médio), franquias e serviços de utilidade pública recentemente privatizados (telecomunicações, principalmente). Mais recentemente, percebeu-se uma tendência significativa por parte do setor financeiro em várias atividades (bancos, financeiras, cartão de crédito, caixas eletrônicos etc.), varejistas, setor automobilístico, setor educacional e até de produtos de consumo na utilização de estudos de *geomarketing*. Abaixo discutimos algumas das aplicações mais comuns deste tipo de inteligência mercadológica nestas empresas:

1. Expansão – Possivelmente a mais comum das aplicações do geomarketing. Esta aplicação consiste em estimar o potencial de regiões consideradas para expansão da rede de distribuição ou prestação de serviço em questão. Inicialmente, análises de geomarketing para expansão se limitavam a identificar concentrações de características geodemográficas previamente especificadas por terem, supostamente, relação com o processo de formação de demanda na região. Mais recentemente vêm surgindo aplicações que utilizam métodos estatísticos para estimar o potencial de demanda dos setores geográficos de interesse. De modo geral, estes estudos utilizam as lojas (também conhecidos como PDVs ou pontos de venda) existentes como unidades de observação. O modelo é aplicado para estimar a relação entre o desempenho de cada PDV e as variáveis geodemográficas das respectivas áreas geográficas de vizinhança. Normalmente, é pré-definida uma área de influência para cada PDV, que presumivelmente captura a maior parte da demanda gerada para aquela loja. Este tipo de estudo é discutido em mais detalhes em Birkin, Clarke e Clarke, 2002: cap. 3. Apesar de superior à análise descritiva pura (que ainda é a prática mais comum entre empresas usuárias e fornecedoras de serviços de geomarketing no Brasil), este tipo de estudo não considera a variabilidade geodemográfica existente dentro da área de influência. Uma alternativa superior é tratar como unidade de observação, ao invés da área de influência (e respectivo PDV), os setores censitários (unidades com cerca de 300 domicílios publicadas pelo IBGE) localizados dentro de cada área de influência estudada. Neste caso, os modelos deverão ser ajustados para estimar a contribuição que cada setor representa no desempenho de cada PDV estudado, com base nas características demográficas e econômicas das respectivas populações, e nas características mercadológicas e urbanísticas (distâncias dos concorrentes, outros serviços, da loja de referência, infraestrutura viária etc.) de cada setor “contribuinte”. Esta metodologia, desenvolvida e aplicada pela Cognatis, é uma adaptação da técnica conhecida com ‘*spatial interaction models*’, também discutida em Birkin, Clarke e Clarke, 2002: cap. 4. No decorrer desse texto, mostramos um caso real de aplicação desta metodologia para o setor de financeiras para crédito popular.
2. Otimização da rede – Decorre da possibilidade de estimar o potencial de um ponto ou região que cada ponto existente (lojas, agências etc.) possa ser avaliado

sob a luz do potencial de seu entorno. Tal avaliação possibilita estimar se uma determinada loja está desempenhando dentro, aquém ou além do esperado, considerando o potencial de sua região. Unidades com desempenho abaixo do esperado podem ser avaliadas com relação a outros aspectos do negócio: gestão, oferta de produtos, aspectos físicos do ambiente etc. Este tipo de abordagem tem inúmeras aplicações, indo da geração de metas com base no potencial à realocação ou exclusão de unidades com baixo desempenho e baixo potencial.

3. Marketing direto – Menos comum no Brasil, mas bastante comum nos EUA e alguns países europeus, o processo de caracterização geodemográfica de regiões censitárias ou, mais popularmente, regiões de CEP, possibilita delimitar geograficamente ações de marketing direto focadas em uma população ou mercado particular. Por exemplo, pode-se enviar uma mala direta para anunciar um novo automóvel apenas para regiões cujas populações apresentem predominantemente características associadas à demanda daquela categoria de automóvel. No Brasil há pouco *Zip Mail Marketing*, como é conhecido nos EUA, talvez por não existirem bases estruturadas de dados geodemográficos organizadas por CEP. Mas isso está mudando e a tendência é a popularização dessa prática.
4. Gestão de força de vendas – Muitas empresas administram forças de vendas em campo, tendo para tal que designar áreas de responsabilidades para a estrutura comercial ou promocional. Aplicando técnicas que estimam o potencial de regiões, passa a ser possível desenhar áreas de atuação de modo a otimizar a alocação de recursos e maximizar as chances de alcançar objetivos comerciais e mercadológicos.
5. Pesquisa de Mercado – Uma premissa básica do geomarketing é a existência de similaridade (observáveis e não-observáveis) entre domicílios e/ou indivíduos residentes ou que habitam uma determinada área geográfica. Nos EUA e Europa foram desenvolvidos sistemas de classificação (segmentação) geomercadológica que possibilitam com facilidade identificar o “tipo” (ou segmento) de uma vizinhança a partir de um endereço ou CEP. Este tipo de metodologia é especialmente útil na execução de pesquisas mercadológicas que necessitam ser representativas para todos ou um conjunto pré-determinado de segmentos geográficos. O sistema possibilita a estratificação da amostra

por segmento, reduzindo custos, e, à medida que o sistema de segmentação se torna conhecido por profissionais do meio, facilita a interpretação dos resultados obtidos. No Brasil, o único projeto de geossegmentação estruturado em desenvolvimento que temos conhecimento está sendo desenvolvido pela Cognatis.

