

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS

Rodrigo Ledur da Silva

**MODELO DE ATRATIVIDADE GEOGRÁFICA NO SETOR DE MATERIAIS
ELÉTRICOS**

Porto Alegre

2017

Rodrigo Ledur da Silva

**MODELO DE ATRATIVIDADE GEOGRÁFICA NO SETOR DE MATERIAIS
ELÉTRICOS**

**Trabalho de conclusão de curso de graduação,
apresentado ao Departamento de Ciências
Administrativas da Universidade Federal do
Rio Grande do Sul, como requisito parcial para
a obtenção do grau de Bacharel em
Administração.**

Orientador: Vinícius Brei

Conceito Final:

Aprovado em ____ de _____ de _____.

BANCA EXAMINADORA

Orientador Prof. Vinicius Brei - UFRGS

AGRADECIMENTOS

À minha família, meu principal agradecimento pelo apoio em todos os momentos. À Carla e ao professor Brei, obrigado pelos ensinamentos e por me mostrarem o que é a área de pesquisa – uma área que não tinha contato e que, hoje, vejo com muito respeito e interesse. Ao professor Pedro Albuquerque, da Universidade de Brasília, também agradeço a cooperação com seu conhecimento avançado sobre o tema. À PS Júnior, obrigado por me mostrar a verdadeira Escola de Administração e por me proporcionar todas as experiências que eu não viveria se não tivesse passado por aquela salinha. A todos os meus amigos, obrigado pelos melhores momentos vividos na UFRGS e, claro, fora dela também. Com certeza, os anos de Administração foram os meus anos de maior aprendizado e que me fizeram ser a pessoa que sou hoje.

RESUMO

Este trabalho foi desenvolvido com o objetivo de identificar as regiões mais atrativas para a venda de interruptores elétricos, bem como as variáveis que possam ter impacto nessa atratividade. Aqui, a atratividade foi mensurada pelas vendas anuais totais de materiais de construção da cidade de São Paulo e, para análise de comparação foram delimitados os 96 distritos da cidade. Portanto, foram investigadas variáveis referentes a 6.790 pontos comerciais da cidade e, juntamente com dados demográficos e de lançamentos residenciais e comerciais, foi elaborada uma base de dados final através de técnicas de georreferenciamento. Utilizou-se sobretudo de análises descritivas e da análise de regressão linear múltipla no software livre RStudio para que fosse proposto um modelo final de atratividade. Os resultados encontrados indicam que existe significância estatística no modelo composto por quatro variáveis – o número de lojas no distrito, o número de funcionários, a presença de grandes redes e o número de domicílios do distrito. Como proposta final, foi elaborado um índice de atratividade através das previsões do modelo, e um mapa final de atratividade ilustrou os resultados. Por fim, os distritos mais atrativos para a venda de interruptores foram identificados de acordo com o modelo final.

Palavras-chave: Atratividade; georreferenciamento; vendas; interruptores elétricos.

ABSTRACT

This study aims to identify the most attractive regions for the commercialization of electric switches, as well as the variables impacting this attractiveness. Here, attractiveness was measured by the annual revenue of construction materials in the city of São Paulo and, for comparison, the 96 city districts were delimited. Therefore, variables referring to 6.790 outlets were investigated and, herewith demographic and commercial and residential launchings data, it was developed a final database, throughout georeferencing techniques. Descriptive analysis and multiple linear regression were performed in the open source software RStudio so that a final attractiveness model was proposed. The results found suggests that there is statistical significance in the model composed by four variables – the number of outlets in the district, the number of employees, the presence of large chains and the number of residences. As a final proposition, it was developed a composite index of attractiveness through the model outcome predictions and a final attractiveness map illustrated the results. Finally, the most attractive districts of São Paulo for the commercialization of electric switches were identified according to the final model.

Palavras-chave: Attractiveness; georeferencing; revenue; electric switches.

LISTA DE ILUSTRAÇÕES

Figura 1 - Hierarquia dos Centros de Comércio.....	12
Figura 2 - Equação Modelo de Huff.....	14
Figura 3 - Equação Modelo de Huff aplicado em Huff 1962.....	15
Figura 4 - Equação 2 Modelo de Huff.....	16
Figura 5 - Vendas do Estabelecimento j.....	16
Figura 6 - Estrutura do Geocódigo do IBGE.....	23
Figura 7 – Número de Lojas por Distritos da Cidade de São Paulo.....	31
Figura 8 – % de Lojas sem Informação de Faturamento – Total por Distrito.....	33
Figura 9 - Mapa de Densidade por Número de Domicílios.....	42
Figura 10 - Mapa de Densidade por Rendimento Mensal Médio	43
Figura 11 - Mapa de Densidade por Número de Lançamentos	44
Figura 12 – Mapa de Atratividade do Modelo Final (à esq.) e Mapa Real (à dir.)	56
Quadro 1 – Variáveis das bases de dados (continua)	17
Quadro 2 – Códigos CNAE da Base de Dados	19
Quadro 3 - Variáveis do Modelo Proposto.....	27
Gráfico 1 - % de Presença das Categorias em Lojas de Materiais de Construção em Geral	21
Gráfico 2 – Número de Lojas por Faixa de Faturamento Anual	29
Gráfico 3 – Distribuição do Faturamento Total dos Distritos (‘000)	30
Gráfico 4 – Relação entre o Faturamento Total e o Número de Domicílios dos Distritos	34
Gráfico 5 – Relação entre o Faturamento Total e Número de Moradores dos Distritos	35
Gráfico 6 – Relação entre o Faturamento Total e o Número de Funcionários nos Distritos	36
Gráfico 7 – Relação entre o Faturamento Total e o Número de Lojas dos Distritos.....	37
Gráfico 8 – Relação entre o Faturamento Total e o Rendimento Médio Mensal nos Distritos	38
Gráfico 9 – Matriz de Correlação entre as Variáveis	39
Gráfico 10 - Número de Comércio por CNAE na Cidade de São Paulo	45
Gráfico 11 - Número de Pontos de Venda por Índice de Atratividade.....	57

LISTA DE TABELAS

Tabela 1 - Coordenadas de 10 Lojas da Base de Dados	22
Tabela 2 – Características Descritivas das Variáveis	32
Tabela 3 – Faturamento Anual Total, Rendimento e Domicílios por Distrito	47
Tabela 4 – Resumo do Modelo Final	50
Tabela 5 – Coeficientes do Modelo Final.....	52
Tabela 6 – Comparação entre Scores Reais e Previsão do Modelo.....	53
Tabela 7 - Índice de Atratividade por Distrito.....	57

SUMÁRIO

1. INTRODUÇÃO	1
2. JUSTIFICATIVA	3
3. OBJETIVOS	5
3.1. OBJETIVO GERAL	5
3.2. OBJETIVOS ESPECÍFICOS	6
4. REVISÃO DA LITERATURA	6
4.1 O IMPACTO DA LOCALIZAÇÃO NAS VENDAS E NO MARKET SHARE	7
4.2 MODELOS ESPACIAIS DE ATRATIVIDADE.....	11
4.2.1 Teoria do Lugar Central	11
4.2.2 Modelos Gravitacionais	13
5. PROCEDIMENTOS METODOLÓGICOS	17
5.1 COLETA E SELEÇÃO DE DADOS	19
5.2 CONSTRUÇÃO DO MODELO INICIAL DE ATRATIVIDADE	24
5.3 PREPARAÇÃO E EXAME INICIAL DA BASE DE DADOS	28
5.4 PLANO DE ANÁLISE DE DADOS.....	40
6. ANÁLISE DOS RESULTADOS	41
6.1 CARACTERÍSTICAS GERAIS.....	41
6.2 MODELO DE ATRATIVIDADE FINAL	49
7. CONSIDERAÇÕES FINAIS	60
8. LIMITAÇÕES DO TRABALHO	62
REFERÊNCIAS	63
APÊNDICES	65

1. INTRODUÇÃO

A importância da conveniência por parte do consumidor vem crescendo devido a fatores como o aumento da sua conscientização, da abundância de opções, a busca pela mobilidade e o surgimento de novos estilos de vida (REIMERS E CLULOW, 2004, p. 207). Assim, para as empresas se faz essencial entender onde focar seus esforços de maneira que o seu público-alvo seja atingido e, para tal, é necessária a informação sobre este público e do meio em que ele está inserido. A medida em que se aprofunda este tema, vemos também que a busca pela distribuição ótima ou pela localização mais atrativa, no caso de um varejista, se torna mais complexa.

Ao nos referirmos à intensidade de distribuição, por exemplo, é importante considerar a categoria em que o produto se enquadra. Para bens de conveniência, maiores níveis de distribuição tendem a elevar as vendas sempre (COUGHLAN ET AL., 2006, p. 114 apud BUCKLIN, SIDDARTH, SILVA-RISSO, 2008 - pode-se mencionar a Coca-Cola como um grande caso de líder de mercado que tem sua força baseada na forte distribuição. Já no caso de bens duráveis como carros e eletrônicos, a relação pode não ser tão clara, visto que outras variáveis entram em cena, como menciona Bucklin et al (2008, p. 474):

Para bens duráveis, compradores podem procurar mais informação sobre os produtos antes de efetuar a compra. O papel do revendedor vai além de fornecer conveniência de espaço e sortimento, mas também incluir suporte de marketing (assistência de vendas, demonstrações de produto) (BUCKLIN, SIDDARTH, SILVA-RISSO, 2008, p. 474)

Desse modo, não só a intensidade de distribuição, mas a localização do ponto de venda onde o produto é ofertado muitas vezes é fundamental. Quando se trata do varejista, a localização possui um papel central no sucesso (CHAN et al. 2007, p. 622) e talvez seja o principal elemento do mix de marketing. Decisões relacionadas ao local onde abrir uma filial ou em qual região direcionar investimentos de logística e distribuição impactam diretamente na performance da empresa. Na Teoria do Lugar Central desenvolvida por Lösch (1954) e Christaller (1933) e uma das bases da teoria da estrutura espacial, é desenvolvida a ideia de zona de influência do bem ou produto:

Devido aos crescentes custos de transporte, a demanda de um produto diminui conforme aumenta a distância entre o mercado consumidor e a localização do consumidor; a partir de uma certa distância a demanda cai a

zero; a distância máxima se chama zona de influência ou amplitude do bem. (BROWN, 1992, BEAVON, 1977 apud ARANHA, 2001)

Considerando-se a premissa levantada pelos autores, se faz importante conhecer o perfil do consumidor e os atributos que contribuem para a demanda potencial de uma localidade. Além disso, vem à luz o fato de que é possível que duas filiais, por exemplo, estejam na mesma zona de influência, acarretando de certa forma numa competição entre os próprios pontos de venda. Aprofundar essa questão, pode nos levar a perguntas ainda mais complexas: promoções em um varejo aumentariam a zona de influência do mesmo? Deve-se escolher uma região de alto potencial de demanda mesmo que isso acarrete estar localizado próximo de um concorrente? Essas questões são importantes tanto para um varejista quanto para um fabricante e podem se tornar mais estratégicas de acordo com o produto em questão.

Diversos estudos de elasticidade foram conduzidos no que se diz respeito ao impacto do preço e da propaganda no market share e nas vendas se comparado a estudos do impacto da distribuição (BUCKLIN, SIDDARTH, SILVA-RISSO. 2008, p. 473). Já sobre a localização do ponto de venda, Titton et al. (2010, p. 33) ressaltam que “há uma escassez de matéria sobre o assunto, obrigando a referenciar estudos muito antigos, apesar da relevância do tema localização para as organizações em geral”. Em estudo voltado para o mercado de postos de gasolina em Cingapura, Chan et al. (2007), através de um modelo econométrico de decisões sobre preço e localização, conseguiram quantificar a dependência da demanda local por características demográficas e, a partir disso, determinar as regiões de maior potencial de demanda. Titton et al. (2010, p. 39) ao estudar a correlação entre variáveis espaciais e o faturamento das lojas de uma rede de drogarias, apresentam que variáveis tidas como importantes para o varejo como um “alto fluxo de pessoas e a movimentação dos carros em baixa velocidade não tiveram correlação significativa e poderia indicar que não se tratam de variáveis a serem analisadas do ponto de vista quantitativo”.

Há de se acrescentar o papel da concorrência e a concentração do mercado, que impactam diretamente os novos entrantes. No caso de franquias, existem, inclusive, cláusulas contratuais que dizem respeito ao raio de distância entre um ponto de venda e outro. Portanto, questões a respeito da localização e seu potencial de demanda podem ser complexas e, se bem analisadas, impactar positivamente na performance da empresa - fabricante ou varejista.

Devido à relevância deste tema, o presente trabalho busca estudar o impacto da localização geográfica nas vendas, na perspectiva dos fabricantes e sua distribuição. Mais especificamente, o intuito é definir **qual a região de maior atratividade para as vendas dos fabricantes no setor de material elétrico?** O estudo é direcionado para a venda de interruptores com foco na cidade de São Paulo, escolha derivada da parceria existente com a empresa Tramontina, que facilitou e enriqueceu a pesquisa através do fornecimento de dados do setor sobre a região estudada.

2. JUSTIFICATIVA

Hoje em dia, no mercado competitivo, as barreiras geográficas foram superadas pelos ganhos de escala e esforços logísticos permitidos pelo progresso tecnológico. Nesse cenário, a concorrência em todos os níveis vem se tornando cada vez mais intensa e os players do mercado disputam por cada detalhe no que se refere a canais de venda e a obtenção de vantagens competitivas e redução de custos.

Apesar de um crescimento do e-commerce, a loja física ainda é de longe o canal mais importante e, para muitos segmentos, o único meio possível de se atingir o consumidor, principalmente em países em desenvolvimento como o Brasil. Também, zonas centrais das grandes cidades sofrem com o crescimento da população e do tráfego diário, tornando a localização e a proximidade geográfica um fator muito importante para empresas atingirem seu público alvo. Do ponto de vista dos fabricantes, uma distribuição “inteligente” é aquela que fornece o produto nos pontos de venda de maior representatividade, portanto se faz essencial identificá-los, bem como as regiões de maior atratividade.

O presente estudo vai ao encontro desse contexto e busca esclarecer o que constitui a atratividade de uma região, bem como as variáveis de maior impacto na mesma. Utilizando como plano de fundo o setor de material elétrico, poder-se-á identificar oportunidades de distribuição, do ponto de vista de um fabricante, além de identificar, de fato, fatores que influenciam na sua performance de vendas. De maneira geral, esse conhecimento ainda é negligenciado e tem carência de uma análise mais aprofundada, visto que que variáveis tidas como muito importantes nem sempre são, de fato, diferenciais de performance (TITTON ET AL., 2010).

Portanto, desenvolver um método para identificação das regiões de maior atratividade para as vendas se faz de grande valor. Ressalta-se que o foco no setor elétrico de São Paulo é uma oportunidade que a cátedra Tramontina Eletrik do Grupo de Pesquisa de Marketing e Consumo viabiliza, à medida que a parceria fornece dados que ajudam a tornar o estudo mais rico. Acredita-se que a contribuição do estudo será para uma tomada de decisão mais estratégica das empresas no que se refere ao *geomarketing* e que se aprofunde cada vez mais o impacto desse tipo de decisão nos resultados das empresas e também do setor público.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Desenvolver um índice de atratividade que determine as melhores regiões para venda de interruptores elétricos para fabricantes na cidade de São Paulo.

3.2. OBJETIVOS ESPECÍFICOS

- I. Identificar as variáveis que possam ter impacto na atratividade de uma região;
- II. Testar a relevância estatística e substantiva de cada variável;
- III. Desenvolver um modelo para calcular a atratividade de uma região.

4. REVISÃO DA LITERATURA

Nesta seção serão revisitados os estudos que formam o esqueleto do assunto abordado neste presente trabalho. A seção foi dividida em dois tópicos principais: o

primeiro fazendo uma revisão do que já foi pesquisado e estimado a respeito do impacto da localização nas vendas e no *market share*; após, será dada uma aprofundada nos modelos espaciais, mais especificamente nos modelos gravitacionais e modelo de Huff, que constituem a base inicial do presente trabalho.

4.1 O IMPACTO DA LOCALIZAÇÃO NAS VENDAS E NO MARKET SHARE

De acordo com Chan, Padmanabhan e Seetharaman (2007, p. 622), a variável localização é “reconhecida na literatura acadêmica como um importante determinante para a performance e competição no varejo”. Essa importância também é reconhecida quando olhamos do ponto de vista do fabricante. Bucklin, Siddarth e Silva-Risso (2008) ao analisarem o impacto da intensidade de distribuição na decisão de compra de um novo carro, mostram que variáveis como a acessibilidade do revendedor, ou seja, estar próximo do comprador, apresenta significativa relevância na sua tomada de decisão. Os autores também destacam que “diferente de preço, publicidade e promoção, a intensidade de distribuição geralmente varia de maneira lenta ao longo do tempo” (BUCKLIN; SIDDARTH E SILVA-RISSO, 2008, p. 474), o que contribui para a importância de uma maior atenção nesse fator. Do ponto de vista de uma loja ou rede, a localização ganha ainda mais relevância, “pois um reposicionamento sempre irá demandar um esforço logístico e de estratégia muito grande (TITTON ET AL., 2010, p. 39). Cheng, Li e Yu (2007), ilustra essa situação no contexto de projetos de construção:

Projetos de construção envolvem uma grande quantia de investimento de capital de longo prazo [...]. Nesse quesito, cada parte do ciclo de vida de um projeto deve ser cuidadosamente determinado antes de ser empreendido. Devido à complicada natureza de um projeto de construção, a tomada de decisão se torna um processo importante, mas difícil. Entre as numerosas decisões a serem tomadas, uma decisão central que precisa ser realizada é sempre subestimada, que é a seleção da localização. (CHENG; LI; YU, 2007, p. 884)

Nesse cenário, estudos da área buscam entender os mecanismos e fatores envolvidos na atratividade de uma determinada localização perante outra. Dolega, Pavlis e Singleton (2016, p. 78) trazem o conceito de zona de influência, que “pode ser definida como a extensão de uma área onde os principais usuários de uma loja ou centro comercial serão tipicamente encontrados”. No entanto, Chan, Padmanabhan e Seetharaman (2007) trazem a importante definição de localização como um atributo horizontal:

Localização é um atributo horizontal (diferentes consumidores iriam discordar sobre o que seria a localização ideal de acordo com o local onde residem), diferentemente de atributos verticais (todos os consumidores iriam concordar que um preço menor é preferível à um preço maior, maior qualidade é preferível à menor qualidade, etc. (CHAN; PADMANABHAN; SEETHARAMAN, 2007, p. 623)

As zonas de influência, portanto, são relativas à localização do consumidor - estar perto de um determinado público, significa estar distante de outro. Inclusive, Chan, Padmanabhan e Seetharaman (2007) ao desenvolverem seu modelo econométrico para localização geográfica e preço no setor de postos de gasolina, mostram que devido à essa limitação na abrangência das zonas de influência, existem áreas que seriam mais lucrativas devido a um maior potencial de demanda no local.

A decisão de onde localizar um produto ou ponto de venda, então, é de fundamental interesse do tomador de decisão, seja ele um órgão público ou privado. Nesse contexto, Teller e Reutterer (2008) avaliaram a atratividade das aglomerações ou clusters de comércio, que podem ser exemplificadas por aglomerações planejadas (ex.: shopping centers) ou aglomerações não planejadas (ex.: ruas ou avenidas comerciais). Independentemente do tipo, um centro comercial pode trazer valor adicional, do ponto de vista do consumidor, se comparado à um ponto de venda único. Facilidades como estacionamento, variedade de serviços e produtos e entretenimento têm consigo um enriquecimento na experiência de compra (KIM ET AL, 2002 apud TELER; REUTTERER, 2008). Já Reimers e Clulow (2004) argumentam sobre o importante atributo da conveniência no processo de compra, cujo crescimento se dá por tendências sociais como o aumento do estilo de vida individual e fatores como a escassez de tempo nos dias de hoje. Teller e Reutterer (2008) ponderam que apesar da cooperação entre pontos de venda na convivência da mesma infraestrutura ou ambiente, os mesmos competem pelo *share of wallet* e tempo dedicado pelos visitantes.

Ao trazer a discussão para um nível mais amplo, alguns estudos mudaram o foco de modelos que previam a localização ideal para abertura de novas lojas para um foco em modelos de mensuração de atratividade entre centros comerciais (FINN; LOUVIERE, 1996; DENNIS; MARSLAND; COCKETT, 2002; CHENG; LI; YU, 2007; TELLER; REUTTERER, 2008; DOLEGA; PAVLIS; SINGLETON, 2016). Dolega, Pavlis e Singleton (2016), ao desenvolverem um modelo de atratividade a nível nacional no Reino Unido, ressaltam que a maioria dos modelos é estimada num contexto geograficamente limitado, o que geralmente seria uma suposição falsa no

mundo real. No entanto, os autores ressaltam que estudos voltados para analisar zonas de influência entre diferentes centros comerciais encaram certas limitações se comparados a estudos com enfoque em negócios individuais. De acordo com os autores, para se estimar a abrangência de uma zona de influência de centros comerciais são necessárias algumas generalizações:

Zonas de influência de diferentes lojas devem variar substancialmente visto que consumidores iriam tipicamente percorrer maiores distâncias para comprar bens de comparação oferecidos por centros comerciais maiores comparado à bens de consumo, geralmente disponíveis localmente (DOLEGA; PAVLIS; SINGLETON, 2016).

Na Teoria do Lugar Central, um dos estudos base que trouxe implicações para os estudos de localização, Christaller (1933) e Lösch (1940), desenvolveram o conceito de hierarquia dos centros de comércio - a organização espacial dos diversos níveis ofertantes sobrepostos. Segundo Christaller (1933) apud Dennis, Marsland e Cockett (2002, p. 187), “para cada bem em particular, o limite da sua abrangência é definido pela distância que os consumidores estão dispostos a percorrer para adquiri-lo”. Apesar de sua reconhecida influência, a teoria, baseada em pressupostos microeconômicos de livre concorrência (ARANHA; FIGOLI, 2001) possui sérias limitações quando se retrata o cenário atual do mercado.

Birkin, Clarke e Clarke (2010) resumem que para a previsão de atratividade de uma localização, são necessários três elementos: demanda, oferta e interação entre os consumidores. Um dos modelos mais duradouros para estimativa de zonas de atratividade é o modelo de Huff (que será melhor explicado nos tópicos a seguir) introduzido em 1962 pelo próprio Huff, que utilizou dois fatores determinantes para o potencial de atratividade de uma loja: o seu tamanho em metros quadrados (diretamente proporcional) e a distância entre as lojas e os consumidores (inversamente proporcional). Além da aplicação original, o modelo já teve diversas extensões incluindo preço e níveis de serviço, horas de atendimento e imagem da loja como atratividade adicional (TELLER; REUTTERER, 2008).

Dolega, Pavlis e Singleton (2016) utilizam o modelo de Huff ao estimarem as zonas de atratividade de centros de comércio a nível nacional. Através da distância e de um índice de atratividade para cada centro, composto por fatores como o tamanho da oferta (S_j) subtraído pelas unidades vacantes (V_j) (pontos sem lojas), diversidade do mix de lojas (R_{mix}), proporção de unidades de lazer (L) e lojas âncora (A), os autores

propuseram um mapeamento das zonas de influência dos centros comerciais na Inglaterra e País de Gales, elencando a hierarquia entre os mesmos. O modelo proposto, como etapa inicial é descrito como:

$$A = S_j - V_j + R_{mix} + L + A \quad (1)$$

Os autores ponderam que, “métodos de distância e tempo de direção são geralmente considerados os mais válidos para cenários de lojas de conveniência” (DOLEGA; PAVLIS; SINGLETON, 2016, p. 79), onde a proximidade é fator prioritário na busca.

Chan, Padmanabhan e Seetharaman (2007) em seu modelo econométrico para localização geográfica e preço no setor de postos de gasolina, estima as zonas de maior potencial de demanda por um método de mínima distância, propondo por fim um modelo de preço para os postos de gasolina. O estudo mostrou que variáveis como renda média, número de motoristas, população, proximidade de aeroporto, centro e avenidas impactam positivamente no potencial de demanda. Os autores também conseguiram estimar o market share das companhias petrolíferas, ilustrando que o mesmo é impactado negativamente pelo custo de deslocamento e preço da gasolina. Tais achados também trazem à luz o conceito de âncoras, locais que podem impactar diretamente na atratividade e, conseqüentemente, nas vendas de uma loja (TELLER; REUTTERER, 2008), como um hospital para uma farmácia próxima; e na imagem da loja, como uma loja de vestuário especializada dentro um centro comercial (FINN; LOUVIERE, 1996).

Titton et al. (2010), ao analisarem a correlação entre faturamento e variáveis espaciais no contexto de uma rede de drogarias, ressaltam a possibilidade de influência de âncoras. Através da análise da demografia local, variáveis de fluxo e proximidade, constatou-se que “a presença em esquina e proximidade de comércio intenso têm correlação significativa com o faturamento” (TITTON ET AL., 2010, p. 39). No entanto, talvez o mais interessante foi que muitas variáveis consideradas importantes, como o alto fluxo de pessoas não apresentaram correlação significativa.

Trazendo uma abordagem no nível desagregado, com a localização exata de cada consumidor, Bucklin, Siddarth e Silva-Risso (2008) desenvolveu um modelo de intensidade de distribuição para montadoras baseando-se em três critérios: (1) acessibilidade do revendedor (distância entre o comprador e o revendedor mais

próximo), (2) concentração do revendedor (número de revendedores próximos de um comprador) e (3) difusão do revendedor (dispersão dos revendedores em relação ao comprador). Baseando-se em uma abordagem de utilidade e probabilidade, os autores comprovaram que o *market share* exibe retornos marginais menores na medida em que a intensidade de distribuição aumenta. Também, mostraram que a elasticidade de escolha pela intensidade de distribuição tem média de aproximadamente .6 - uma importante visão para o setor de consumo de bens duráveis.

4.2 MODELOS ESPACIAIS DE ATRATIVIDADE

De acordo com Aragão (2005, p. 19), “um modelo espacial de mercado descreve um fenômeno através da incorporação da dimensão espacial para auxílio à tomada de decisão pelos analistas de negócios”. Teller e Reutterer (2008) dividem as bases conceituais dos estudos espaciais de atratividade em três linhas: (1) abordagens com base na teoria da interação espacial, conhecidas como ‘modelos de gravidade’; (2) modelos de atratividade de centros de comércio baseados na teoria da randômica; e (3) modelos multiplicativos de interação competitiva. Nesta seção pretende-se dar enfoque nos modelos gravitacionais, mais especificamente no modelo de Huff, descrevendo anteriormente a Teoria do Lugar Central devido à sua importância como base teórica para os modelos existentes.

4.2.1 Teoria do Lugar Central

A Teoria do Lugar Central, desenvolvida por Lörsch e Christaller, “já foi considerada a teoria da estrutura espacial [do mercado] mais inovadora e de maior sucesso” (ARANHA; FIGOLI, 2001, p. 7). De acordo com Aragão (2005, p. 23), “a principal característica deste modelo espacial é a sua natureza qualitativa”, sem nenhum modelo matemático atrelado a mesma. Seus pressupostos são descritos por Brown (1992) e Beavon (1977) *apud* Aranha e Figoli (2001, p. 8) como:

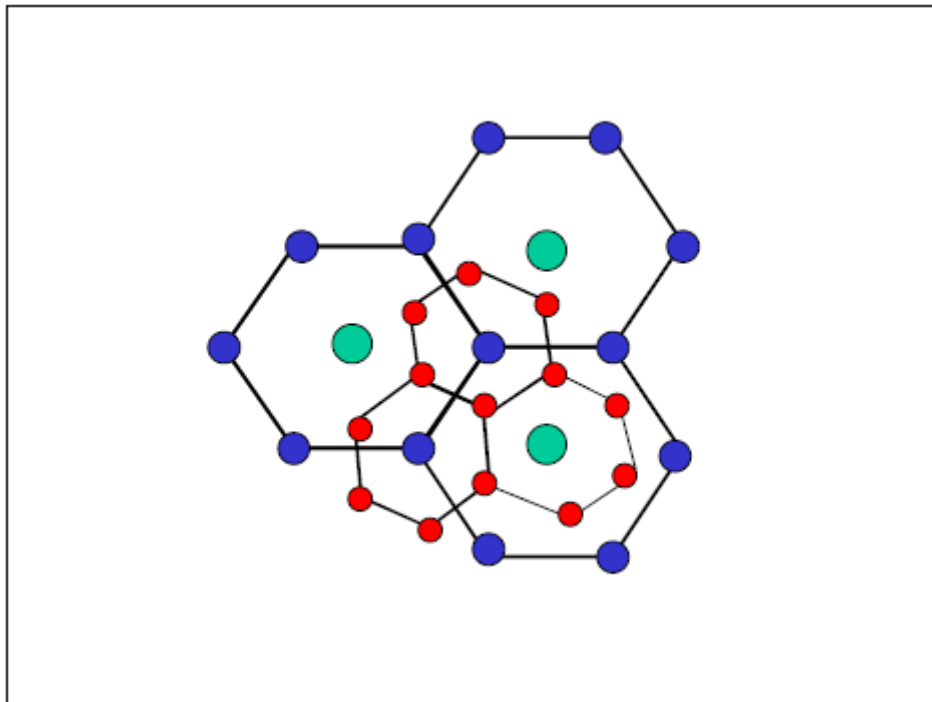
- Quanto aos consumidores, assume-se que são pequenos, igualmente afluentes, perfeitamente bem informados, racionais e tomam decisões buscando maximizar sua utilidade;

- Quanto aos fornecedores, admite-se que são pequenos, racionais e tomam suas decisões buscando maximizar seu lucro; vendem *fob*¹ e operam com custos equivalentes num ambiente em que o capital é móvel e não há barreiras de entrada;
- Quanto à geografia, postula-se um espaço hegemônico, em que os custos de transporte são uniformes em todas as direções e onde tanto os consumidores quanto os fornecedores estão uniformemente distribuídos; finalmente, com relação ao comportamento geográfico dos consumidores, pressupõe-se que fazem expedições de compras com objetivo único, comprando um único produto na localização mais próxima que o ofereça.

A partir destes pressupostos, conclusões como a zona de influência e a hierarquia dos centros de comércio foram definidos. O consumidor, então estaria disposto a percorrer distâncias maiores quando a mercadoria é de maior valor, buscando melhores preços e qualidade (ARAGÃO, 2005).

Figura 1 - Hierarquia dos Centros de Comércio

¹ FOB é a abreviatura da frase inglesa “Free On Board”, que significa “Posto a Bordo”. Nas operações mercantis de compra e venda, a cláusula FOB atribui as despesas decorrentes do transporte por conta do comprador.



Fonte: ARAGÃO (2005)

A teoria apresenta sérias limitações para se estabelecer zonas de influência, segundo Dolega, Pavlis e Singleton (2016, p. 79), como “a afirmação de uma distribuição relativamente uniforme da população e, portanto, uma estática distribuição de bens”. No entanto, Tang e Ingene (2000) apud Dennis, Marsland e Cockett (2002) observaram na prática a hierarquia dos centros de comércio, em seu estudo realizado em Shanghai, o que mostra que, de certo modo, a teoria ainda permanece influente e atual.

4.2.2 Modelos Gravitacionais

Segundo Aragão (2005, p. 30), os modelos gravitacionais de interação espacial são uma evolução da Teoria do Lugar Central, visto que “consideram, além da distância, outros fatores que influenciam a escolha do consumidor”. Segundo Commenges e Giraud (2016, p. 1), “Os modelos gravitacionais foram aplicados em diversas áreas (geografia, economia, demografia) para modelar uma grande variedade de fluxos (deslocamento, comércio, migração)”. O termo usado para nomear os modelos segue a metáfora da física, no qual os fluxos são considerados a força gravitacional entre dois

corpos e a influência do local, o potencial gravitacional (COMMENGES; GIRAUD, 2016)

Como explicam Aranha e Figoli (2001, p. 27), “a utilização de modelos gravitacionais para a localização de pontos comerciais e previsão de vendas visa estimar como o potencial de gastos dos consumidores será atraído para cada ponto de rede varejista”. Esses modelos são calibrados por variáveis da localização, como população, preço e produtos oferecidos (DOLEGA; PAVLIS; SINGLETON, 2016), dependendo do contexto utilizado. Dentre os diversos modelos gravitacionais, dar-se-á enfoque no modelo de Huff, devido à sua importância ao longo do tempo e sua aplicabilidade no presente estudo.

4.2.2.1 Modelo de Huff

O modelo de Huff, proposto em 1962 por David Huff, “considera que a força de atração exercida sobre os consumidores por um determinado centro comercial é baseada em fatores probabilísticos” (ARAGÃO, 2005, p. 26). De acordo com Huff e McCallum (2008), o modelo adquiriu sua longevidade de aplicação devido à três fatores principais:

Primeiro, o modelo é conceitualmente atrativo. A base lógica do modelo faz sentido e seu resultado pode ser comunicado facilmente e de maneira entendível. Segundo, o modelo é relativamente fácil de ser operacionalizado. As computações são diretas assim que os valores das variáveis e parâmetros são especificados. A terceira razão de sua popularidade é sua aplicabilidade em um largo campo de problemas [...] (HUFF; MCCALLUM, 2008, p. 2)

Sua premissa básica é de que a probabilidade de escolher uma alternativa em particular é proporcional à utilidade percebida da mesma. Portanto, “é improvável que qualquer alternativa seja selecionada exclusivamente a não ser que outras alternativas não existam” (HUFF, 2003, p. 34). A fórmula no seu nível mais generalista, pode ser visualizada abaixo:

Figura 2 - Equação Modelo de Huff

$$P_{ij} = U_{ij} / \sum_{k \in N_i} U_{ik}$$

Fonte: Huff e McCallum (2008)

Onde P_{ij} é a probabilidade de um indivíduo i selecionar a alternativa j dada a utilidade de j relativa à soma das utilidades das n alternativas consideradas pelo indivíduo i . A utilidade deve ser definida empiricamente e pode ser definida como o conjunto de variáveis que podem impactar em uma determinada escolha. Em sua primeira aplicação do modelo, Huff definiu a utilidade de uma loja pelo seu tamanho em metros quadrados (diretamente proporcional) e a distância entre a loja e a residência dos consumidores (inversamente proporcional).

Com a hipótese de que a distância e o tamanho da loja eram mais importantes para alguns produtos do que para outros, Huff (2013) explica que as duas variáveis devem ser ponderadas por um parâmetro. Por exemplo, é esperado que a maior distância de uma loja apresente um efeito mais negativo para o consumidor no caso de bens de conveniência (ex.: refrigerante), do que para bens de comparação (ex.: carro). Para a aplicação deste caso, com as ponderações, a probabilidade de um consumidor localizado em i selecionar a loja j pode ser estimada como:

Figura 3 - Equação Modelo de Huff aplicado em Huff 1962

$$P_{ij} = S_j^\alpha D_{ij}^\beta / \sum_{k \in N_i} S_k^\alpha D_{ik}^\beta$$

Fonte: Huff e McCallum (2008)

Onde S_j é a metragem quadrada da área de vendas de uma loja, D_{ij} é a distância de i para j , e α e β são os parâmetros S_j e D_{ij} .

Contudo, Aranha e Figoli (2001) mostram que a calibragem do modelo tem sido um fator negativo devido à atribuição muitas vezes arbitrária dos parâmetros. Huff e McCallum (2008) apontam que essa solução por aproximação tornava impossível a validação da significância estatística desses parâmetros, o que pode acarretar em resultados errados. Grande parte do problema se deve à falta de acesso dados de origem, ou seja, das preferências dos consumidores. Além disso, o modelo infere que se sabe a utilidade de todas as alternativas possíveis, porém geralmente as empresas não possuem os dados da concorrência (ARANHA; FIGOLI, 2001; HUFF; MCCALLUM, 2008).

Desde o surgimento do modelo de Huff, avanços foram feitos na estimativa de parâmetros, que agora permitem a associação estatística dos mesmos. Diversas variáveis podem ser incluídas no modelo para que se determine medidas de

acessibilidade e atratividade. De acordo com Huff (2013), elas podem ser divididas em controláveis (ex.: preço, formato da loja) e não controláveis (ex.: renda, distribuição da população). De maneira geral, o modelo pode ser expresso da seguinte maneira:

Figura 4 - Equação 2 Modelo de Huff

$$P_{ij} = \left(\prod_{h=1}^H A_{hj}^{\gamma_h} \right) D_{ij}^{\lambda} / \sum_{j=1}^n \left(\prod_{h=1}^H A_{hj}^{\gamma_h} \right) D_{ij}^{\lambda}$$

Fonte: Huff e McCallum (2008)

Onde P_{ij} é a probabilidade de um consumidor na área geográfica i , escolher o estabelecimento j ; A_{hj} é a medida da característica h ($h=1,2,..H$) que reflete a atração do estabelecimento h ; γ é um parâmetro para a sensibilidade de P_{ij} associada com a variável de atração h ; D_{ij} é a medida de acessibilidade do estabelecimento j para um consumidor localizado em i ; λ é um parâmetro para a sensibilidade de P_{ij} com respeito à acessibilidade; e n é o número de estabelecimentos.

Com isso, pode ser determinado o gasto esperado em cada região determinada pelo estudo:

Figura 5 - Vendas do Estabelecimento j

$$E_{ij} = (P_{ij})(B_i)$$

Fonte: Huff e McCallum (2008)

Onde E_{ij} é o gasto que será realizado na região i para o estabelecimento j ; e B_i é o gasto total possível na região i .

5. PROCEDIMENTOS METODOLÓGICOS

Esta seção pretende dar enfoque nos procedimentos metodológicos utilizados para a criação do modelo de atratividade para venda de interruptores na cidade de São Paulo, além de descrever os critérios que foram utilizados. As fontes de dados para a realização deste estudo podem ser divididas em três pontos principais:

- (1) A base de dados sobre pontos de vendas na área de materiais de construção é realizada pela Associação Nacional dos Comerciantes de Material de Construção (ANAMACO) e foi obtida por meio da cátedra Tramontina Eletrik. Trata-se de uma pesquisa quantitativa de abrangência nacional, realizada por meio de entrevistas com varejistas através do CATI (*Computer Assisted Telephone Interviews*). Os dados continham informações referentes a 13.089 pontos de venda com atuação na região metropolitana de São Paulo e seu exercício no ano de 2015 (<http://pesquisas.anamaco.com.br/>).
- (2) Dados demográficos obtidos pelo site da prefeitura de São Paulo referente ao ano de 2012 e do Censo 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE);
- (3) E dados da Inteligência Empresarial da Construção (ITC) (<http://www.itc.etc.br/>) referentes a lançamentos residenciais e comerciais na cidade de São Paulo. A pesquisa forneceu a identificação (nome e endereço) de todos os empreendimentos realizados no ano de 2016 na cidade e também foi obtida através da cátedra Tramontina Eletrik.

No Quadro 1, é possível verificar um resumo de todas as variáveis disponíveis para este trabalho, sua descrição, escala e a fonte de onde foi obtida.