6. Análise de risco – Outra aplicação da segmentação geodemográfica. Características regionais ou de áreas intra-urbanas podem ser usadas em modelos que objetivam estimar o risco de inadimplência de um cliente *prospect*. A principal idéia aqui é que as variáveis ecológicas funcionem como *proxy* para variáveis individuais não-observáveis mas associadas ao risco individual. Em países como os EUA, esta prática é regulamentada para impedir que *bureaus* de crédito pratiquem o que veio a ser chamado de “*red lining*”, ou seja, negar sistematicamente crédito para uma população com base em sua localização. Entretanto, é permissível usar variáveis geodemográficas em combinação a outras, com o objetivo de aumentar a eficácia do modelo que considera também variáveis individuais e domiciliares.
7. Mercado Imobiliário – A abordagem geomercadológica tem se mostrado útil em pelo menos duas aplicações para este mercado: análise de vocação e análise de valoração. No primeiro caso é feito um estudo para identificar a maior ‘vocação’ de um determinado espaço ou imóvel disponível. Por exemplo, um empreendedor proprietário de uma gleba no interior de São Paulo pode estar em dúvida se deve usar o espaço para construir um loteamento de casas de padrão popular, um centro de serviços ou shopping center, ou um hotel. O estudo de vocação consistiria, neste caso, no levantamento das características demográficas e econômicas ambientais do entorno, presença de serviços concorrentes ou complementares etc., para montar uma matriz de decisão e priorização entre as alternativas. Aplicações para avaliações de “precificação” consistem em análises das relações entre dinâmicas e características demográficas microrregionais e formação de preços imobiliários. Este tipo de análise é pouco difundido no Brasil, apesar do grande potencial para o planejamento de empresas empreendedoras imobiliárias. Uma revisão deste tipo de aplicação para os EUA pode ser encontrada em Wachter, S. *et al* (2005).

Integração do Geomarketing e CRM Analítico

Conforme apresentado anteriormente, dois novos movimentos de aplicação de técnicas analíticas nas áreas de marketing e planejamento de negócios vêm ganhando força nas últimas décadas: o CRM analítico e o geomarketing. Entretanto, apesar da grande sinergia entre estes dois campos de inteligência mercadológica, empresas e analistas raramente integram técnicas e abordagens das duas áreas. De fato, muitas empresas possuem estruturas organizacionais independentes para o geomarketing e CRM analítico (independentemente de como forem denominadas internamente). Há, entretanto, vantagens em integrar as duas abordagens. Nessa seção discutimos sucintamente como isso pode ser feito, e como a integração metodológica pode beneficiar tanto o geomarketing quanto o CRM analítico.

Aplicação de técnicas estatísticas no geomarketing

Uma grande limitação da abordagem tradicional do geomarketing é a falta de uma metodologia estruturada para “amarrar” os fatores geomercadológicos (demográficos, concorrência, distâncias etc.) dentro do processo de estimativa do potencial mercadológico da região ou área geográfica estudada. Algumas empresas utilizam índices que combinam vários fatores geodemográficos e mercadológicos através de um sistema de ponderação, de modo a refletir a percepção que o analista e/ou empresa contratante tem com relação à importância relativa de cada fator. Além de incorrerem freqüentemente em problemas metodológicos oriundos da má definição de escalas e/ou falta de padronização de medidas (que na maioria das vezes passa despercebida tanto pela empresa especializada ou analista como pela empresa contratante), esta abordagem é limitada pela forma arbitrária e subjetiva pelo qual o sistema de ponderação e forma funcional de combinação de fatores são construídos. Outra limitação do geomarketing tradicional é a falta de capacitação por parte das empresas especializadas em geomarketing para lidarem com informações transacionais e cadastrais de clientes existentes e registrados pela empresa contratante, quando esta informação existe e está disponível.

A utilização de dados e técnicas tradicionalmente utilizadas em aplicações de CRM analítico como as discutidas anteriormente neste texto é uma resposta natural e adequada para as limitações supracitadas. O conceito é simples. As variáveis geomercadológicas (demográficas e de mercado) devem ser tratadas como fatores explicativos em modelos multivariados desenhados para explicar o desempenho de unidades (pontos de vendas etc.) observadas. Na Figura 4, exibimos um diagrama ilustrando este conceito.

Figura 4
Aplicação de modelos multivariados no geomarketing



Normalmente, as informações para a construção do modelo são inicialmente georreferenciadas, ou seja, espacializadas, para depois serem organizadas tabularmente, conforme a estrutura geográfica de interesse, e somente então preparadas para serem utilizadas como insumos pelos modelos estatísticos.