Quadro 1 – Variáveis das bases de dados (continua)

Variável	Escala	Descrição	Fonte
UF	Nominal	Unidade Federativa em que o Comércio estava localizado	Censo (2010)
Código da UF	Ordinal	Código IBGE para a Unidade Federativa	Censo (2010)
Município	Nominal	Município em que o Comércio estava localizado	Censo (2010)
Código do Município	Ordinal	Código do IBGE para o município	Censo (2010)
Distrito	Nominal	Distrito em que o Comércio estava localizado	Censo (2010)
Código do Distrito	Ordinal	Código do IBGE para o distrito	Censo (2010)

Código do Setor Censitário	Ordinal	Código do IBGE para o setor censitário em que o Comércio estava localizado	Censo (2010)
Número de Moradores	Razão	Número de moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes no setor censitário	Censo (2010)
Número de Domicílios	Razão	Número de domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes no setor censitário	Censo (2010)
Rendimento Mensal Médio	Razão	Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) no setor censitário	Censo (2010)
Endereço do Comércio	Nominal	Logradouro e número do comércio	ANAMACO (2015)
CEP do Comércio	Nominal	Código de Endereçamento Postal do comércio	ANAMACO (2015)
CNPJ do Comércio	Nominal	Cadastro Nacional de Pessoa Jurídica do comércio	ANAMACO (2015)
Razão Social do Comércio	Nominal	Razão Social do comércio	ANAMACO (2015)
Código de Atividade Econômica (CNAE)	Nominal	Tipo de atividade primária do comércio (i.e., atacado, varejo de materiais elétricos, varejo de materiais hidráulicos)	ANAMACO (2015)
Porte do Comércio	Nominal	Porte do comércio dividido em pequeno, médio e grande, baseado no número de funcionários do estabelecimento	ANAMACO (2015)
Faixa de Faturamento do Comercio	Ordinal	Intervalo de faturamento anual do comércio, dividido em seis faixas	ANAMACO (2015)
Número de Funcionários	Razão	Número de funcionários na unidade de comércio	ANAMACO (2015)
Número de Funcionários Total	Razão	Número Total de funcionários do comércio em todas as unidades (no caso de uma rede)	ANAMACO (2015)
Índice de Desenvolvimento Humano	Razão	Índice de Desenvolvimento Humano por distrito, variando de 0 a 1.	Prefeitura de SP (2012)
Nome dos Lançamentos	Nominal	Identificação dos lançamentos residenciais e comerciais na cidade de São Paulo	Inteligência Empresarial da Construção (ITC) (2016)
Endereço dos Lançamentos	Nominal	Logradouro dos lançamentos residenciais e comerciais na cidade de São Paulo	Inteligência Empresarial da Construção (ITC) (2016)

Fonte: Criação própria

A partir destes dados secundários, foi construída a base de dados final que será utilizada neste trabalho. Os tópicos a seguir têm o objetivo de definir de maneira mais detalhada estes dados e o processo de coleta e análise percorridos. Serão divididos em:

coleta e seleção de dados, construção do modelo inicial de atratividade, preparação e exame inicial da base de dados e plano de análise dos dados

5.1 COLETA E SELEÇÃO DE DADOS

A base de dados dos pontos de venda comerciais da ANAMACO, foi o alicerce para a elaboração do modelo de atratividade. Cada observação dos 13.089 pontos de vendas continha, entre outras variáveis, a Classificação Nacional de Atividade Econômica (CNAE), o que permitiu a identificação da atividade primária de cada comércio, que variava entre varejos e atacados de materiais de construção, materiais elétricos, ferramentas, entre outros. A CNAE “é um instrumento de padronização nacional de códigos de atividade econômica e dos critérios de enquadramento utilizados pelos diversos órgãos da Administração Tributária do país” (RECEITA FEDERAL, 2014). Coordenada pela Receita Federal e orientação técnica do IBGE, a classificação tem uma estrutura de cinco dígitos, acrescidos de dois dígitos em casos de detalhamento de subclasses. No Quadro 2 abaixo, pode-se verificar a lista de códigos da CNAE encontrados na base de dados bem como a porcentagem de lojas da base encontradas para cada um:

Quadro 2 – Códigos CNAE da Base de Dados

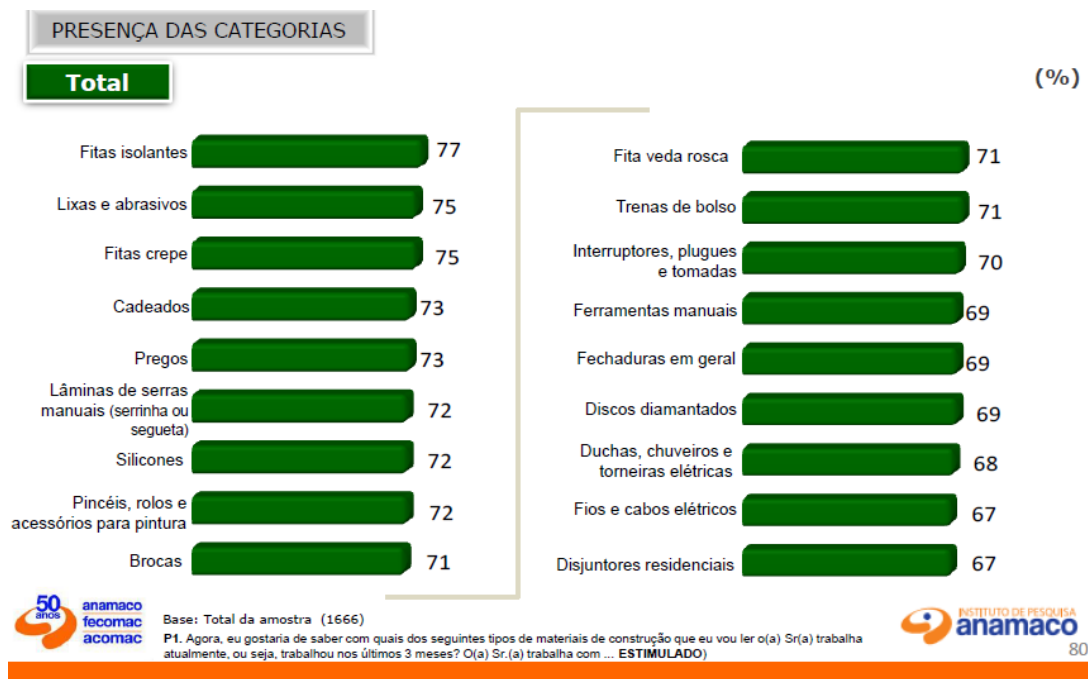
Código CNAE	Atividade Econômica	% de Lojas da Base
46.71-1	Comércio atacadista de madeira e produtos derivados	0,5%
46.72-9	Comércio atacadista de ferragens e ferramentas	2,1%
46.73-7	Comércio atacadista de material elétrico	1,1%
46.74-5	Comércio atacadista de cimento	0,0%
46.79-6/01	Comércio atacadista de tintas, vernizes e similares	0,3%
46.79-6/02	Comércio atacadista de mármore e granitos	0,5%
46.79-6/04	Comércio atacadista especializado de materiais de construção não especificados anteriormente	0,5%
46.79-6/99	Comércio atacadista de materiais de construção em geral	1,2%
47.41-5	Comércio varejista de tintas e materiais para pintura	5,4%
47.42-3	Comércio varejista de material elétrico	5,0%

47.44-0/01	Comércio varejista de ferragens e ferramentas	13,2%
47.44-0/02	Comércio varejista de madeira e artefatos	3,7%
47.44-0/03	Comércio varejista de materiais hidráulicos	1,8%
47.44-0/04	Comércio varejista de cal, areia, pedra britada, tijolos e telhas	0,5%
47.44-0/05	Comércio varejista de materiais de construção não especificados anteriormente	15,8%
47.44-0/99	Comércio varejista de materiais de construção em geral	48,3%

Fonte: cnae.ibge.gov

A partir disso, foram selecionados apenas os pontos de venda da cidade de São Paulo, região alvo do estudo, e realizou-se uma análise das CNAEs de modo que fossem selecionadas as atividades econômicas que tivessem relação com a venda de interruptores elétricos. De acordo com o Fórum Pesquisa ANAMACO, estudo aberto da associação referente ao exercício de 2014, produtos como interruptores, plugues, tomadas, fios e cabos elétricos estão presentes em cerca de 70% dos comércios varejistas, conforme o Gráfico 1 (<http://novo.anamaco.com.br/dadosSetor-interna.aspx?uid=21>). O dado foi obtido em pesquisa a nível nacional com mais de 1.600 lojas varejistas entrevistadas, todas de portes variados e com especializações diversas (generalista, elétrico, hidráulico, etc.).

Gráfico 1 - % de Presença das Categorias em Lojas de Materiais de Construção em Geral



Fonte: ANAMACO (2016)

Também, é importante ressaltar que é comum os comércios possuírem outros CNAEs como atividades secundárias, fato que foi verificado através da consulta por CNPJ de diversos comércios da base de dados no site da Receita Federal. Visto isso, foi adotada uma estratégia mais conservadora e optou-se por excluir somente os comércios varejistas e atacadistas de tintas, materiais de pintura, vernizes e similares (CNAEs 4741-5 e 4679-6/01), por serem pontos de venda especializados neste segmento, conforme informado pela própria ANAMACO. Com a exclusão de lojas de outras cidades e das lojas especializadas citadas acima, o número de pontos de venda da base de dados final foi de 8.616 mil.

Como o intuito aqui é construir um modelo de atratividade para determinar as melhores regiões para venda de interruptores elétricos, os 96 distritos de São Paulo capital, definidos no site da prefeitura municipal (<http://www.prefeitura.sp.gov.br>), foram escolhidos como as regiões de análise, devido a possibilidade de uma maior riqueza de dados disponíveis. Um distrito pode ser definido como as unidades administrativas dos municípios. Sua criação, desmembramento ou fusão dependem de

leis municipais previstos em lei complementar estadual (IBGE, 2010). Se trata da primeira unidade territorial reduzida após a unidade município.

Assim, foi necessário o processo de agrupamento destes pontos de venda por distrito de São Paulo. Para se identificar em qual distrito cada loja se localizava, utilizou-se da ferramenta Google Maps e do site BatchGeo (<https://pt.batchgeo.com/>) de modo a se obter as coordenadas de cada ponto. As duas ferramentas fornecem as coordenadas no formato grau decimal, através do endereço solicitado, tornando possível localizar os comércios no mapa. No entanto, para o procedimento no site BatchGeo, é gerado um arquivo no formato KML, utilizado para exibir dados geográficos, que só pôde ser analisado com o software livre denominado Quantum GIS 2.4 (QGIS), um sistema de informação geográfica. Na Tabela 1, pode ser visualizado um exemplo das coordenadas geradas para dez lojas:

Tabela 1 - Coordenadas de 10 Lojas da Base de Dados

Latitude	Longitude	Razão Social da Loja
-23.7076484	-46.6120508	ZIF MATERIAL PARA CONSTRUCAO LTDA - ME
-23.5362355	-46.650487	BELFIX COMERCIAL LTDA EPP
-23.6465338	-46.7118751	NEO IMPORTACAO E EXPORTACAO DE REVESTIMENTOS LTDA
-23.4988346	-46.6573998	OPA CRIANDO ARTE COM DE PRODUTOS PARA ARTE SANATO LT
-23.4014526	-46.3218983	TRANSYOSHIO COMERCIO E TRANSPORTES LTDA ME
-23.6663902	-46.5404671	JOSE LUIS CERQUEIRA DOS ANJOS EPP
-23.4552186	-46.6725437	SERVILO E MILER SERRALHERIA LTDA ME
-23.6785846	-46.6092392	B.R. BESSON MATS PARA CONSTRUCAO LTDA
-23.6274008	-46.5388161	APML DO BRASIL COM IMP E EXP EIRELI ME
-23.5572316	-46.5244434	DEUSDETE RODRIGUES VIEIRA ME
-23.4520852	-46.7505645	JOADI COSTA SANTOS PORTOES ME
-23.4502562	-46.519059	JO SANI DEL GRANDE ME
-23.550114	-46.6077085	SCALLA GESSO COMERCIO IMP EXP LTDA

Fonte: Criação própria

As coordenadas de cada loja foram exportadas para um arquivo CSV (*Comma-separated value*) e, devido ao grande volume de dados, o software livre RStudio, voltado para gráficos e computação estatísticos, foi utilizado para automatizar o agrupamento das lojas por distrito. O procedimento foi utilizar os micro dados da malha

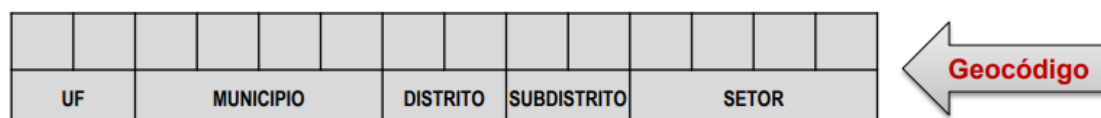
territorial do Censo 2010, disponibilizados no site do IBGE (ftp://geofpt.ibge.gov.br/recortes_para_fins_estatisticos/malha_de_setores_censitarios/censo_2010/base_de_faces_de_logradouros/) e cruzá-los com as coordenadas de cada comércio. Através do pacote *rgdal* do RStudio, utilizado para a elaboração de mapas, foi possível converter as coordenadas em polígonos para que fosse identificado dentro da malha territorial do Censo 2010, o setor censitário e o distrito ao qual cada loja pertencia.

A malha territorial definida pelo Censo do IBGE é dividida entre: Unidades da Federação, Municípios, Distritos, Subdistritos, Bairros e Setores Censitários. O setor censitário é o menor nível de agregação da pesquisa, com limites de 250 a 350 domicílios para áreas urbanas e de 100 a 200 para áreas rurais. O IBGE o define como:

A menor unidade territorial, com limites físicos identificáveis em campo, com dimensão adequada à operação de pesquisas e cujo conjunto esgota a totalidade do território nacional, o que permite assegurar a completa cobertura do país” (IBGE, 2011)

Nos documentos referentes as malhas territoriais do Brasil, disponibilizados pelo IBGE, cada malha possui um código específico, denominado geocódigo e uma hierarquia entre si. Estes códigos foram utilizados para uma melhor identificação e uso dos micro dados do Censo 2010, que são disponibilizados no formato *shapefile*, formato de arquivo voltado para mapas. Como se pode ver na Figura 6, a estrutura do geocódigo acompanha a especificidade da malha territorial, sendo o setor censitário a unidade territorial com o maior número de dígitos (14). No caso de São Paulo, a estrutura do geocódigo consta com um dígito adicional - para a Unidade Federativa de São Paulo o código referente é “35”, seguido pelo município São Paulo, de código “50308”. Posteriormente os distritos, com dígitos de “01” a “96” e assim por diante. Estes geocódigos, foram utilizados neste trabalho para agrupar as lojas por distrito.

Figura 6 - Estrutura do Geocódigo do IBGE



Fonte: IBGE (2011)

A partir do cruzamento entre a base das lojas e do Censo 2010, também foi possível agregar variáveis demográficas a nível setor censitário do Censo 2010. As variáveis de rendimento médio dos domicílios, número de moradores, domicílios e agora, o número de pontos de venda, foram então agrupadas ao nível distrito. Os mesmos procedimentos descritos acima foram utilizados para definir o número de lançamentos comerciais e residenciais de cada distrito e, assim, agregar estes dados para a base de dados final. Os códigos executados no RStudio estão disponibilizados nos apêndices do trabalho, para uma melhor análise.

5.2 CONSTRUÇÃO DO MODELO INICIAL DE ATRATIVIDADE

A elaboração do modelo inicial de atratividade tem sua base no modelo de Huff construído por Dolega, Pavlis e Singleton (2016), que cria um índice de atratividade entre regiões, supondo uma relação linear entre variáveis quantitativas tidas como importantes na escolha de destino de compra do consumidor. No entanto, é importante ressaltar que o modelo aqui se restringiu apenas à variável de atratividade determinada pelos autores, sem contemplar o modelo de Huff completo, que também considera a variável distância entre o consumidor e as lojas para determinar as zonas de atratividade do varejo. O motivo desta restrição ocorre visto que, para o modelo de atratividade composto neste trabalho, a atratividade não é pensada do ponto de vista de um novo varejista que almeja entrar no mercado, mas sim do ponto de vista dos fabricantes (e.g., uma região com maior número de pontos de venda tende a ser mais atrativa para um fabricante, porém menos atrativa para um novo varejista). Dessa forma, o modelo de Dolega foi utilizado somente como base para escolha de algumas variáveis, pois o método de análise utilizado posteriormente foi o de regressão linear.

O objetivo geral no presente trabalho foi de entender quais as regiões de maior atratividade para a venda de interruptores elétricos na cidade de São Paulo, através da criação de um índice de atratividade, e identificar as variáveis de maior impacto neste índice. Muitos estudos voltados para a elaboração de modelos de atratividade comparam seus resultados com dados de pesquisa *survey* realizadas com o público-alvo (e.g. consumidores ou empresas). Assim, é possível testar a capacidade preditiva do modelo

(DOLEGA; PAVLIS; SINGLETON, 2016; DENNIS, 2002; MERRINO; NARRAFATE, 2015).

A partir da base de dados da ANAMACO, utilizou-se da informação das faixas de faturamento anual de cada ponto de venda de materiais de construção como proxy para a sua atratividade. Ressalta-se que o fato desta variável ser ordinal e dividida em seis intervalos impossibilitou uma maior especificidade e diferenciação entre as lojas. Como estes intervalos apresentavam variação de maneira disforme (e.g., o sexto intervalo era mais de seis vezes superior ao primeiro intervalo), o método proposto então, foi de identificar o ponto médio de cada faixa de faturamento e atribuir este valor médio para cada loja:

- Para os pontos de venda com faturamento de até 500 mil reais anuais, foi definido o faturamento de 250 mil reais/ano;
- Para os pontos de venda com faturamento de 500 até 750 mil reais anuais, foi definido o faturamento de 625 mil reais/ano;
- Para os pontos de venda com faturamento de até 750 até 2,4 milhões de reais anuais, foi definido o faturamento de 1.575 milhão reais/ano;
- Para os pontos de venda com faturamento de até 2,4 até 10 milhões de reais anuais, foi definido o faturamento de 6,2 milhões de reais/ano;
- Para os pontos de venda com faturamento de até 10 a 23 milhões de reais anuais, foi definido o faturamento de 16,5 milhões de reais/ano;
- Para os pontos de venda com faturamento acima de 23 milhões de reais anuais, foi definido o faturamento de 23 milhões reais/ano, visto não haver um limite estabelecido;

Dessa maneira, os faturamentos anuais estabelecidos para cada loja foram somados em cada um dos 96 distritos da cidade de São Paulo, de modo a se obter o “faturamento total” dos distritos, definido aqui como sua atratividade.

Com o intuito de estimar a atratividade para vendas de interruptores, o pressuposto aqui é de que existe uma relação linear entre o faturamento total de cada distrito e as vendas de interruptores, ou seja, quanto maiores forem as vendas de materiais de construção em geral na região, maiores serão as vendas para interruptores. Para isso, foram selecionados os comércios mais propensos a venda de interruptores, procedimento explicado na seção anterior. Ainda assim, tanto esta hipótese quanto o

método de ponto médio para estabelecer o faturamento anual de cada loja entrarão como uma limitação do presente estudo, que será melhor descrita na seção final do trabalho.

Definido o faturamento total de cada distrito (variável dependente), as variáveis explicativas (independentes) do modelo podem ser divididas em dois grupos: referentes aos pontos de vendas e referentes à população do distrito. Pode-se fazer um paralelo com os fatores de oferta e demanda, respectivamente, dois importantes componentes no conceito de atratividade mencionados por Birkin (2010). O objetivo de agregar o segundo grupo de variáveis foi identificar o impacto das características relacionadas à população do distrito na atratividade do mesmo. Assim, as variáveis utilizadas para teste do modelo inicial foram: o número de pontos de venda de materiais de construção por distrito (Smc), a presença de grandes redes (G), o número de funcionários total no distrito (F), o número de domicílios (D), os lançamentos residenciais e comerciais (Lrc) e a renda média dos domicílios (Rmd).

O número de pontos de venda de materiais de construção no distrito (Smc) foi considerado na hipótese de que uma maior oferta de pontos de venda resulta em maiores opções, aumentando a probabilidade de deslocamento ao distrito. Em estudos voltados para a atratividade a nível loja, o clássico método utilizado é da área de vendas do ponto, no entanto, para centros comerciais, Dolega, Pavlis e Singleton (2016) também utilizam a soma dos pontos de venda em seu modelo.

A presença de grandes redes (G) se baseia na ideia de que uma rede maior possui uma maior visibilidade e uma marca mais forte, inclusive podendo desempenhar o papel de loja âncora na região. Portanto, um maior reconhecimento do consumidor, pode resultar em vendas maiores em relação aos outros pontos de venda concorrentes. A hipótese se faz a luz de achados de Finn e Louviere (1996), que através de um estudo investigativo por meio de *surveys* identificaram diversas variáveis de imagem que impactam no *share of choice* do consumidor em relação a shopping centers. Aqui, foram consideradas grandes redes, todas as lojas da base que apresentaram o porte “Grande”, que também se enquadravam no faturamento máximo de 24 milhões de reais/ano. Desse modo, foi aplicada uma variável binária, atribuindo um valor de 0 para distritos que não apresentavam a presença de nenhuma grande rede e de 1 para distritos que apresentassem ao menos uma grande rede.

O número de funcionários (F) foi considerado a partir da hipótese de que lojas com um maior quadro de funcionários possuem um custo maior e que este deve ser proporcional ao seu faturamento. Lojas maiores, possuem então, um maior número de funcionários para atender um maior número de clientes e, um distrito que tiver um número total de funcionários maior, tende a possuir um maior faturamento total.

Por fim, o número de domicílios da região (D), os lançamentos residenciais e comerciais (Lrc), e o rendimento médio em reais (Rmd) são possíveis fatores que podem impactar na demanda de interruptores de uma região e, portanto, podem ter relação com as vendas totais de interruptores no distrito.

O modelo inicial que será testado para determinar a atratividade do distrito pode ser visto na equação abaixo:

$$Ad = Smc + G + F + D + Lrc + Rmd \quad (3)$$

No Quadro 3, pode-se ver um resumo do modelo proposto inicialmente, com todas as variáveis relacionadas ao nível distrito:

Quadro 3 - Variáveis do Modelo Proposto

Variável	Escala	Descrição	Fonte
Atratividade do Distrito (Ad)	Razão	Atratividade do Distrito, medida pelo faturamento anual agregado das lojas do distrito.	ANAMACO (2015)
Número de Lojas (Smd)	Razão	Soma do número de Lojas de materiais de construção no distrito.	ANAMACO (2015)
Presença de Grandes Redes (G)	Nominal	Variável binária, atribuindo 0 para distritos sem a presença de grandes redes e 1 para distritos com ao menos uma grande rede.	ANAMACO (2015)
Funcionários (F)	Razão	Soma do número de funcionários das lojas do distrito.	ANAMACO (2015)
Domicílios (D)	Razão	Número de domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes no distrito.	Censo (2010)
Lançamentos Residenciais e Comerciais (Lrd)	Razão	Número total de lançamentos residenciais e comerciais no distrito.	Inteligência Empresarial da Construção (ITC) (2016)
Rendimento Médio Mensal	Razão	Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) no distrito.	Censo (2010)

Fonte: Criação própria

5.3 PREPARAÇÃO E EXAME INICIAL DA BASE DE DADOS

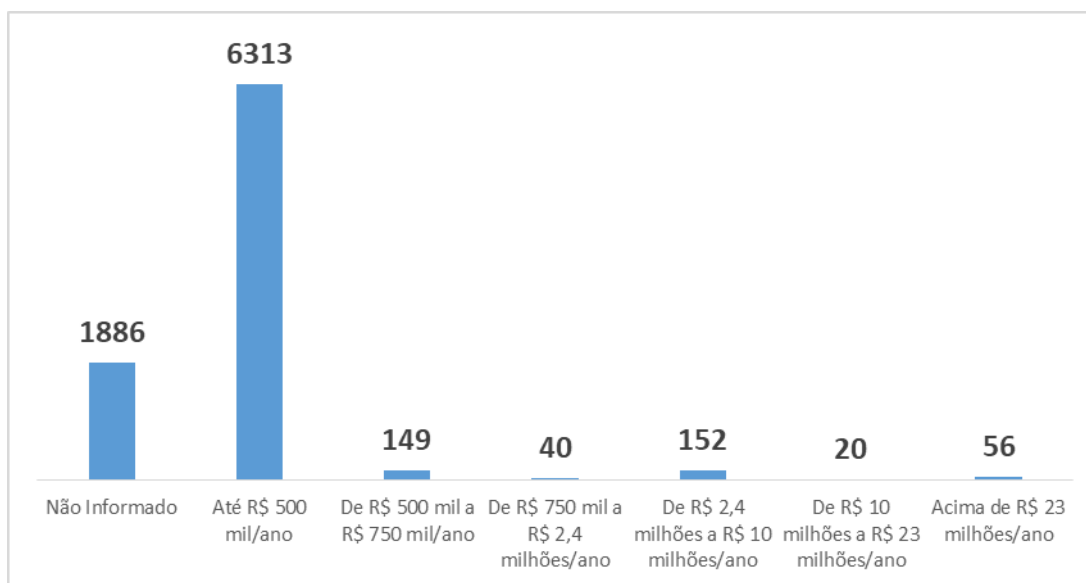
A análise da base de dados final se utilizará de métodos quantitativos para estabelecer seus resultados. A estratégia analítica foi de primeiramente obter um melhor entendimento das variáveis através de um exame inicial, para que na segunda etapa fosse realizada uma análise inferencial do modelo e suas variáveis, através do método de regressão linear. Inicialmente, então, realizou-se uma análise descritiva das variáveis e a relação entre as mesmas, de modo a se identificar suas características iniciais. De acordo com Hair (2010), o objetivo desta etapa “é tanto revelar o que não está aparente como retratar os dados atuais, porque efeitos ocultos são facilmente negligenciados”.

Assim, foram realizadas, sobretudo, técnicas de análise gráfica devido a facilidade de leitura das informações para uma grande quantidade de dados. Os efeitos procurados neste primeiro momento foram (1) a forma distribuição dos dados, (2) a identificação de *missings* e *outliers* e (3) efeitos de multicolinearidade no modelo.

A forma de distribuição dos dados permite que seja visualizado o comportamento de uma variável, sua grandeza, extremos e outras características. Além disso, é importante para verificar se existe a característica de normalidade na variável dependente para que a técnica de regressão linear seja realizada posteriormente. Conforme Hair (2010), “se a variação da distribuição normal é suficientemente grande, todos os testes estatísticos são válidos, já que a normalidade é necessária para usar o F e teste t”.

Antes de analisar então, a distribuição de faturamento total dos distritos de São Paulo, a variável dependente do modelo, foi importante entender o número de comércios dentro de cada intervalo de faturamento. Verifica-se no Gráfico 2 que existe uma grande predominância de pontos de venda menores com faturamento de até 500 mil reais/ano (definidos com ponto médio de 250 mil reais/ano), que representaram cerca de 73% dos pontos de venda estudados e são a maioria na cidade de São Paulo. Deve-se ressaltar também dos 8.616 pontos de venda, 1.906 não apresentaram dados de faturamento, o que será analisado posteriormente ao verificar os *missings* (dados faltantes) e seu impacto no resultado deste trabalho.

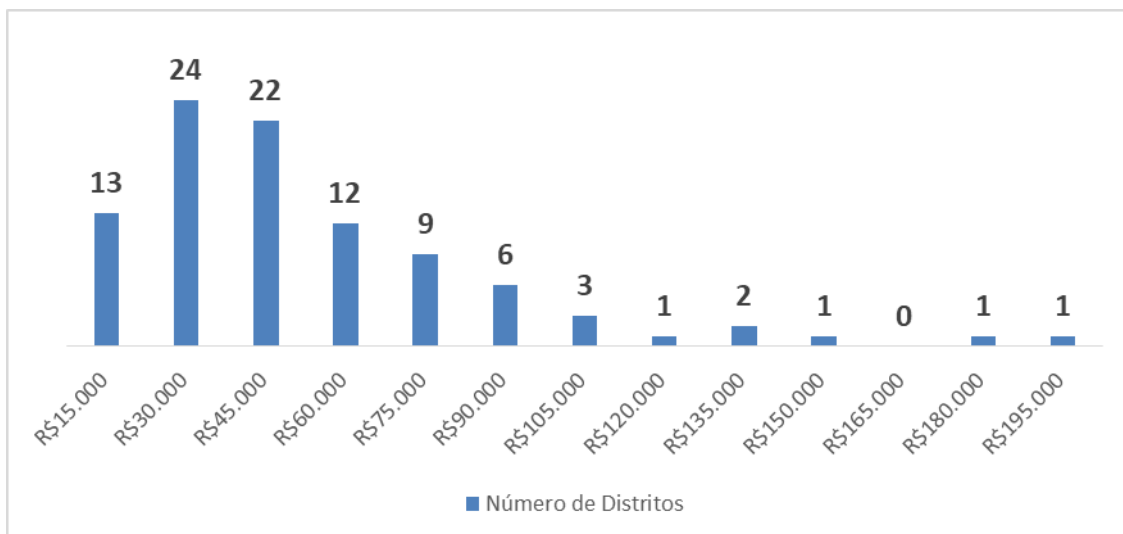
Gráfico 2 – Número de Lojas por Faixa de Faturamento Anual



Fonte: Criação Própria

Com este entendimento, pode-se retornar para a variável dependente do modelo. Considerando apenas as lojas que apresentaram a informação do intervalo de faturamento, o total de lojas analisadas foi de 6.790. No gráfico abaixo, verifica-se que a distribuição do faturamento total dos distritos se situou entre os limites de 4.625 milhões de reais/ano e 191 milhões de reais/ano, e aparenta ter uma assimetria positiva (à direita). Isso se deve ao fato mencionado acima, de um número muito alto de comércios com uma faixa de faturamento menor e um número restrito de comércios com um faturamento muito superior, que podem ser considerados *outliers*. É importante ressaltar que este formato de distribuição pode prejudicar os resultados da análise final, visto que uma distribuição normal é o requisito que garante a confiança dos testes.

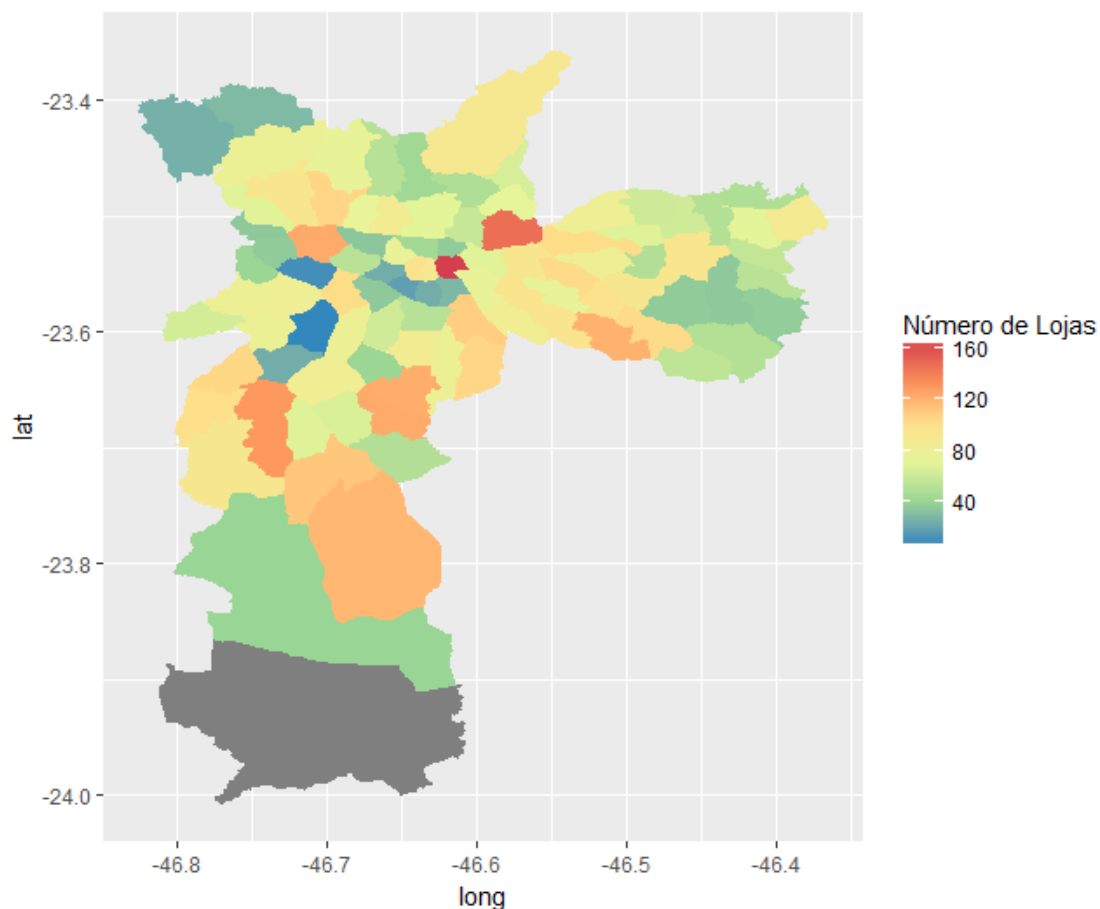
Gráfico 3 – Distribuição do Faturamento Total dos Distritos ('000)



Fonte: ANAMACO (2015)

Entre os 96 distritos de São Paulo, somente o distrito de Marsilac, distrito rural na região sul, não apresentou nenhum ponto de venda. Todas as análises a seguir foram realizadas considerando, então, os 95 distritos que apresentaram algum comércio. Na figura 7, é possível analisar um mapa de densidade dos distritos em relação ao número de lojas que cada um apresentou. Marsilac aparece em cinza na região sul do mapa. Em vermelho, pode-se encontrar o distrito de Brás, que apresentou a maior concentração das lojas de materiais de construção de São Paulo, com 167 pontos. Por outro lado, distritos em azul, como Morumbi e Alto de Pinheiros apareceram com o menor número de lojas, 7 e 10 respectivamente. Cabe ressaltar que também estão entre os distritos com maior poder aquisitivo, como será analisado abaixo.

Figura 7 – Número de Lojas por Distritos da Cidade de São Paulo



Fonte: ANAMACO (2015)

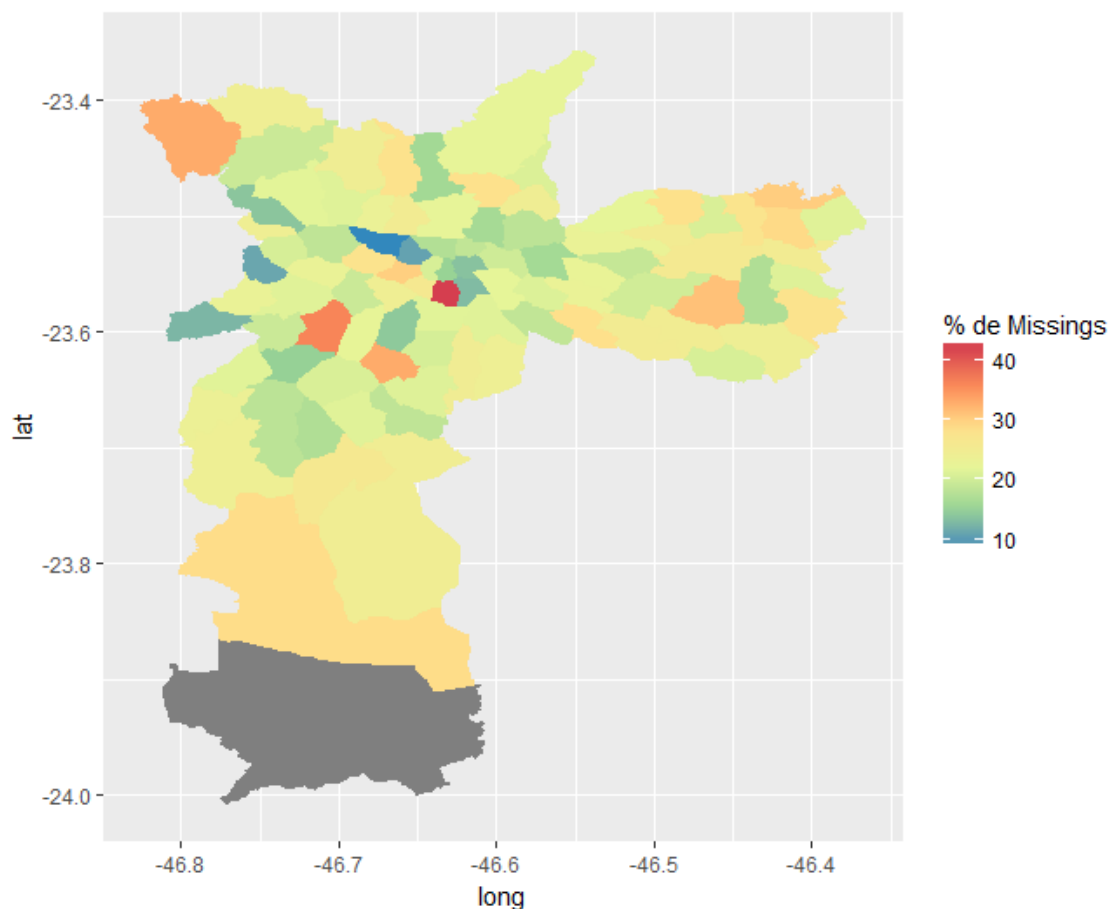
Para as variáveis independentes do modelo foram analisadas as estatísticas descritivas de médias, desvios padrão e limites, para verificar as características de cada uma. Pode-se perceber que os limites máximos em variáveis como o faturamento total e o rendimento mensal médio, por exemplo, ultrapassam até três desvios padrão, estando bem distantes da média da cidade. No primeiro caso, o distrito de Brás se destaca com um faturamento total de 191,4 milhões de reais/ano, muito acima da média entre os distritos de São Paulo. Já para o rendimento médio mensal, distritos com população de maior poder aquisitivo como Morumbi e Moema, ficam muito fora da média da cidade, apresentando um rendimento de 10.503,90 mil reais/mês e 8.115,67 mil reais/mês, respectivamente. Na tabela 2 pode-se ver as características gerais das variáveis selecionadas para análise.

Tabela 2 – Características Descritivas das Variáveis

VARIÁVEL	MÍN	MÁX	MÉDIA	DESVIO PADRÃO
Faturamento Total ('000)	4.625	191.400	45.280	35.310
Nº de Pontos de Venda	7	167	71	32
Nº Total de Funcionários	142	2.920	767	500
Rendimento Mensal Médio	721,2	10.503,90	2.441,90	1952,70
Nº de Domicílios	5.543	103.070	37.599	19.825
Lançamentos Residenciais e Comerciais	0	60	5	10

Fonte: Censo (2010), ANAMACO (2015), ITC (2016)

Como visto anteriormente, a base de dados contou com 1.886 casos de pontos de venda sem informação declarada de faturamento. Devido a isso, foi analisada a distribuição destes *missings* entre os distritos e foi possível verificar que os casos estão bem divididos entre os distritos, com nenhum apresentando mais de 2% do total de casos sem informação. Após esta primeira análise, o segundo passo foi verificar se havia algum distrito com uma grande proporção de pontos de venda sem informação em relação aos seus pontos de vendas totais. Dentre os distritos, o mais afetado foi o distrito Liberdade, que pode ser visualizado em vermelho na figura 8 com 43% dos pontos de venda sem a informação de faturamento (17 dos 40 pontos de venda), mas de maneira geral, pode-se verificar no gráfico abaixo que grande parte dos distritos apresentou entre 20% e 30%. Deve-se ressaltar, o potencial impacto negativo da falta de informação destes pontos, pois além disso poderiam retratar uma situação de “dados faltantes não aleatórios” conforme Hair (2010), na hipótese, por exemplo, de que a maioria destes comércios poderiam pertencer a intervalos de faturamentos maiores, por exemplo.