A utilização de informações geográficas apresenta certas dificuldades adicionais para o processo de estimação dos parâmetros do modelo, particularmente quando são utilizados modelos lineares do tipo OLS (mínimos quadrados ordinários). Uma delas é a possível existência de autocorrelação espacial entre as regiões com proximidade geográfica. Autocorrelação espacial é um problema comum neste tipo de análise, freqüentemente resultante da correlação entre fatores *não observados* e, conseqüentemente, não especificado nos modelos, entre áreas com proximidade geográfica. Uma das conseqüências da autocorrelação espacial é a violação da premissa de independência nas distribuições

de “resíduos” entre observações, fundamental para garantir as propriedades desejáveis dos modelos de regressão linear. Este problema também aparece em modelos de séries temporais que utilizam dados de regiões, conforme discutido em Dowd, M. e LeSage, J.P. (1997) e dados não históricos em Wachter S. *et al.* (2005).

Independentemente das dificuldades metodológicas, é fato que a aplicação de técnicas multivariadas em dados geomercadológicos integrados a dados de clientes (oriundos das bases de CRM analítico) estão se tornando cada vez mais comuns e inevitavelmente se tornarão a norma para este tipo de campo de análise de inteligência de negócios. Ao final deste artigo exploramos com detalhes um caso de sucesso da Cognatis nesta área.

Aplicação de informações geodemográficas no CRM analítico

Da mesma forma que aplicações do geomarketing tradicional podem se beneficiar com os dados e técnicas usadas por desenvolvedores de análises de CRM analítico, o inverso também acontece. Nesta seção discutimos uma forma simples, importante e pouco utilizada, pela qual informações geodemográficas podem ser integradas a análises de CRM analítico que tradicionalmente não utilizam informações espacializadas.

Informações geodemográficas de uma área ou vizinhança, tanto em forma bruta (proporção de jovens, renda *per capita* etc.) ou reduzida (segmentação geodemográfica de vizinhanças tipo Acorn ou Psyte), podem servir como uma poderosa descrição das características dos indivíduos residentes na região. A premissa é que variáveis ambientais estejam relacionadas ao comportamento individual de duas formas: diretamente, devido à influência que o ambiente demográfico e mercadológico exerce no comportamento individual, e indiretamente devido ao processo de auto-seleção nas decisões residenciais dos indivíduos. Em ambos os casos, podem-se utilizar informações microambientais para enriquecer o nível de conhecimento do indivíduo, neste caso, clientes ou consumidores, e conseqüentemente melhorar os modelos preditivos usados em aplicações de CRM analítico.

Estudo de Caso – Modelo de Expansão desenvolvido pela Cognatis

Este estudo de caso mostra as soluções desenvolvidas pela Cognatis Consultoria Mercadológica para o plano de expansão com inteligência geomercadológica solicitado por uma empresa atuante no mercado de empréstimos pessoais para segmentos de baixa renda, ligado a um grande conglomerado financeiro.

Este projeto aliou conceitos de geomarketing com modelagem estatística utilizando dados demográficos e mercadológicos. A resposta final do projeto foi a indicação dos melhores pontos para a abertura de lojas, a partir do faturamento esperado, aplicados a todos os setores censitários do Estado de São Paulo.

Histórico

Desde o início dessa década, o setor de empréstimos pessoais para a população de baixa renda vem experimentando um forte crescimento. Este mercado era inicialmente caracterizado por empresas financeiras de médio porte focadas em um segmento populacional até então de pouco interesse para os grandes bancos. Entretanto, a partir de 2001 esse cenário começa a ser modificado através da consolidação entre concorrentes, impulsionado principalmente pela aquisição destas empresas por bancos de varejo de grande porte. Hoje, exemplos dos principais representantes deste setor incluem a Fininvest (Unibanco), a Finasa (Bradesco), a Taii (Itaú), o CitiFinancial (CitGroup), Aymoré (Banco Real ABN), Losango (HSBC), BMG, GE Money, e Ibi, do grupo C&A.

Apesar da consolidação de algumas empresas neste setor, o mercado de empréstimos pessoais apresenta enorme concorrência entre as empresas. As taxas de juros praticadas são muito elevadas, bem superiores à média de mercado, devido principalmente à alta inadimplência que caracteriza o mercado de baixa renda. Entretanto, o grande volume de demanda, por tratar de população demograficamente generosa e até recentemente excluída da oferta de serviços financeiros, vem permitindo um alto grau de elasticidade de preço neste mercado, razão pela qual a prática de juros extremamente altos não impediu o rápido crescimento do setor. Naturalmente, este cenário tende a mudar. À medida que os espaços do mercado vão sendo ocupados, e com a perspectiva de queda na taxa de juros básicos nos próximos anos, as empresas começam a perceber que,

para assegurar o sucesso no futuro próximo, torna-se fundamental melhorar continuamente os processos de avaliação de risco, para praticar taxas melhores ajustadas ao risco individual do cliente. Percebe também a necessidade de avaliar com profundidade o potencial de cada novo ponto de venda antes de decidir por sua abertura. O projeto desenvolvido pela Cognatis está relacionado com esta necessidade de avaliar o potencial para expansão.