Figura 8 – % de Lojas sem Informação de Faturamento – Total por Distrito

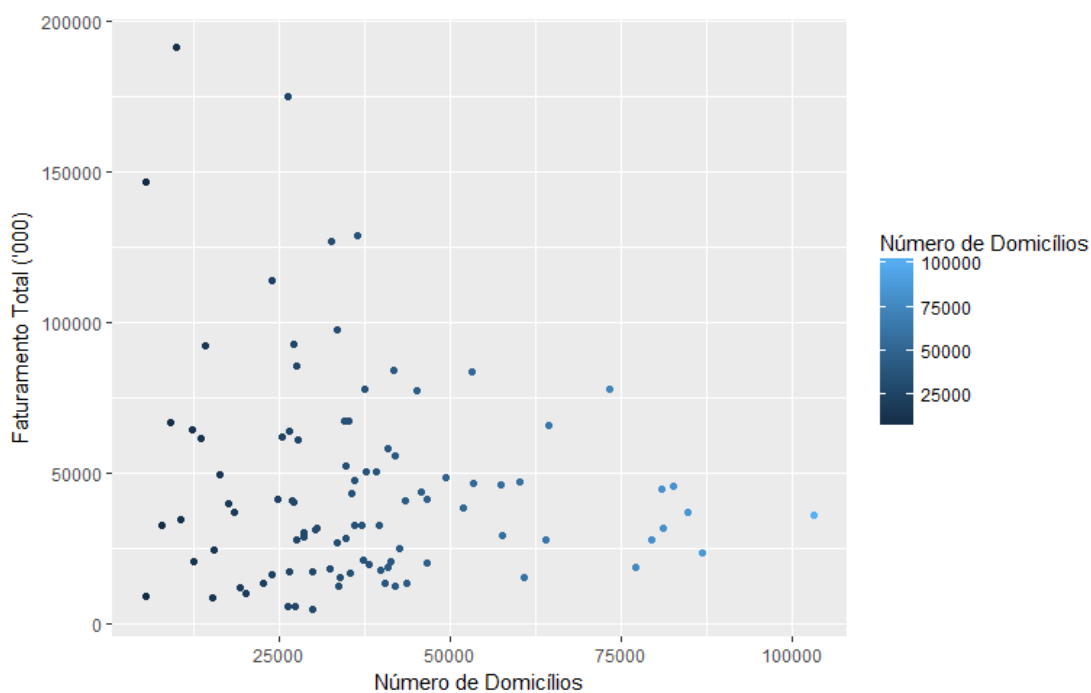
Fonte: ANAMACO (2015)

A linearidade entre as variáveis foi testada através do gráfico *scatterplot* de dois eixos. Para não realizar um número extenso de gráficos, optou-se por ilustrar somente o cruzamento da variável dependente com as variáveis independentes. O *scatterplot* é importante pois permite verificar se existe relação linear entre variável dependente e as variáveis explicativas, além de detectar a presença de *outliers*. Os *outliers* podem ser definidos como “uma observação substancialmente diferente das outras (i.e., tem valor extremo) em uma ou mais características” (HAIR, 2010). Estas observações podem distorcer os resultados de uma análise ou, pelo contrário, representar uma parcela da população em análise que não está representada na amostra. É importante ressaltar que os dados dos pontos de venda estudados serão vistos como o universo de São Paulo, sem o objetivo de expansão dos mesmos. Para o modelo inicial testado, não foi retirado nenhum *outlier*.

No gráfico 4, pode-se verificar a dispersão dos 95 distritos na relação do seu faturamento total com o número de domicílios. Apesar dos pontos apresentarem uma dispersão forte, pode-se verificar uma relação aparentemente negativa entre as duas variáveis, ou seja, a medida que o número de domicílios é maior, parece reduzir o faturamento total nos distritos. Hipoteticamente, a relação pode fazer sentido visto que regiões muito residenciais, podem apresentar um número menor de comércios de materiais de construção; além disso, comércios maiores, que tendem a ter um faturamento proporcionalmente grande, podem buscar uma alocação em regiões mais afastadas, devido ao custo menor do terreno. O ponto no canto esquerdo acima no gráfico é referente ao distrito Brás, que apresentou o faturamento anual mais alto entre os distritos e também um número de domicílios abaixo da média (10.110 domicílios).

Na outra extremidade, pode-se verificar o distrito de Grajaú, um dos maiores distritos de São Paulo, com mais de 10 mil domicílios e um dos menores faturamentos para o setor de materiais de construção. No entanto, é possível perceber que boa parte dos casos se situa no quadrante inferior esquerdo, onde se situa a média da cidade de São Paulo, de 37 mil domicílios.

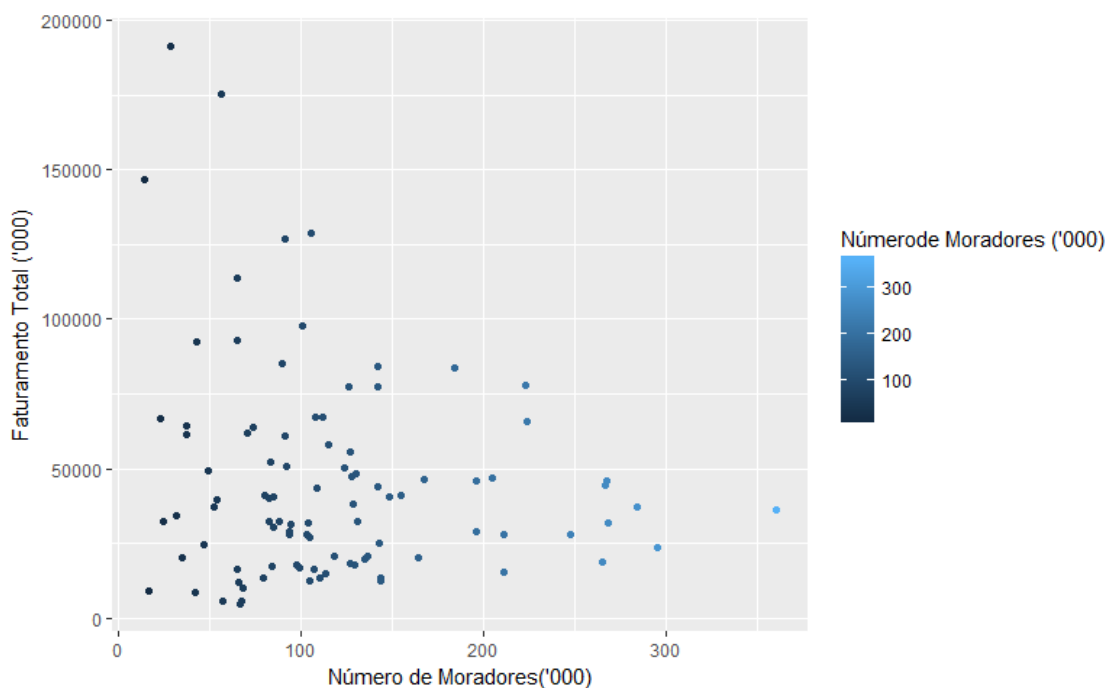
Gráfico 4 – Relação entre o Faturamento Total e o Número de Domicílios dos Distritos



Fonte: Censo (2010), ANAMACO (2015)

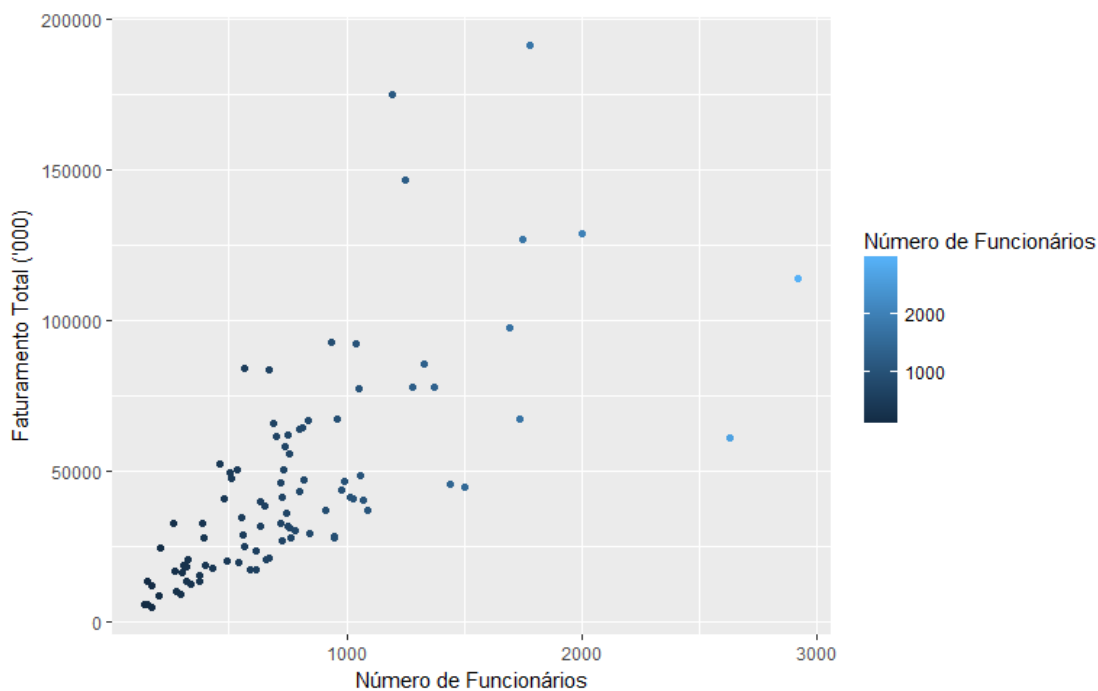
Também se optou por realizar um comparativo com o número de moradores do distrito. O gráfico 5 apresenta uma dispersão muito semelhante à do número de domicílios, ilustrando que são variáveis muito relacionadas, como já esperado. Posteriormente, foi verificado o índice de correlação das variáveis.

Gráfico 5 – Relação entre o Faturamento Total e Número de Moradores dos Distritos



Fonte: Censo (2010), ANAMACO (2015)

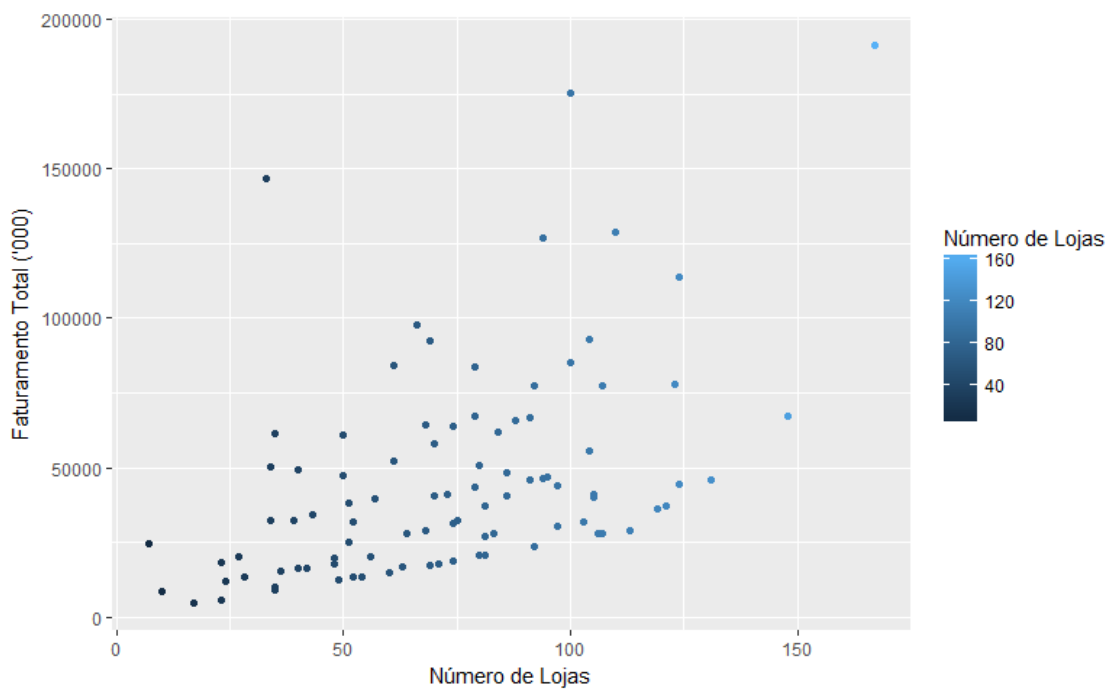
No gráfico 6, percebe-se uma relação mais linear e positiva com o número de funcionários por loja nos 95 distritos e na medida em que os valores aumentam, também cresce a dispersão entre os pontos. Pode-se verificar que o distrito de Brás, com o maior faturamento, também apresenta um dos maiores números de funcionários. Os distritos com o maior número de funcionários foram Lapa e São Miguel, com 2.920 e 2.628 funcionários respectivamente. Os dois distritos são os pontos que aparecem à direita do quadrante, um pouco afastados do restante.

Gráfico 6 – Relação entre o Faturamento Total e o Número de Funcionários nos Distritos

Fonte: ANAMACO (2015)

No cruzamento entre o faturamento e o número de lojas, percebe-se também uma relação linear positiva entre as variáveis, apesar da dispersão ser um pouco mais acentuada do que na relação com o número de funcionários. O distrito Vila Maria, que aparece com o segundo maior número de lojas do setor de materiais de construção, ocupa somente a 16ª posição no ranking de faturamento. Do lado oposto, os distritos com o menor número de lojas são Morumbi e Alto de Pinheiros, distritos com alto poder aquisitivo de São Paulo. O distrito de Brás, novamente, aparece como o distrito mais distoante da média, com os maiores valores de faturamento e número de lojas.

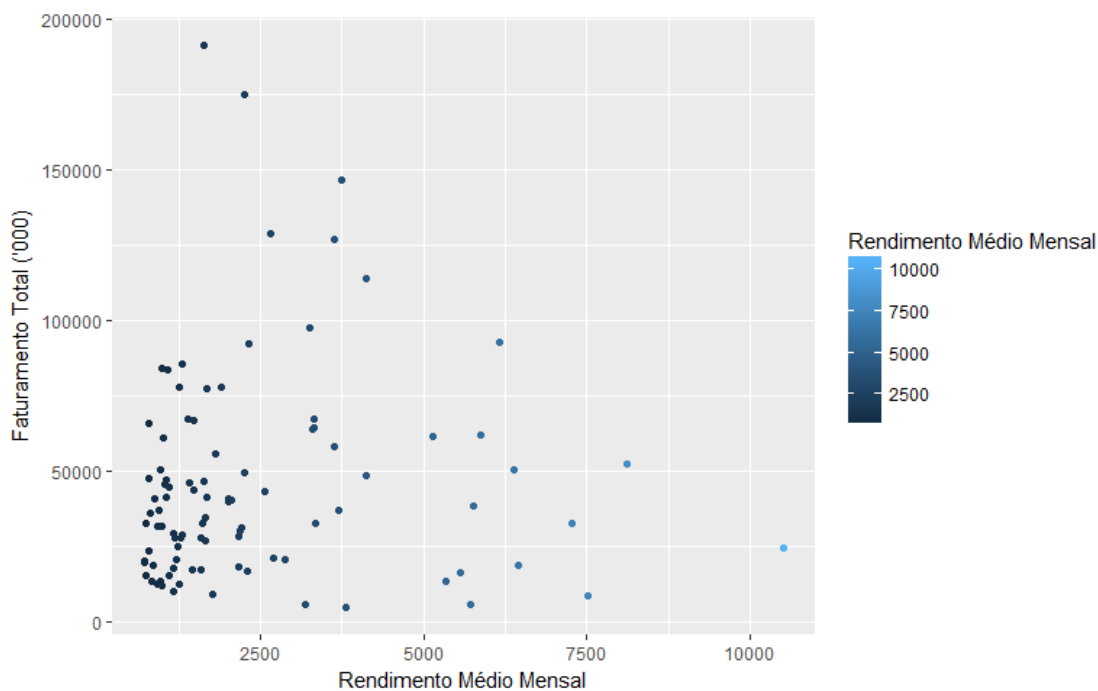
Gráfico 7 – Relação entre o Faturamento Total e o Número de Lojas dos Distritos



Fonte: ANAMACO (2015)

No gráfico 8, ao analisar o rendimento médio mensal dos distritos, percebe-se uma aglomeração de pontos em distritos com rendimento médio de até R\$ 2.500,00. Na medida em que o valor no eixo do x aumenta, o faturamento total nos distritos tem uma leve tendência de queda, apesar da relação não ser muito clara. Morumbi, Moema e Alto de Pinheiros são os distritos com o maior rendimento médio mensal, atingindo valores de 10.503, 8.115 e 7.512 reais mensais, respectivamente. Percebe-se que todos os distritos com um rendimento superior a 5.000 reais mensais apresentam faturamentos menores para o setor de materiais de construção.

Gráfico 8 – Relação entre o Faturamento Total e o Rendimento Médio Mensal nos Distritos



Fonte: Censo (2010), ANAMACO (2015)

Após a análise dois a dois entre as principais variáveis e a variável dependente, também foi realizado o teste de correlação entre todas as variáveis do modelo inicial. Fortes correlações entre variáveis independentes não são desejadas e o efeito é conhecido como colinearidade (entre duas) ou multicolinearidade (entre mais de duas). Esta etapa é importante no processo de seleção e exclusão das variáveis para definir o modelo final.

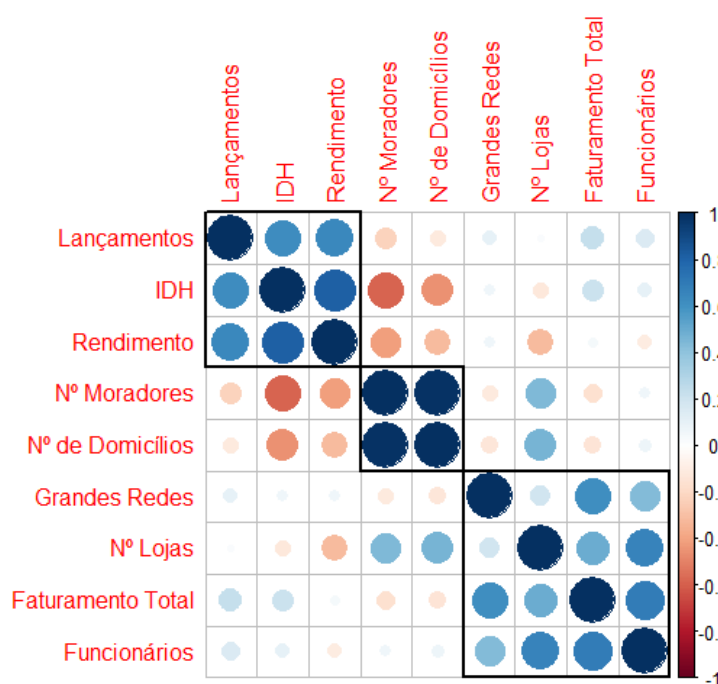
O objetivo da matriz de correlação apresentada no gráfico 11 então, foi identificar as variáveis levantadas que apresentam altos índices de correlação, seja negativa ou positiva. Para valores maiores, foram atribuídas as cores vermelho (valores negativos) e azul (valores positivos) para os círculos, ambas se tornando mais acentuadas na medida em que a correlação fosse mais forte. Na linha diagonal, pode-se perceber a correlação perfeita de 1, visto que é o cruzamento entre as próprias variáveis. O Índice de Desenvolvimento Humano (IDH) foi adicionado para verificar se existiria diferenciação com a variável rendimento médio, já que a renda é um fator considerado na composição do índice. Da mesma maneira, o número de moradores foi adicionado para realizar a comparação com o número de domicílios. Como se pode analisar no

Gráfico 9, ambas tiveram uma correlação positiva muito forte, de 0,81 (IDH vs rendimento médio) e 0,98 (número de moradores vs número de domicílios).

É interessante identificar a relação negativa entre o número de domicílios e o rendimento médio do distrito, indicando que distritos de maior poder aquisitivo são menos populosos e também no sentido oposto - no caso de favelas e aglomerados muito populosos em situações de menor renda. Também é possível ver a relação negativa entre o número de lojas e o rendimento médio do distrito, corroborando com o fato de distritos como Morumbi e Alto de Pinheiros possuírem poucas lojas do setor.

Com relação a variável dependente (“Faturamento Total”) pode-se verificar que no cruzamento com as outras variáveis, o número de funcionários, o número de lojas e a presença de grandes redes apresentam correlação mais representativa. Já o número de lançamentos residenciais e comerciais, o rendimento médio e o número de domicílios apresentaram uma correlação fraca. Para a próxima etapa de análise, optou-se, por excluir as variáveis “IDH” e “número de moradores”, que apresentaram alta correlação com as variáveis de número de domicílios e rendimento médio mensal. Todas as outras variáveis da matriz foram mantidas, visto apresentaram correlação de 0,7 ou abaixo (positivo ou negativo), valor de corte considerado para a exclusão.

Gráfico 9 – Matriz de Correlação entre as Variáveis



Fonte: Censo (2010), ANAMACO (2015), ITC (2016)

5.4 PLANO DE ANÁLISE DE DADOS

Com uma melhor caracterização da base de dados final, foi possível partir para a principal etapa analítica deste trabalho - estabelecer o melhor modelo de atratividade através da análise de regressão linear múltipla. Este método é utilizado para analisar a relação entre uma variável dependente, neste trabalho sendo a atratividade do distrito (faturamento total), e diversas variáveis dependentes, assumindo uma relação linear entre elas. O objetivo por trás da regressão linear é identificar a capacidade do conjunto de variáveis de prever a variabilidade da variável dependente que se está identificando. A análise de regressão é amplamente utilizada em estudos “na área de modelos de previsão em negócios, variando desde modelos econométricos que preveem a economia nacional baseada em inputs (nível de renda, investimentos, etc.), até modelos de performance de uma empresa (...)” (HAIR, 2010).

No contexto do presente trabalho, este método é relevante na medida em que identifica se existe relação entre as variáveis explicativas adotadas e a atratividade do distrito, bem como a grandeza dessa relação. O objetivo final é determinar o melhor modelo que permita determinar a atratividade de cada distrito da cidade de São Paulo para a venda de interruptores.

Portanto, foram realizadas três etapas no processo de análise: inicialmente foi percorrida uma análise das características gerais do universo de materiais de construção na cidade de São Paulo, após as variáveis e lojas serem definidas na seção anterior. Após, foi realizada a análise de regressão múltipla com as variáveis delimitadas, que permitiu verificar se existe significância estatística entre o conjunto de variáveis explicativas e a variável dependente (faturamento do distrito). Nesta segunda etapa, verificou-se a melhor combinação de variáveis explicativas para o modelo final e, assim, sua capacidade de previsão. Por fim, foi desenvolvido o índice de atratividade, segmentando os distritos mais atrativos para a venda de interruptores na cidade de São Paulo. O resultado final do modelo com os índices gerados na previsão foi ilustrado no mapa de atratividade que foi elaborado.

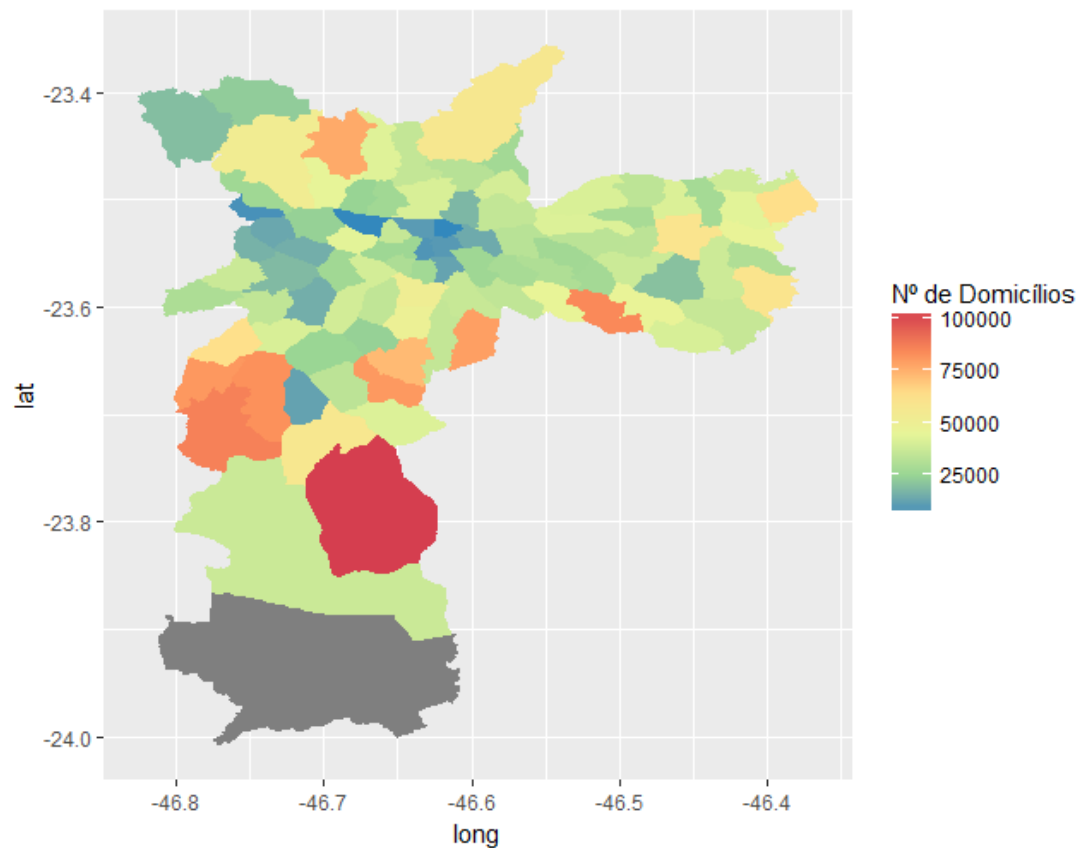
A partir das etapas de análise descritas, foi possível caracterizar o modelo de atratividade proposto de modo a testar sua capacidade preditiva. Após as análises dos resultados, foram apresentadas, então, as considerações finais e limitações do estudo.

6. ANÁLISE DOS RESULTADOS

6.1 CARACTERÍSTICAS GERAIS

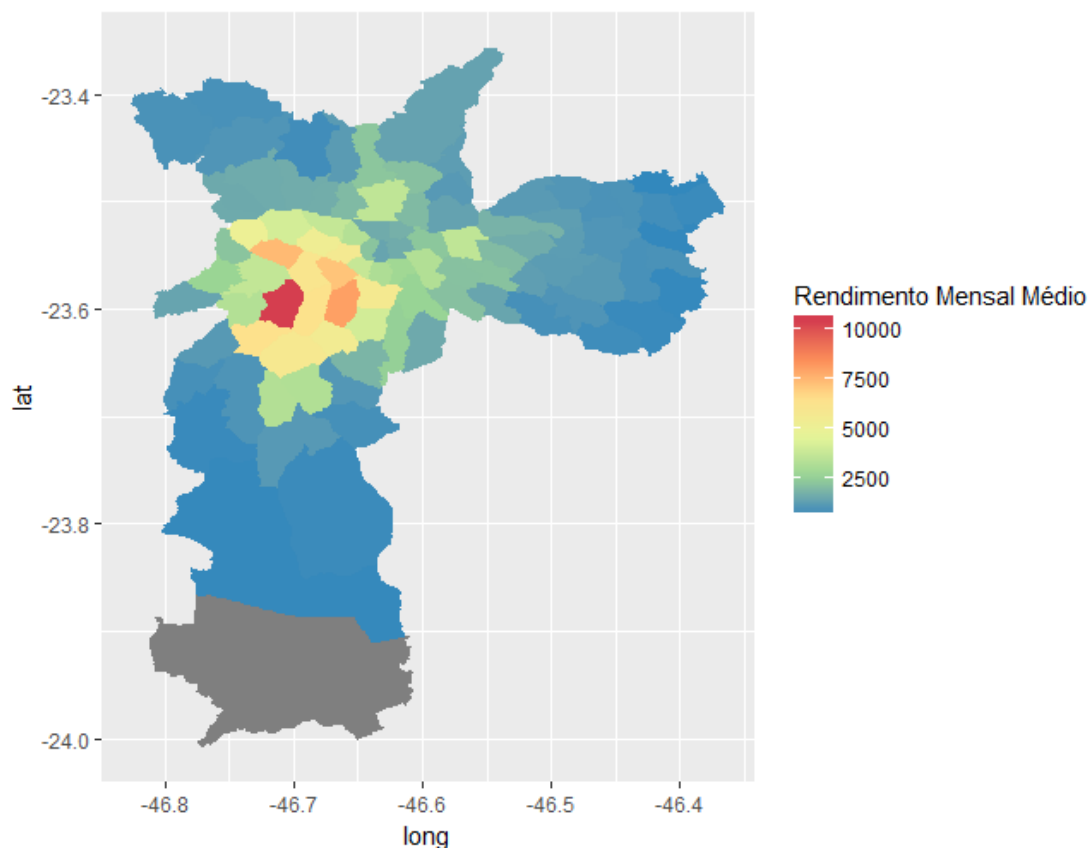
De acordo com os dados do Censo 2010, último dado oficial reportado, a cidade de São Paulo, nosso universo estudado, possui 3.571.928 milhões de domicílios particulares. A estimativa é de que este número já tenha alcançado quase 4 milhões em 2017 de acordo com a Fundação SEADE (<http://produtos.seade.gov.br/produtos/projpop/index.php>). Dentre os distritos da capital, o distrito Grajaú, localizado na Zona Sul, aparece com o maior número de domicílios, representando 2,9% do total da cidade de São Paulo, sendo também o distrito mais populoso e com uma das maiores áreas geográficas (92km²). Por outro lado, o distrito com o menor número de domicílios particulares é o distrito de Pari, também por ter uma das menores áreas geográficas com 2,75km². O distrito fica localizado na região central próxima ao centro histórico da capital e possui somente 5.543 mil domicílios.

Na Figura 9, pode ser visualizado o mapa de densidade de São Paulo e seus distritos. Os distritos com maiores números de domicílios aparecem em tons de laranja e vermelho e de maneira oposta, os distritos em azul, com menos de 2.500 domicílios.

Figura 9 - Mapa de Densidade por Número de Domicílios

Fonte: Censo (2010)

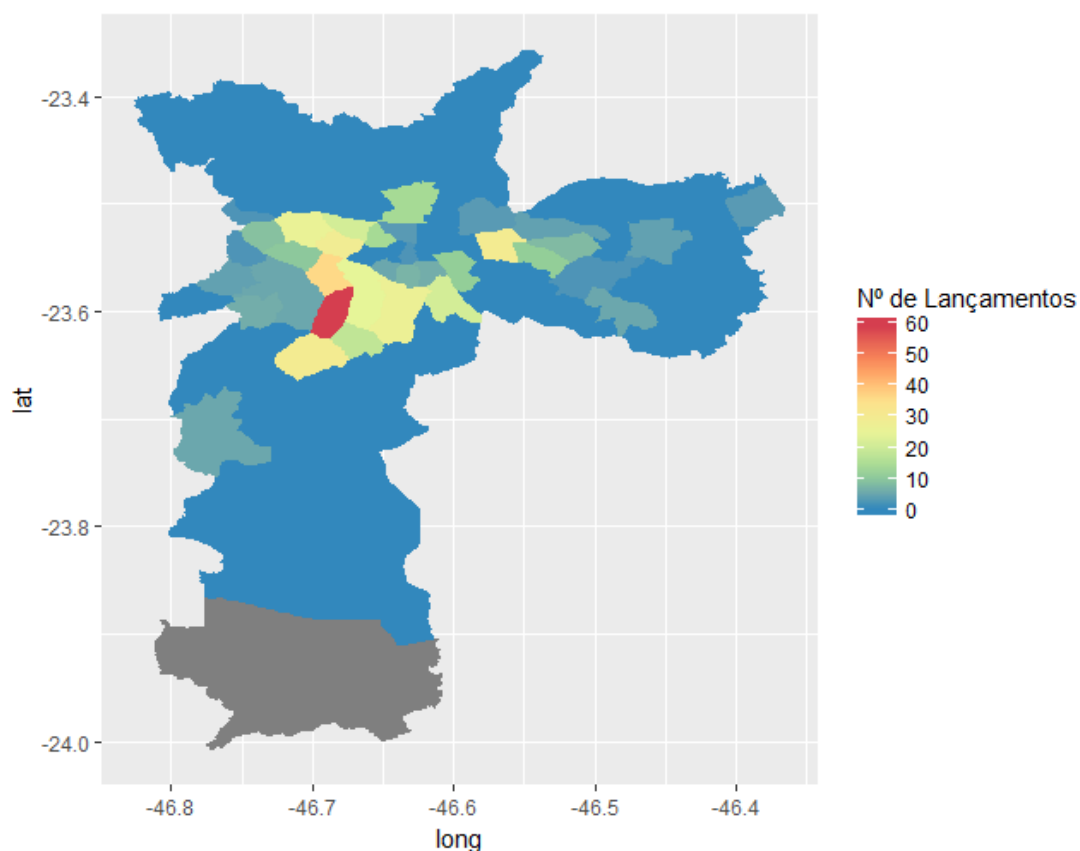
Ao se analisar a Figura 10, fica mais clara a aglomeração entre distritos quando o critério de densidade é o rendimento médio. Pode-se verificar o distrito de Morumbi em vermelho, com o maior rendimento médio, rodeado de distritos com valores semelhantes, nas cores laranja e amarelo. Nos distritos periféricos ao redor, a grande maioria apresenta rendimentos médios de até 2.500 reais/mês. Dez distritos apresentaram um menor rendimento médio abaixo de 1.000 reais/mês, com destaque para os distritos Jardim Helena e Lajeado, que apresentaram os menores rendimentos, de 721,25 reais/mês e 730,80 reais/mês, respectivamente.

Figura 10 - Mapa de Densidade por Rendimento Mensal Médio

Fonte: Censo (2010)

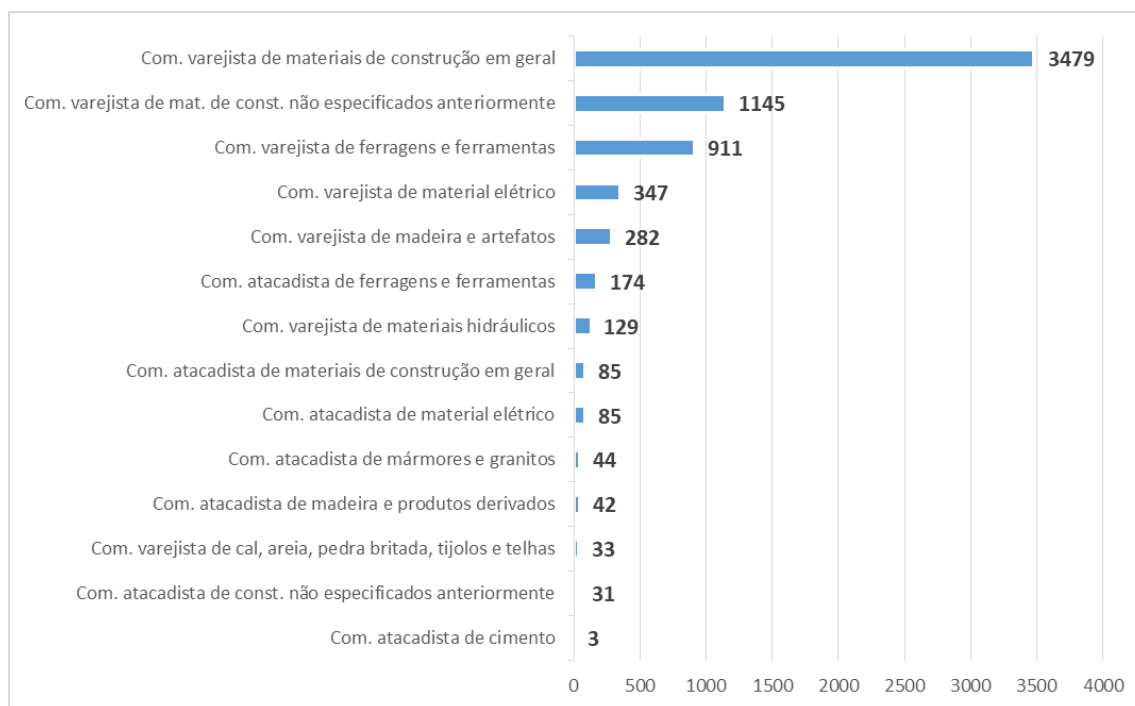
É interessante analisar a similaridade da Figura 11 do mapa de densidade para o número de lançamentos com o mapa anterior do rendimento médio. O distrito com o maior número de lançamentos no ano de 2016 foi Itaim Bibi, com 60 empreendimentos. É necessário ressaltar a limitação da variável, visto que não foram considerados o porte dos empreendimentos, no entanto, percebe-se a relação do maior número de lançamentos e distritos com maior poder aquisitivo. Dos 10 distritos com maior rendimento médio mensal, 6 estão entre os distritos com o maior número de empreendimentos residenciais e comerciais.

Figura 11 - Mapa de Densidade por Número de Lançamentos



Fonte: ITC (2016)

Analisando o universo de materiais de construção, é interessante observar como os diferentes tipos de loja estão presentes em São Paulo em números absolutos. No gráfico 12, fica clara a predominância do varejo de classificação de atividade econômica generalista, que representa 68% das lojas, seguido pelas ferragens (17%) e material elétrico (13%). Como já esperado, os atacadistas aparecem em número menor, sendo o comércio atacado de cimento o CNAE com menor presença. Os distritos com a maior variedade de comércios foi Carrão, seguido de Aricanduva e Freguesia do Ó, no entanto nenhum distrito apresentou todas as 14 atividades econômicas listadas na base de dados.

Gráfico 10 - Número de Comércio por CNAE na Cidade de São Paulo

Fonte: ANAMACO (2015)

Ao se analisar a importância das atividades econômicas dentro das faixas de faturamento na Tabela 3, percebe-se que as duas atividades generalistas também são maioria, sem exceção. A porcentagem dos CNAEs de comércio varejista de materiais de construção em geral (47.44-0/99) e de não especificados anteriormente (47.44-0/05), inclusive aumenta na medida em que as faixas de faturamento aumentam também. Dentro das grandes redes de comércio presentes na faixa de faturamento de 24 milhões ou mais, fazem parte da análise as lojas Construdecor S.A., Leroy Merlin, Votorantim, Saint Gobain e Ferramentas Gerais.