Metodologia

Para este projeto de expansão foi desenvolvida uma metodologia que une técnicas de geomarketing (localização, cálculo de distâncias etc.) e técnicas estatísticas. O método usado para estimar o potencial de áreas geográficas foi um modelo de regressão múltipla que estima, separadamente, o impacto das características geodemográficas das áreas onde residem e trabalham clientes com a contribuição que estas áreas tiveram no desempenho de cada ponto de venda observado. Porém, devido à presença de muitos setores geográficos sem clientes residentes ou com endereços comerciais (uma situação comum, considerando o baixo nível de agregação dos setores geográficos utilizados), foi necessário aplicar aos modelos uma correção para distribuição censurada da variável dependente (neste caso, censuradas por baixo no valor zero). Tal correção, popularizada por Tobin (1958), se baseia na estimativa da demanda latente e não observada, cuja observação é censurada em zero. A seguir, discutimos em mais detalhes os principais passos e características da metodologia desenvolvida.

O projeto foi estruturado em quatro grandes etapas:

- Etapa 1: Mapeamento da demanda observada na rede, através das características geodemográficas dos clientes e da população residente ou trabalhadora.
- Etapa 2: Modelagem estatística. Estabelecer a relação entre o transacional observado e as características geodemográficas dos setores censitários.
- Etapa 3: Aplicação do modelo estatístico para o Estado de São Paulo. Identificação das áreas de maior potencial.
- Etapa 4: Checagem manual de cada ponto recomendado levando-se em conta características mercadológicas das áreas em estudo.

Foram utilizadas as seguintes informações transacionais, mercadológicas e geodemográficas:

- Cognatis GeoPop : Informações demográficas e socioeconômicas:
 - Informações do IBGE (Censo, Amostra, POF) compiladas, transformadas, interpoladas e estimadas pela Cognatis.
- Cognatis GeoPop Day: Informações sobre trabalhadores:
 - Informações do Ministério do Trabalho (Rais, Caged) compiladas, transformadas, interpoladas e estimadas pela Cognatis.
- Informações transacionais da financeira:
 - Endereços dos clientes (residencial e comercial).
 - Valor dos contratos.
 - Características sociodemográficas dos clientes (idade, renda, sexo, ocupação).
- Informações de mercado:
 - Localização das lojas da financeira em estudo e de seus principais concorrentes.
- Informações sobre agências bancárias, comércios, shopping centers e outros pólos de atração de fluxo de pessoas coletadas pela Cognatis.

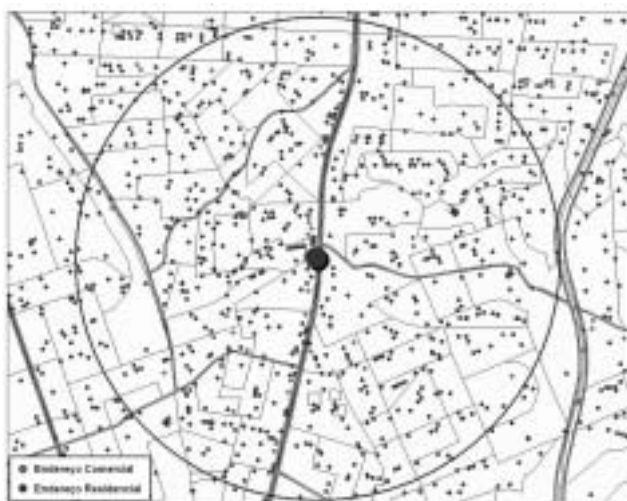
Mapeamento da Demanda

O mapeamento da demanda consiste na identificação da contribuição de cada setor censitário para o faturamento de cada loja existente. Para este estudo limitamos o raio de influência de cada loja em “N” quilômetros, garantindo que a área de influência de cada loja fosse suficiente para abranger no mínimo 60% dos clientes plotados. Limitamos o período de observação aos seis últimos meses, diminuindo assim possíveis distorções oriundas de sazonalidades e/ou alterações nas políticas de concessão de crédito implementadas ao longo do tempo. Lojas recém-inauguradas, sem histórico suficiente para gerar estimativas de faturamento robustas também foram excluídas. Finalmente, um fator medindo o tempo de existência de cada loja foi incluído nos modelos, buscando minimizar a heterogeneidade entre lojas devido às diferenças de “idade” entre lojas.

O primeiro passo consistiu no mapeamento geográfico (plotagem) dos endereços dos clientes em mapas digitais. Quando disponíveis, foram utilizados

tanto os endereços residenciais como comerciais. Uma vez plotados em mapas, foi possível identificar, através de um *software* de GIS, qual dentre os endereços de um mesmo cliente se localiza mais próximo à loja à qual o cliente se encontra registrado: o endereço residencial ou comercial. Através dessa identificação, a base de clientes foi dividida em duas: Clientes Residenciais e Clientes Comerciais. A principal premissa assumida é que um cliente cujo endereço residencial é mais próximo à loja de consumo que o endereço comercial, foi influenciado pela proximidade entre residência e loja na escolha daquela loja. Da mesma forma, clientes cujos endereços comerciais são mais próximos às respectivas lojas de opção, foram influenciados pela conveniência trazida pela proximidade entre loja e local de trabalho em suas decisões. Exibimos no Mapa 1 a localização de endereços comerciais e residenciais de um conjunto de clientes, com relação a uma loja hipotética, considerando apenas o endereço mais próximo para cada cliente.