Tabela 3 – Importância da Atividade Econômica por Faixa de Faturamento

CNAE	Até 500 mil	De 500 mil a R\$ 750 mil	De 750 mil a 2,4 milhões	De 2,4 a 10 milhões	De 10 a 23 milhões	Acima de 23 milhões
Comércio varejista de materiais de construção em geral	50,4%	60,4%	67,5%	82,2%	95,0%	64,3%
Comércio varejista de materiais de construção não especificados anteriormente	17,3%	15,4%	15,0%	5,3%	5,0%	32,1%

Comércio varejista de ferragens e ferramentas	14,3%	4,0%	2,5%	1,3%	0,0%	0,0%
Comércio varejista de material elétrico	5,3%	4,0%	7,5%	0,0%	0,0%	3,6%
Comércio varejista de madeira e artefatos	4,3%	6,0%	0,0%	0,7%	0,0%	0,0%
Comércio atacadista de ferragens e ferramentas	2,6%	3,4%	0,0%	5,3%	0,0%	0,0%
Comércio varejista de materiais hidráulicos	2,0%	2,0%	0,0%	0,0%	0,0%	0,0%
Comércio atacadista de material elétrico	1,3%	0,7%	2,5%	1,3%	0,0%	0,0%
Comércio atacadista de madeira e produtos derivados	0,6%	0,7%	2,5%	0,0%	0,0%	0,0%
Comércio atacadista de mármore e granitos	0,6%	2,7%	0,0%	0,7%	0,0%	0,0%
Comércio varejista de cal, areia, pedra britada, tijolos e telhas	0,5%	0,7%	0,0%	0,0%	0,0%	0,0%
Comércio atacadista especializado de materiais de construção não especificados anteriormente	0,5%	0,0%	2,5%	0,7%	0,0%	0,0%
Comércio atacadista de tintas, vernizes e similares	0,3%	0,0%	0,0%	2,6%	0,0%	0,0%
Comércio atacadista de cimento	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Fonte: ANAMACO (2015)

Por fim, abaixo pode ser visualizada a Tabela 4 sintetizando as principais variáveis a nível distrito, para uma melhor comparação e identificação de cada um.

Tabela 4 – Faturamento Anual Total, Rendimento e Domicílios por Distrito

Distritos	Faturamento Anual Total ('000)	Número de Lojas	Rendimento Médio Mensal	Número de Domicílios
BRÁS	R\$ 191.400	167	R\$ 1.644	10.110
REPÚBLICA	R\$ 175.200	100	R\$ 2.256	26.344
BARRA FUNDA	R\$ 146.875	33	R\$ 3.747	5.622
IPIRANGA	R\$ 129.075	110	R\$ 2.649	36.572
TATUAPÉ	R\$ 126.775	94	R\$ 3.630	32.734
LAPA	R\$ 113.775	124	R\$ 4.128	24.085
CAMPO GRANDE	R\$ 97.775	66	R\$ 3.262	33.618
PINHEIROS	R\$ 92.950	104	R\$ 6.170	27.083
BELÉM	R\$ 92.425	69	R\$ 2.328	14.323
ARICANDUVA	R\$ 85.425	100	R\$ 1.312	27.661
VILA JACUÍ	R\$ 84.250	61	R\$ 996	41.658
JARAGUÁ	R\$ 83.575	79	R\$ 1.071	53.238
JABAQUARA	R\$ 77.750	123	R\$ 1.891	73.200
CIDADE LIDER	R\$ 77.650	92	R\$ 1.254	37.561
FREGUESIA DO Ó	R\$ 77.550	107	R\$ 1.677	45.123
VILA MARIA	R\$ 67.200	148	R\$ 1.401	35.242
VILA SÔNIA	R\$ 67.050	79	R\$ 3.321	34.658
SÉ	R\$ 66.850	91	R\$ 1.471	9.098
ITAIM PAULISTA	R\$ 65.625	88	R\$ 797	64.319
SOCORRO	R\$ 64.375	68	R\$ 3.316	12.429
MOOCA	R\$ 63.825	74	R\$ 3.293	26.456
SANTO AMARO	R\$ 61.975	84	R\$ 5.873	25.374
VILA LEOPOLDINA	R\$ 61.625	35	R\$ 5.131	13.589
SÃO MIGUEL	R\$ 61.025	50	R\$ 1.022	27.868
SANTANA	R\$ 58.100	70	R\$ 3.627	40.975
PENHA	R\$ 55.450	104	R\$ 1.816	41.896
MOEMA	R\$ 52.075	61	R\$ 8.116	34.821
ITAIM BIBI	R\$ 50.500	80	R\$ 6.384	39.230
JOSÉ BONIFÁCIO	R\$ 50.425	34	R\$ 966	37.832
JAGUARÉ	R\$ 49.200	40	R\$ 2.255	16.390
SAÚDE	R\$ 48.500	86	R\$ 4.113	49.278
IGUATEMI	R\$ 47.525	50	R\$ 786	36.151
ITAQUERA	R\$ 46.875	95	R\$ 1.069	60.185
PIRITUBA	R\$ 46.450	94	R\$ 1.631	53.342
TREMEMBÉ	R\$ 45.875	91	R\$ 1.419	57.372
JARDIM SÃO LUÍS	R\$ 45.775	131	R\$ 1.045	82.615

CIDADE ADEMAR	R\$ 44.400	124	R\$ 1.114	80.858
SÃO LUCAS	R\$ 43.800	97	R\$ 1.488	45.770
CURSINO	R\$ 43.250	79	R\$ 2.567	35.554
LIMÃO	R\$ 41.200	73	R\$ 1.684	24.861
SÃO MATEUS	R\$ 41.175	105	R\$ 1.053	46.692
VILA CURUÇÁ	R\$ 40.625	70	R\$ 892	43.486
CASA VERDE	R\$ 40.575	86	R\$ 2.006	27.035
CARRÃO	R\$ 40.225	105	R\$ 2.049	27.115
VILA GUILHERME	R\$ 39.750	57	R\$ 2.017	17.750
VILA MARIANA	R\$ 38.250	51	R\$ 5.759	51.822
BUTANTÃ	R\$ 37.050	81	R\$ 3.693	18.542
SAPOPEMBA	R\$ 36.950	121	R\$ 941	84.686
GRAJAÚ	R\$ 36.075	119	R\$ 808	103.070
BOM RETIRO	R\$ 34.250	43	R\$ 1.648	10.620
SANTA CECÍLIA	R\$ 32.525	75	R\$ 3.346	35.951
PARELHEIROS	R\$ 32.500	39	R\$ 750	37.140
JAGUARA	R\$ 32.400	34	R\$ 1.606	7.935
JARDIM PAULISTA	R\$ 32.300	34	R\$ 7.265	39.701
CAPÃO REDONDO	R\$ 31.700	103	R\$ 934	81.033
GUAIANASES	R\$ 31.600	52	R\$ 992	30.547
VILA FORMOSA	R\$ 31.150	74	R\$ 2.221	30.412
ÁGUA RASA	R\$ 30.200	97	R\$ 2.182	28.652
CIDADE DUTRA	R\$ 29.000	113	R\$ 1.172	57.691
PONTE RASA	R\$ 28.900	68	R\$ 1.302	28.667
VILA PRUDENTE	R\$ 28.025	83	R\$ 2.172	34.707
JAÇANÃ	R\$ 27.900	64	R\$ 1.277	27.564
SACOMÃ	R\$ 27.875	107	R\$ 1.599	79.473
CAMPO LIMPO	R\$ 27.825	106	R\$ 1.188	63.867
VILA MATILDE	R\$ 26.950	81	R\$ 1.661	33.491
CACHOEIRINHA	R\$ 25.025	51	R\$ 1.236	42.683
MORUMBI	R\$ 24.500	7	R\$ 10.504	15.448
JARDIM ÂNGELA	R\$ 23.375	92	R\$ 783	86.894
RIO PEQUENO	R\$ 20.750	80	R\$ 2.697	37.308
CANGAIBA	R\$ 20.625	81	R\$ 1.214	41.250
CAMBUCCI	R\$ 20.350	27	R\$ 2.876	12.645
LAJEADO	R\$ 19.950	56	R\$ 731	46.653
JARDIM HELENA	R\$ 19.650	48	R\$ 721	38.247
BRASILÂNDIA	R\$ 18.500	74	R\$ 869	76.997
VILA ANDRADE	R\$ 18.400	23	R\$ 6.458	40.826
TUCURUVI	R\$ 17.950	48	R\$ 2.173	32.540

VILA MEDEIROS	R\$ 17.750	71	R\$ 1.173	39.905
SÃO DOMINGOS	R\$ 17.250	69	R\$ 1.593	26.617
RAPOSO TAVARES	R\$ 16.875	63	R\$ 1.457	29.865
MANDAQUI	R\$ 16.450	42	R\$ 2.306	35.416
CAMPO BELO	R\$ 16.325	40	R\$ 5.568	24.049
CIDADE TIRADENTES	R\$ 15.325	36	R\$ 749	60.740
ERMELINO MATARAZZO	R\$ 15.000	60	R\$ 1.102	34.029
SÃO RAFAEL	R\$ 13.500	54	R\$ 847	40.589
PERDIZES	R\$ 13.375	52	R\$ 5.349	43.635
PERUS	R\$ 13.325	28	R\$ 966	22.788
PEDREIRA	R\$ 12.250	49	R\$ 925	42.055
ARTUR ALVIM	R\$ 12.250	49	R\$ 1.268	33.712
ANHANGUERA	R\$ 11.950	24	R\$ 984	19.437
PARQUE DO CARMO	R\$ 10.075	35	R\$ 1.179	20.285
PARI	R\$ 9.125	35	R\$ 1.772	5.543
ALTO DE PINHEIROS	R\$ 8.450	10	R\$ 7.513	15.408
LIBERDADE	R\$ 5.750	23	R\$ 3.187	27.314
CONSOLAÇÃO	R\$ 5.750	23	R\$ 5.724	26.339
BELA VISTA	R\$ 4.625	17	R\$ 3.803	29.967

Fonte: ANAMACO (2015)

6.2 MODELO DE ATRATIVIDADE FINAL

Para se definir o modelo de atratividade final, foi realizada a análise de regressão múltipla com a combinação de variáveis delimitadas nas seções anteriores – o número de lojas, o número de funcionários, o rendimento médio mensal, o número de domicílios, o número de lançamentos residenciais e comerciais e a presença de grandes redes. A análise de regressão múltipla permite verificar se existe significância estatística entre o conjunto de variáveis explicativas e a variável dependente (faturamento do distrito). O objetivo aqui foi determinar a combinação entre as variáveis que define o melhor modelo com capacidade de prever o faturamento dos distritos.

Assim, foi realizado no RStudio o processo de seleção *stepwise backward*, que é um método automatizado de seleção entre modelos de regressão, no qual se inicia um teste com todas as variáveis e posteriormente são realizados testes com a exclusão de

cada uma das variáveis por vez (por isso *backward*). A partir disso, o modelo que apresentar melhor qualidade comparado aos outros é proposto.

O critério utilizado aqui foi o Critério de Informação de Akaike (AIC), que determina o modelo que melhor se ajusta à variável dependente. Para cada possível modelo, é atribuído um valor de AIC e o modelo que apresentar o menor valor é o que apresenta melhor ajuste. A partir deste critério e das variáveis delimitadas, o modelo que melhor se ajustou apresentou um valor AIC de 3.206,5, o menor valor entre os modelos e a única variável excluída foi o rendimento médio mensal, pois não apresentou significância estatística.

Assim, avaliando o resultado do modelo final de previsão de atratividade na Tabela 5, pode-se perceber na coluna ANOVA que ele se mostrou estatisticamente significativo – apresentando uma Sig. menor que 5% e resultados plausíveis. Ao se verificar a sumarização do modelo, o coeficiente de determinação R^2 é de 0,674, ou seja, o conjunto de variáveis tem a capacidade de explicar 67% da variabilidade da variável dependente.

Tabela 5 – Resumo do Modelo Final

SUMARIZAÇÃO DO MODELO			ANOVA	
R ²	R ² AJUSTADO	ERRO PADRÃO DA ESTIMATIVA	F	Sig.
0,674	0,655	20.071.000	36,84	0,000

Fonte: Adaptado de RStudio

Na Tabela 6, pode-se verificar que quatro das cinco variáveis explicativas se mostraram estatisticamente significativas (Sig.<5%)– a presença de grandes redes, o número de funcionários, o número de lojas e o número de domicílios. Para a variável lançamentos residenciais e comerciais, não foi possível realizar inferências, visto que não atingiu o critério de significância estatística.

Pode-se verificar através do sinal dos coeficientes que somente o número de domicílios apresentou relação negativa com a variável dependente. Desta forma, o modelo ilustra que a cada nova unidade de domicílio em um distrito, o faturamento para

materiais de construção, e interruptores, cai em 466,90 reais/ano. Apesar do resultado talvez não esperado, é possível fazer algumas suposições: regiões muito residenciais, podem apresentar um menor número de comércios de materiais de construção; além disso, comércios maiores, podem buscar uma alocação em regiões mais afastadas, devido a ganhos de custo do terreno, como já mencionado na seção 5.3. Vale lembrar, que o *scatterplot* realizado na mesma seção, entre faturamento total do distrito e o número de domicílios relação, apesar de uma grande dispersão já apresentava sinais de relação negativa. Este resultado também traz a luz a questão do deslocamento do consumidor e a distância que ele está disposto a percorrer. Visto que a relação entre o faturamento dos comércios e o número de domicílios é negativa, surge a hipótese de que os distritos com faturamentos maiores poderiam estar atendendo demandas de outros distritos.

Já o número de lojas, o número de funcionários e a presença de grandes redes tiveram relação positiva com o a variável dependente. Para o número de lojas, o aumento de uma unidade, resulta num aumento de 359.190,20 mil reais/ano no faturamento total de um distrito. No caso do número de funcionários, o acréscimo de um novo funcionário resulta em uma variação de 22.880,00 mil reais/ano no faturamento do distrito.

A variável “Grandes Redes”, por se tratar de uma variável nominal – sendo zero para ausência de grandes redes; e 1 para presença de ao menos uma grande rede no distrito – deve ter seu coeficiente interpretado como a diferença entre o faturamento de distritos que apresentam grandes redes e distritos que não apresentam. Assim, percebe-se que esta diferença é em média de 26.806.082 milhões de reais/ano a mais para distritos com grandes redes, ilustrando o grande impacto destas no faturamento da região.

Tabela 6 – Coeficientes do Modelo Final

VARIÁVEL	COEFICIENTE	ERRO PADRÃO	t	Sig.
(Constante)	8.282.158	5.883.678	1,408	0,162
Grandes Redes	26.806.082	5.050.186	5,308	0,000
Número de Funcionários	22.880	6.592	3,471	0,000
Número de Lojas	359.190,20	106.785	3,364	0,001
Número de Domicílios	-466,90	131	-3,561	0,000

Fonte: Adaptado de RStudio

Dessa forma, o modelo final de atratividade proposto, é definido na equação abaixo:

$$Ad = Smc + G + F + D + Lrc \quad (4)$$

6.3 TESTE DO MODELO E CRIAÇÃO DO ÍNDICE DE ATRATIVIDADE FINAL

A partir das análises de regressão e da delimitação do modelo final de atratividade, a etapa final de teste do modelo e elaboração do índice de atratividade foi realizada. O objetivo aqui foi de verificar a sua assertividade de previsão comparado aos dados reais de faturamento total de cada distrito. Após, foi elaborado o índice de atratividade, que permitiu identificar de maneira mais clara os distritos mais atrativos para a venda de interruptores elétricos. Para tal, foi calculado o faturamento médio anual dos distritos na cidade de São Paulo, que apresentou o valor de 45,31 milhões de reais/ano, e o faturamento de cada distrito foi, então, dividido por este valor e multiplicados por 100, de modo a se obter o índice.

$$\text{Índice de Atratividade} = \text{Faturamento do Distrito} / \text{Faturamento Médio da cidade} \quad (3)$$

Na coluna “Previsão”, da tabela 6, é possível analisar a previsão de faturamento total do modelo final, comparada ao faturamento real de cada distrito. Dos 95 distritos analisados, 20 apresentaram predição com diferença de 10% acima ou abaixo (valores em verde na coluna “% Representada”). Os distritos com representação abaixo de 90%

estão caracterizados com valores em vermelho e os distritos superestimados (>110%) estão com os valores destacados em laranja.

Pode-se perceber na Tabela 7, uma maior concentração de valores subestimados em distritos que apresentaram um faturamento maior. Já os valores superestimados, apresentaram concentração maior nos distritos de faturamento mais baixo. Os distritos que apresentaram a melhor previsão, entre 98% e 102%, foram: Pinheiros, Jardim Helena, Iguatemi, São Lucas, Sapopemba, Anhanguera, Freguesia do Ó, Vila Leopoldina, Jaçanã e Tucuruvi.

Tabela 7 – Comparação entre Scores Reais e Previsão do Modelo

Distritos	Faturamento Anual Total ('000)	Previsão do Modelo ('000)	% Representada
BRÁS	R\$ 191.400	R\$ 131.057	68%
REPÚBLICA	R\$ 175.200	R\$ 59.176	34%
BARRA FUNDA	R\$ 146.875	R\$ 80.492	55%
IPIRANGA	R\$ 129.075	R\$ 110.953	86%
TATUAPÉ	R\$ 126.775	R\$ 104.485	82%
LAPA	R\$ 113.775	R\$ 144.658	127%
CAMPO GRANDE	R\$ 97.775	R\$ 81.859	84%
PINHEIROS	R\$ 92.950	R\$ 94.615	102%
BELÉM	R\$ 92.425	R\$ 77.027	83%
ARICANDUVA	R\$ 85.425	R\$ 89.252	104%
VILA JACUÍ	R\$ 84.250	R\$ 50.432	60%
JARAGUÁ	R\$ 83.575	R\$ 53.894	64%
JABAQUARA	R\$ 77.750	R\$ 74.405	96%
CIDADE LIDER	R\$ 77.650	R\$ 82.718	107%
FREGUESIA DO Ó	R\$ 77.550	R\$ 76.480	99%
VILA MARIA	R\$ 67.200	R\$ 112.608	168%
VILA SÔNIA	R\$ 67.050	R\$ 71.364	106%
SÉ	R\$ 66.850	R\$ 56.577	85%
ITAIM PAULISTA	R\$ 65.625	R\$ 53.572	82%
SOCORRO	R\$ 64.375	R\$ 72.335	112%
MOOCA	R\$ 63.825	R\$ 71.648	112%
SANTO AMARO	R\$ 61.975	R\$ 81.563	132%
VILA LEOPOLDINA	R\$ 61.625	R\$ 60.562	98%
SÃO MIGUEL	R\$ 61.025	R\$ 100.167	164%
SANTANA	R\$ 58.100	R\$ 63.015	108%
PENHA	R\$ 55.450	R\$ 71.638	129%
MOEMA	R\$ 52.075	R\$ 33.220	64%
ITAIM BIBI	R\$ 50.500	R\$ 57.245	113%