Mapa 1
Dispersão de clientes comerciais e residenciais



Uma vez plotados e classificados os clientes, foi possível identificar, para cada ponto mapeado, o setor censitário ao qual cada cliente pertence. Em seguida, os clientes, e respectivas contribuições (neste caso utilizamos o volume de

empréstimo) foram agregados em cada setor, resultando na contribuição agregada de cada setor no faturamento de cada loja. Esta será, como veremos, a variável dependente na equação de potencial estimada posteriormente.

No Mapa 2, mostramos a contribuição hipotética dos setores censitários para o faturamento da loja.

Mapa 2
Contribuição dos setores censitários para o faturamento da loja



O passo seguinte foi estudar como variáveis geodemográficas da população residente e da população trabalhadora influenciam a contribuição de cada setor.

Modelagem Estatística

Conceitos e Premissas

O objetivo desta etapa é desenvolver um modelo de estimativa de demanda potencial para cada par de “setor-loja” estudado. Inicialmente, foram aplicados modelos de regressão múltipla, onde a contribuição de cada setor à cada loja foi assumida depender linearmente da combinação de características geodemográficas observadas da população de cada setor, da distância do setor à loja em questão, e de características mercadológicas e urbanísticas como

proximidade às lojas concorrentes, lojas do mesmo grupo (“canibalização”), centros comerciais etc. Foram desenvolvidos dois modelos totalmente independentes: o modelo denominado *residencial*, que estima a contribuição apenas de clientes classificados como “residenciais” na etapa anterior, e obviamente utiliza no “lado direito da equação” apenas características geodemográficas da população residente em cada setor; e o modelo denominado *comercial*, que estima a contribuição dos clientes “comerciais” com base nas características da população que trabalha na região, mas não necessariamente residem nas mesmas. Em ambos os modelos, consideramos as distâncias físicas entre os setores censitários, bem como variáveis mercadológicas mencionadas anteriormente (canibalização, concorrência etc.).

O modelo de regressão múltipla em questão é dado por

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + e_i \quad i = 1, 2, \dots, n \quad (1)$$

onde:

- Y_i é a i -ésima observação da variável dependente; $X_{1i}, X_{2i}, \dots, X_{ki}$ são as i -ésimas observações sobre cada uma das k variáveis regressoras e e é o termo de erro;
- β_1 é o coeficiente de declividade de X_1 , β_2 é o coeficiente de declividade de X_2 etc. O coeficiente β_1 é a variação esperada em Y resultante da variação unitária em X_{1i} mantendo constantes $X_{2i}, X_{3i}, \dots, X_{ki}$. Os coeficientes dos outros X'_s são interpretados do mesmo modo;
- o intercepto β_0 é o valor esperado de Y quando todos os X'_s são iguais a zero;
- a suposição básica para (1) é que $e \sim \text{Normal}(\mathbf{0}, \sigma^2 I)$.

As principais especificações são:

- A **unidade de observação** utilizada é o par *setor-loja*, incluindo todos os setores censitários dentro do raio de N quilômetros, onde N refere-se à distância mínima necessária para cobrir pelo menos 60% dos clientes de cada loja.

- A **variável dependente** utilizada foi o faturamento gerado pelos clientes no setor *i* para a loja *j*. Com objetivo de diminuir os possíveis efeitos de sazonalidade, foi utilizada a média mensal de contribuição referente aos últimos seis meses de observação. Este tratamento foi idêntico em ambos os modelos: residencial e comercial.
- As **variáveis independentes** utilizadas incluíram características demográficas e econômicas das populações *residencial* e *comercial* (separadamente em cada modelo); características do mercado, como proximidade à concorrência ou outras lojas da mesma empresa; proximidade a outras unidades de varejo ou centros comerciais, shopping centers etc.

Ajuste para observação censurada

A principal dificuldade encontrada neste projeto foi a grande quantidade de setores censitários vazios, ou seja, setores onde nenhum indivíduo gerou qualquer transação na loja de influência nos últimos seis meses. Esta situação é comum em modelos geodemográficos deste tipo, pois, por maior que seja o grau de penetração de um produto ou serviço cuja demanda procura-se estimar, e considerando o baixo nível de agregação dos setores censitários (que incluem em média 300 domicílios em áreas urbanas), é de se esperar que não haja demanda observada em vários dos setores estudados. Em muitos casos, a proporção de setores sem demanda observada, ou seja, a proporção de zeros, pode chegar a 60 ou 70 por cento das observações válidas.

É evidente que uma quantidade relativa grande de “zeros” implicará um forte viés na estimativa dos coeficientes neste caso. O problema é conhecido, principalmente em estimativas de características populacionais cuja observação depende do comportamento estudado. O exemplo mais conhecido que apresenta esse tipo de complexidade refere-se a modelos que buscam estimar a renda potencial do trabalho da população feminina. Neste caso, como não é possível observar o rendimento ou horas trabalhadas por mulheres não participantes no mercado de trabalho, e por ser a decisão de trabalhar ou não dependente, entre outras coisas, do rendimento encontrado no mercado, observa-se uma censura (devido à auto-seleção) na variável dependente que impossibilita a estimativa consistente dos coeficientes.

Aparentemente, a solução mais simples para o problema de “excesso de zeros” seria simplesmente eliminá-los, ou seja, desconsiderar do modelo os setores com *zero* demanda. Mas sabemos que esta medida não resolve o problema, pois estaríamos apenas truncando a distribuição da variável dependente arbitrariamente e, conseqüentemente, violando a premissa básica para garantir consistência e eficiência em modelos OLS, de independência e uniformidade na dispersão do termo de resíduo.

A solução adotada procura modelar a parte da distribuição não observada a partir da parte observada da distribuição. Este modelo se tornou conhecido por **Tobit**, pois foi aplicado pioneiramente nas ciências sociais em 1958 pelo economista James Tobin, Tobin (1958), com o propósito de estimar o número de horas trabalhadas por mulheres. Desde então, o mesmo método, ou pequenas variações, foi aplicado em diversas situações, desde modelos de estimativa de número de casos extraconjugais (onde a censura em zero é evidente) até estimativas de “dias na prisão” por indivíduo. Em linhas gerais o modelo de regressão baseado em dados censurados é dado por

$$y_i^* = \beta'x_i + e_i \text{ para } i = 1, 2, \dots, n \text{ observações (2)}$$

onde

$$y_i = 0 \quad \text{se } y_i^* \leq 0$$
$$y_i = y_i^* \quad \text{se } y_i^* > 0.$$

Por simplicidade, a especificação estatística do modelo Tobit não é descrita neste capítulo, mas o leitor interessado encontrará uma discussão detalhada no capítulo 22 de Greene (1993) e em referências encontradas naquele capítulo.

Foram consideradas várias transformações e interações. Em modelos geodemográficos, é importante permitir que as principais variáveis sociodemográficas interajam livremente com as variáveis espaciais, como distâncias à loja ou aos concorrentes. O processo de especificação é complexo e fundamental para o sucesso do modelo. E a não linearidade, adicionada pelo ajuste do modelo Tobit, dificulta substancialmente a interpretação dos resultados. A seguir, são apresentadas algumas simulações descrevendo os efeitos de variáveis selecionadas, bem como resultados gerais dos modelos.

Resultados

Os Gráficos a seguir exemplificam o impacto da distância entre o setor e loja de faturamento, o impacto do aumento de famílias de classe A (critério Anep) e do efeito do número de concorrentes na área de influência. Nos três casos, os resultados foram gerados por simulações, considerando o efeito linear (OLS) e ajuste não linear do modelo Tobit. Devido ao grande número de interações utilizadas, foi necessário fixar os valores das demais variáveis em cada simulação. Nas simulações feitas, foram utilizados os valores médios das respectivas populações.

Gráfico 1 representa o impacto da distância entre um setor censitário e a loja de referência, na contribuição deste setor para o faturamento da loja. Os valores e escalas foram alterados para garantir a confidencialidade das informações, mas as formas funcionais foram preservadas. Conclui-se que o impacto da distancia na demanda é negativo e marginalmente decrescente, consistente com os modelos conhecidos como “gravitacionais”, bastante populares nas aplicações de gemorarketing. Estes modelos assumem que o impacto da distância entre consumidores e pontos de vendas, na demanda por serviços ou produtos será sempre negativo e exponencial, convergindo para zero.

Gráfico 1
Comportamento de aumento da distância no faturamento
modelo residencial

Efeito do aumento da distância no faturamento estimado

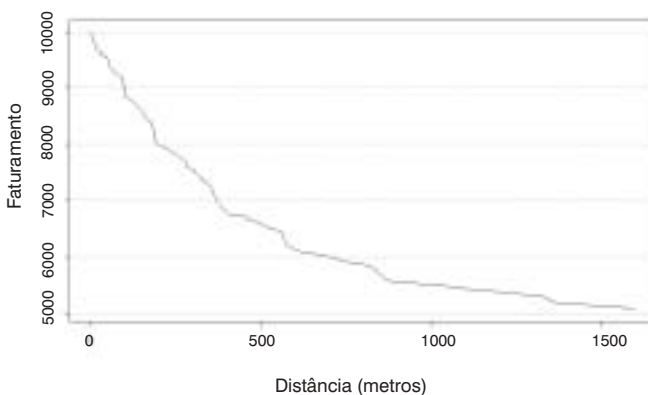
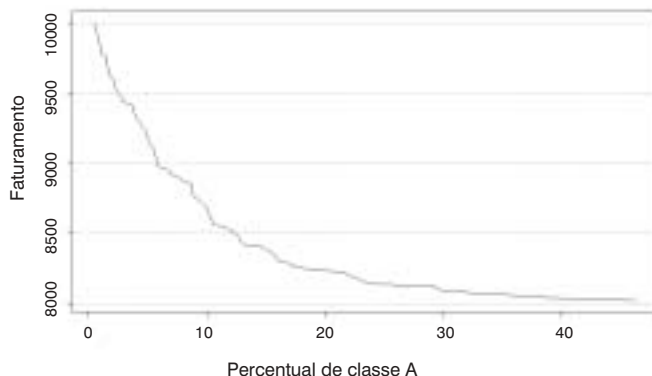


Gráfico 2 descreve a relação entre a proporção da população de alta renda (classe A) e demanda. O impacto negativo da fração de alta renda é

Gráfico 2

Comportamento do incremento da classe social no faturamento modelo residencial

Efeito do aumento do percentual de famílias de classe A no faturamento estimado



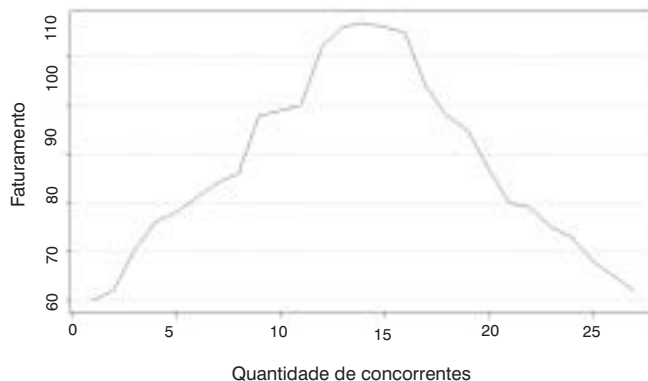
consistente com o posicionamento das financeiras populares, dirigidas para setores de menor renda.

Gráfico 3 apresenta o impacto da proximidade da concorrência na demanda do setor. Interessantemente, percebe-se que o efeito é positivo entre zero e poucos concorrentes, passando a ser negativo para quantidades de

Gráfico 3

Comportamento do incremento da quantidade de concorrentes no faturamento modelo residencial

Efeito do aumento de quantidade de concorrentes no faturamento estimado



concorrentes maiores. O resultado é consistente com a percepção de vários planejadores que, ter “alguma concorrência” nas proximidade é salutar para o negócio, pois aumenta a atratividade da região. A medida em que a quantidade de concorrentes na região aumenta, entretanto, o efeito negativo (resultado do rateio da demanda regional entre os concorrentes) passa a dominar.

As estimativas de faturamento total apresentados pelos modelos, quando comparados aos faturamentos reais observados apresentaram um erro absoluto médio de cerca de 6%, sendo que o modelo residencial apresentou um erro um pouco menor e o modelo comercial apresentou um erro de mais de 10%. Estes resultados variam, é claro, de caso a caso, mas pode-se esperar, em circunstâncias normais, variações em torno destes valores.

Potencialização de Mercado

Uma vez testados e validados os modelos iniciou-se o processo de “escoragem” de todos os setores censitários da área de estudo, ou seja, o processo de estimativa de potencial. Para tanto foram aplicados os modelos residencial e comercial em todos os setores censitários nas cidades estudadas. Importante ressaltar que o processo de estimativa de potencial de cada setor envolveu a estimativa da contribuição de todos os setores vizinhos (dentro de um raio pré-determinado) para aquele setor. Estas contribuições precisaram ser recalculadas para cada novo setor cujo potencial foi estimado, pois a cada mudança de setor de potencialização (ou seja, o setor onde se estuda abrir uma nova loja) faz-se necessário recalcular completamente a matriz de distâncias, e conseqüentemente as contribuições individuais dos setores inclusos na área de influência.

Em cada caso de estimativa de potencial por setor, o faturamento esperado foi calculado pelo somatório do faturamento estimado pelo modelo comercial e residencial em cada setor dentro da área de influência, e acrescentado a esta soma a estimativa de “captura externa”, ou seja, a contribuição esperada oriunda de todos setores que ficaram fora da área de influência considerada. A contribuição relativa da captura externa tende a ser relativamente constante entre lojas, em relação à contribuição dos setores dentro da área de influência, o que permitiu que usássemos uma “correção” uniforme (em termos relativos) para incluir a contribuição externa às áreas de influência de cada setor potencializado. Por fim,

os resultados foram agrupados por distritos ou bairros, quando disponíveis, possibilitando estimar o *ranking* dos melhores locais para abertura de novas lojas entre todas as áreas estudadas.

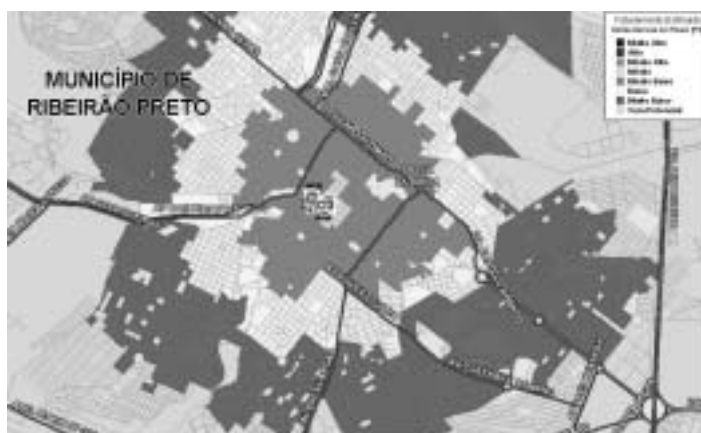
Recomendações Finais ao Cliente

A parte final deste trabalho foi a apresentação dos resultados dos modelos, indicando áreas com alto potencial de abertura de novas lojas.

Esta etapa leva em consideração outras características, não demográficas, que não entraram nos modelos estatísticos. Os modelos podem ter indicado como setor censitário de maior potencial uma área residencial próxima a alguns quarteirões de uma grande avenida, por exemplo. Nesta etapa, nosso papel foi de ajuste, ou seja, recomendar a melhor localização próxima aos setores de maior potencial apontados pelos modelos.

O Mapa 3 ilustra os contornos de uma macrorregião com seu respectivo potencial de faturamento.

Mapa 3
Macrolocalização – um local com potencial dos setores censitários
(mapa ilustrativo)



O passo seguinte foi apontar a melhor localização dentro desta região, levando em conta, além do potencial estimado, outras variáveis que não entraram no modelo, como acessos, visibilidade etc. O Mapa 4 ilustra o detalhe de uma região recomendada pela Cognatis.

Mapa 4
Localização recomendada com faturamento estimado dos setores censitários



Conclusão

É cada vez mais comum que empresas em diversas indústrias e áreas de atuação, utilizem métodos analíticos quantitativos até recentemente limitados aos redutos acadêmicos, em seus processos de decisões estratégicas e táticas. Neste texto discutimos as origens e tendências desta prática, dando destaque a dois campos de aplicação: o CRM analítico e geomarketing. Em ambos destes campos, profissionais com formação e experiência em análise demográfica podem fazer contribuições importantes. A aproximação formal e estruturada entre o mundo acadêmico e o de negócios é de fundamental importância para a continuidade do desenvolvimento destes campos de aplicação, por um lado, e aumento do leque de oportunidades para os profissionais em demografia por outro.

O estudo de caso apresentado ilustrou como aplicações de geomarketing podem se beneficiar se integradas a técnicas de modelagem estatística para estimar mercados potenciais. Entre as principais dificuldades encontradas a maior foi a

existência de diversos setores censitários sem informações transacionais, caracterizando demanda censurada. O método escolhido para lidar com esta dificuldade é baseado no modelo Tobit, que corrige distribuições de variáveis dependentes censuradas por baixo, resultando em um modelo bastante satisfatório, com baixo erro absoluto médio (estimado – observado). Esperamos ver no futuro próximo a crescente popularização na aplicação de métodos como este, ou similares, com o propósito de apoiar com “inteligência” objetiva e independente as decisões estratégicas que empresas competitivas precisam tomar para prosperar em um mundo de negócios cada vez mais “cientificado”.

Referências

BATEY, P.; BROWN, P. From Human ecology to customer targeting: the evolution of geodemographics. In: LONGLEY, P.; CLARKE, G. (Eds.) *GIS for Business and Service Planning* New York: Wiley & Sons, 1995.

BATTY, M.; SHIODE, N. Population growth dynamics in cities, countries and communication systems. In: LONGLEY, P. A.; BATTY, M. (Eds.). *Advanced Spatial Analysis*. Redlands: ESRI Press, 2003.

BERRY, M. J. A.; LINOFF, G. S. *Mastering Data Mining* London: Wiley & Sons, 2000.

BIRKIN, M.; CLARKE, G.; CLARKE, M. *Retail Geography & Intelligent Network Planning* London: Wiley & Sons, 2002.

DAVID, M.; LONGLEY, P. Data sources and their geographical integration. In: LONGLEY, P.; CLARKE, G. (Eds.). *GIS for Business and Service Planning* New York: Wiley & Sons, 1995.

DOWD, Mr. R.; LE SAGE, J. P. Analysis of Spatial Contiguity Influences on State Price Level Information. *International Journal of Forecasting* v. 13, p. 245-253, 1997.

GREENE, R. P.; STAGER, J. C. Techniques and Methods of GIS for Business. In: PICK, J. B. (Eds.). *Geographic Information Systems in Business*. London: Idea Group, 2005.

GREENE, W. H. *Econometric Analysis*, Prentice Hall. New York: Prentice Hall, 1993.

LONGLEY, P. A.; CLARKE, G. Applied geographical information systems: developments and prospects. In: LONGLEY, P.; CLARKE, G. (Eds.). *GIS for Business and Service Planning* New York: Wiley & Sons, 1995.

MARTIN, D. Censuses and the modeling of population in GIS. In: LONGLEY, P.; CLARKE, G. (Eds.). *GIS for Business and Service Planning* New York: Wiley & Sons, 1995.

OPENSHAW, S. Marketing spatial analysis: a review of prospects and technologies relevant to marketing. In: LONGLEY, P.; CLARKE, G. (Eds.). *GIS for Business and Service Planning* New York: Wiley & Sons, 1995.