JOSÉ BONIFÁCIO	R\$ 50.425	R\$ 41.857	83%
JAGUARÉ	R\$ 49.200	R\$ 27.303	55%
SAÚDE	R\$ 48.500	R\$ 50.271	104%
IGUATEMI	R\$ 47.525	R\$ 47.908	101%
ITAQUERA	R\$ 46.875	R\$ 61.332	131%
PIRITUBA	R\$ 46.450	R\$ 39.818	86%
TREMEMBÉ	R\$ 45.875	R\$ 57.487	125%
JARDIM SÃO LUÍS	R\$ 45.775	R\$ 49.692	109%
CIDADE ADEMAR	R\$ 44.400	R\$ 49.393	111%
SÃO LUCAS	R\$ 43.800	R\$ 44.087	101%
CURSINO	R\$ 43.250	R\$ 65.101	151%
LIMÃO	R\$ 41.200	R\$ 39.462	96%
SÃO MATEUS	R\$ 41.175	R\$ 49.242	120%
VILA CURUÇÁ	R\$ 40.625	R\$ 50.844	125%
CASA VERDE	R\$ 40.575	R\$ 50.049	123%
CARRÃO	R\$ 40.225	R\$ 61.824	154%
VILA GUILHERME	R\$ 39.750	R\$ 35.021	88%
VILA MARIANA	R\$ 38.250	R\$ 27.131	71%
BUTANTÃ	R\$ 37.050	R\$ 51.315	139%
SAPOEMBA	R\$ 36.950	R\$ 37.148	101%
GRAJAÚ	R\$ 36.075	R\$ 19.953	55%
BOM RETIRO	R\$ 34.250	R\$ 59.297	173%
SANTA CECÍLIA	R\$ 32.525	R\$ 40.007	123%
PARELHEIROS	R\$ 32.500	R\$ 37.752	116%
JAGUARA	R\$ 32.400	R\$ 26.304	81%
JARDIM PAULISTA	R\$ 32.300	R\$ 16.714	52%
CAPÃO REDONDO	R\$ 31.700	R\$ 21.886	69%
GUAIANASES	R\$ 31.600	R\$ 29.859	94%
VILA FORMOSA	R\$ 31.150	R\$ 37.962	122%
ÁGUA RASA	R\$ 30.200	R\$ 47.663	158%
CIDADE DUTRA	R\$ 29.000	R\$ 41.203	142%
PONTE RASA	R\$ 28.900	R\$ 32.091	111%
VILA PRUDENTE	R\$ 28.025	R\$ 43.605	156%
JAÇANÃ	R\$ 27.900	R\$ 27.348	98%
SACOMÃ	R\$ 27.875	R\$ 31.304	112%
CAMPO LIMPO	R\$ 27.825	R\$ 33.998	122%
VILA MATILDE	R\$ 26.950	R\$ 41.196	153%
CACHOEIRINHA	R\$ 25.025	R\$ 19.602	78%
MORUMBI	R\$ 24.500	R\$ 36.947	151%
JARDIM ÂNGELA	R\$ 23.375	R\$ 16.652	71%
RIO PEQUENO	R\$ 20.750	R\$ 36.409	175%
CANGAIBA	R\$ 20.625	R\$ 33.197	161%
CAMBUCCI	R\$ 20.350	R\$ 21.674	107%
LAJEADO	R\$ 19.950	R\$ 17.897	90%

JARDIM HELENA	R\$ 19.650	R\$ 20.000	102%
BRASILÂNDIA	R\$ 18.500	R\$ 8.068	44%
VILA ANDRADE	R\$ 18.400	R\$ 4.531	25%
TUCURUVI	R\$ 17.950	R\$ 17.585	98%
VILA MEDEIROS	R\$ 17.750	R\$ 24.971	141%
SÃO DOMINGOS	R\$ 17.250	R\$ 34.689	201%
RAPOSO TAVARES	R\$ 16.875	R\$ 30.514	181%
MANDAQUI	R\$ 16.450	R\$ 13.012	79%
CAMPO BELO	R\$ 16.325	R\$ 24.793	152%
CIDADE TIRADENTES	R\$ 15.325	R\$ 1.482	10%
ERMELINO MATARAZZO	R\$ 15.000	R\$ 22.504	150%
SÃO RAFAEL	R\$ 13.500	R\$ 16.097	119%
PERDIZES	R\$ 13.375	R\$ 25.656	192%
PERUS	R\$ 13.325	R\$ 11.202	84%
PEDREIRA	R\$ 12.250	R\$ 13.983	114%
ARTUR ALVIM	R\$ 12.250	R\$ 17.832	146%
ANHANGUERA	R\$ 11.950	R\$ 11.787	99%
PARQUE DO CARMO	R\$ 10.075	R\$ 17.630	175%
PARI	R\$ 9.125	R\$ 25.061	275%
ALTO DE PINHEIROS	R\$ 8.450	R\$ 12.965	153%
LIBERDADE	R\$ 5.750	R\$ 9.589	167%
CONSOLAÇÃO	R\$ 5.750	R\$ 7.771	135%
BELA VISTA	R\$ 4.625	R\$ 6.108	132%

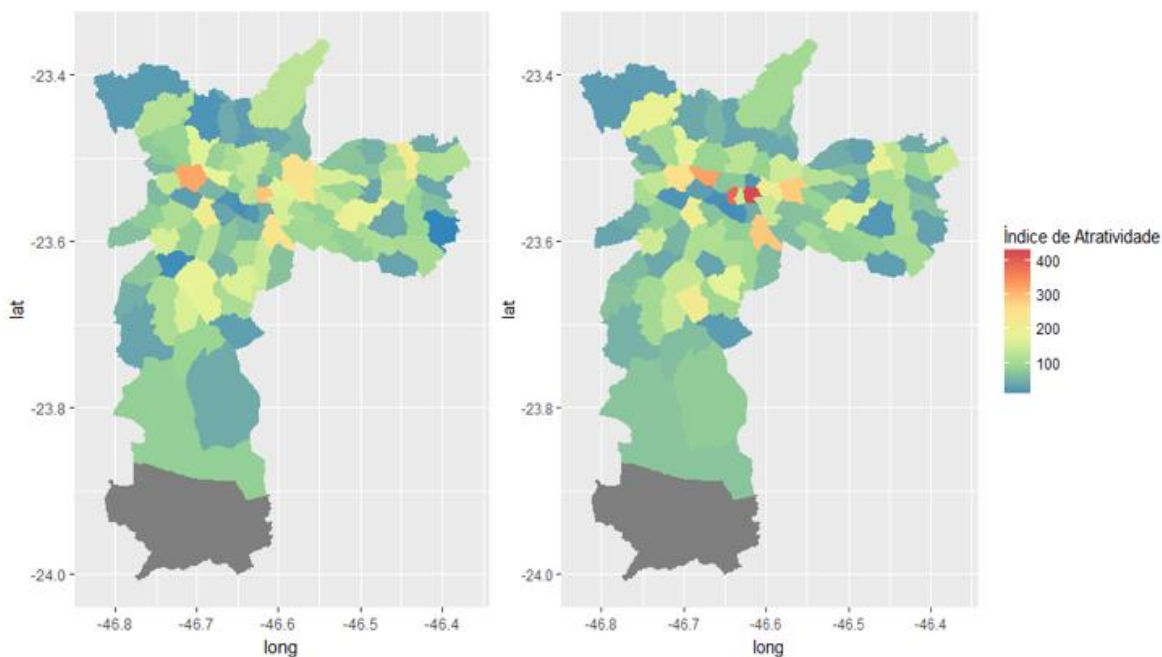
Fonte: ANAMACO (2015)

Após esta primeira análise, foi desenvolvido o índice de atratividade para cada distrito de São Paulo capital. O resultado final do modelo com os índices gerados na previsão foi ilustrado no mapa de atratividade final, onde se pode verificar de uma maneira visual as regiões mais atrativas para a venda de materiais de construção e interruptores, de acordo com a previsão do modelo. Na figura 12, pode-se ver lado a lado o mapa de atratividade estimado pelo modelo (à esquerda), aqui denominado “mapa de previsão” e o mapa de atratividade criado através dos faturamentos reais (à direita), aqui denominado “mapa real”.

Os distritos em tons amarelo, laranja e vermelho foram os que apresentaram uma atratividade superior à média da cidade de São Paulo (valores superiores a 100). Já as regiões em verde e azul apresentaram uma atratividade abaixo da média da cidade. Visualmente, o modelo do índice atratividade elaborado parece apresentar uma boa performance. Pode-se verificar, no entanto, que o modelo previu alguns distritos com um índice de atratividade acima do real, na região central (em amarelo). Por outro lado,

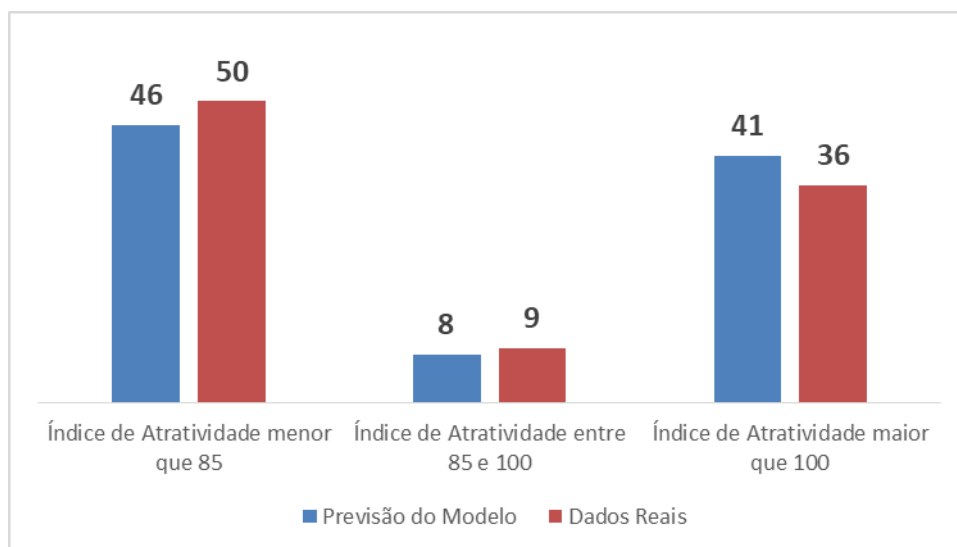
fica claro que o mapa de previsão não apresenta distritos com densidade máxima (em laranja e vermelho), como Brás e República, visualizados no mapa real.

Figura 12 – Mapa de Atratividade do Modelo Final (à esq.) e Mapa Real (à dir.)



Fonte: Criação própria

De acordo com a previsão do modelo, 41 distritos apresentaram uma atratividade acima da média da cidade, ou seja, seriam regiões alvo para a distribuição de um fabricante. Dentre eles, os três distritos com maior índice foram Lapa (319), Brás (289) e Vila Maria (249), conforme o Gráfico 11. Comparando os as barras azuis e vermelhas, fica mais claro que o modelo previu um número maior de distritos com índice superior à 100 do que os dados reais. De maneira geral, pode-se dizer que o modelo de previsão superestimou os índices de atratividade dos distritos. No entanto, fica clara a proporção em ambos os gráficos, de um pequeno número de distritos com índice de atratividade entre 85 e 100 e um maior peso para os dois extremos. Diferentemente da previsão do modelo, os índices de atratividade gerados apresentaram o distrito de Brás (423), República (387) e Barra Funda (324) como os mais atrativos para a venda de interruptores.

Gráfico 11 - Número de Pontos de Venda por Índice de Atratividade

Fonte: Criação Própria

Segue abaixo, na Tabela 8 um compilado dos índices de atratividade previstos no modelo e os índices elaborados a partir dos dados reais de faturamento anual:

Tabela 8 - Índice de Atratividade por Distrito

Distrito	Índice de Atratividade Real	Índice de Atratividade Previsto
BRÁS	423	289
REPÚBLICA	387	131
BARRA FUNDA	324	178
IPIRANGA	285	245
TATUAPÉ	280	231
LAPA	251	319
CAMPO GRANDE	216	181
PINHEIROS	205	209
BELÉM	204	170
ARICANDUVA	189	197
VILA JACUÍ	186	111
JARAGUÁ	185	119
JABAQUARA	172	164
CIDADE LIDER	171	183
FREGUESIA DO Ó	171	169
VILA MARIA	148	249
VILA SÔNIA	148	158
SÉ	148	125

ITAIM PAULISTA	145	118
SOCORRO	142	160
MOOCA	141	158
SANTO AMARO	137	180
VILA LEOPOLDINA	136	134
SÃO MIGUEL	135	221
SANTANA	128	139
PENHA	122	158
MOEMA	115	73
ITAIM BIBI	112	126
JOSÉ BONIFÁCIO	111	92
JAGUARÉ	109	60
SAÚDE	107	111
IGUATEMI	105	106
ITAQUERA	104	135
PIRITUBA	103	88
TREMEMBÉ	101	127
JARDIM SÃO LUÍS	101	110
CIDADE ADEMAR	98	109
SÃO LUCAS	97	97
CURSINO	96	144
SÃO MATEUS	91	109
LIMÃO	91	87
VILA CURUÇÁ	90	112
CASA VERDE	90	111
CARRÃO	89	137
VILA GUILHERME	88	77
VILA MARIANA	84	60
BUTANTÃ	82	113
SAPOEMBA	82	82
GRAJAÚ	80	44
BOM RETIRO	76	131
SANTA CECÍLIA	72	88
PARELHEIROS	72	83
JAGUARA	72	58
JARDIM PAULISTA	71	37
GUAIANASES	70	66
CAPÃO REDONDO	70	48
VILA FORMOSA	69	84
ÁGUA RASA	67	105
CIDADE DUTRA	64	91
PONTE RASA	64	71
VILA PRUDENTE	62	96
SACOMÃ	62	69

JAÇANÃ	62	60
CAMPO LIMPO	61	75
VILA MATILDE	60	91
CACHOEIRINHA	55	43
MORUMBI	54	82
JARDIM ÂNGELA	52	37
RIO PEQUENO	46	80
CANGAIBA	46	73
CAMBUCI	45	48
LAJEADO	44	40
JARDIM HELENA	43	44
BRASILÂNDIA	41	18
VILA ANDRADE	41	10
TUCURUVI	40	39
VILA MEDEIROS	39	55
SÃO DOMINGOS	38	77
RAPOSO TAVARES	37	67
CAMPO BELO	36	55
MANDAQUI	36	29
CIDADE TIRADENTES	34	3
ERMELINO MATARAZZO	33	50
PERDIZES	30	57
SÃO RAFAEL	30	36
PERUS	29	25
ARTUR ALVIM	27	39
PEDREIRA	27	31
ANHANGUERA	26	26
PARQUE DO CARMO	22	39
PARI	20	55
ALTO DE PINHEIROS	19	29
LIBERDADE	13	21
CONSOLAÇÃO	13	17
BELA VISTA	10	13

Fonte: ANAMACO (2015)

Dessa forma, o modelo aparenta um moderado nível de previsão para os índices de atratividade a partir das variáveis utilizadas e os coeficientes apresentados foram plausíveis de interpretação. Através dos mapas gerados, pode-se dizer que existe uma aparente aglomeração de distritos mais atrativos na região central da cidade, no entanto não é possível afirmar esta relação. Esta avaliação poderia ser alvo de um novo estudo acerca deste tema.

7. CONSIDERAÇÕES FINAIS

Através do presente estudo, pode-se concluir que o acesso a dados secundários pode ter um impacto significativo na performance e estratégia de uma empresa quando o objetivo é delimitar as melhores regiões para vendas de seus produtos. No contexto do setor de materiais elétricos, mais especificamente para a venda de interruptores, buscou-se caracterizar a atratividade de cada distrito da capital de São Paulo estabelecendo uma hierarquia entre as 96 unidades territoriais através de um índice de atratividade.

A primeira etapa se resumiu na coleta de dados secundários de diferentes fontes que pudessem ser relacionadas ao setor e à demografia da região. A cidade de São Paulo e seus distritos foi definida como alvo do estudo devido à sua representatividade e a maior disponibilidade de dados através da cátedra Tramontina Eletrik.

A partir do processo de coleta de dados, que envolveu grandes volumes de dados, foi realizado um extenso processo de mineração e seleção de dados a partir dos objetivos estabelecidos. Este processo envolveu técnicas de georreferenciamento e análises de dados em diferentes plataformas e softwares livres – o software Rstudio, para computação voltado para gráficos e computação estatísticos; o software denominado Quantum GIS 2.4 (QGIS), o site BatchGeo, voltados para de informação geográficas; e, por fim, o Microsoft Excel. Junto disso, foram selecionados os dados de interesse para o propósito do trabalho – estabelecer um índice de atratividade entre regiões para a venda de interruptores -, o que necessitou um bom conhecimento do setor de materiais de construção em geral.

Na elaboração do modelo de atratividade inicial, baseado inicialmente nos estudos de Dolega, Pavlis e Singleton (2016), foram utilizados os dados de faturamento do comércio como proxy para atratividade e seis variáveis explicativas - o número de pontos de venda de materiais de construção por distrito (Smc), o número de funcionários (F), a presença de grandes redes (G), o número de domicílios (D), os lançamentos residenciais e comerciais (Lrc), a renda média no distrito (Rmd).

A partir disso foram realizadas as análises descritivas e a análise de regressão linear, principal etapa do presente trabalho que delimitou o modelo final composto por quatro variáveis - o número de pontos de venda de materiais de construção por distrito (Smc), o número de funcionários (F), a presença de grandes redes (G) e o número de

domicílios (D). Estas variáveis foram utilizadas para prever o faturamento dos distritos de São Paulo para o setor de materiais de construção e, assim, os distritos mais atrativos para a venda de interruptores.

Os resultados à luz da literatura sobre modelos de atratividade trazem informações relevantes sobre o setor de materiais de construção. Primeiro, o fato da relação entre número de domicílios e a atratividade do distrito ser negativa, permite inferir que para estas categorias de produtos, a conveniência, ou seja, estar próximo do consumidor, não é clara. Bucklin, Siddarth e Silva-Risso (2008), mencionam que existem muitos estudos na área de produtos não duráveis (conveniência), mas pouco se sabe sobre produtos com uma distribuição mais seleta. Através dos resultados aqui obtidos, surge a questão sobre em qual ocasião o consumidor busca estes produtos. Assim, um modelo de Huff completo, que considera a atratividade de cada loja atrelada a sua distância dos compradores poderia ser relevante para identificar a zona de influência de lojas do setor.

Por fim, o modelo final apresentou um R^2 de 0,674, ou seja, a capacidade de prever 67,4% da variabilidade do faturamento entre os distritos. A performance do modelo apresentou limitações, mas demonstrou que é possível obter resultados significativos quando se tem acesso a uma grande quantidade de dados.

Portanto, além do modelo final de previsão elaborado, espera-se contribuir cientificamente para esse campo de estudo, ainda pouco explorado no Brasil. Através, principalmente da revisão da literatura realizada e do contexto real de implementação buscou-se, por fim, obter uma abordagem prática que possa auxiliar os gestores na tomada de decisão.

8. LIMITAÇÕES DO TRABALHO

Se faz necessário ressaltar aqui as limitações do presente trabalho, que devem ser consideradas ao analisar seus resultados. Como principal aspecto relacionado ao método proposto, a variável dependente do modelo de atratividade definida como os pontos médios de faturamento das lojas apresentou uma generalização, principalmente entre as maiores faixas de faturamento. Além disso, a presença majoritária de lojas com a menor faixa de faturamento, acarretou em uma variável dependente com pouca variabilidade, que apresentou uma forma de distribuição normal com assimetria positiva (à direita).

No aspecto referente às bases de dados, também é importante ressaltar a sua limitação. Uma parcela bastante representativa da base da ANAMACO apresentou falta de informação. Este fator tem impacto direto no retrato da realidade do setor de materiais de construção, podendo trazer resultados diferentes do esperado no modelo, visto que este realiza a previsão somente a partir dos dados disponíveis.

Por fim, a suposição de que o faturamento total de materiais de construção em um distrito tem relação linear com as vendas de interruptores é limitada e, mesmo com uma grande presença desta categoria em todos os tipos de loja, não há assertividade de que todas comercializam estes produtos.

Como oportunidades para próximos estudos na área, sugere-se a busca por variáveis que possam explicar de maneira mais completa as vendas do setor. Também, com a relação negativa entre o faturamento do distrito e seu número de domicílios, surge o questionamento sobre a zona de influência de uma loja do setor de materiais de construção. O quanto o consumidor está disposto a percorrer para realizar a compra de produtos destes segmentos? Seriam estes relacionados a bens de conveniência? Estas são algumas questões que podem ser debatidas e desenvolvidas em próximos trabalhos.

REFERÊNCIAS

- ARAGÃO, Paulo Sergio Sampaio de et al. GeoMarketing: Modelos e sistemas, com aplicações em telefonia. 2005.
- ARANHA, Francisco; FIGOLI, Susana. Geomarketing: memórias de viagem. **São Paulo**, p. 1-73, 2001.
- BIRKIN, Mark; CLARKE, Graham; CLARKE, Martin. Refining and Operationalizing Entropy- Maximizing Models for Business Applications. 商业应用模式下熵最大化模型的应用与改进. **Geographical Analysis**, v. 42, n. 4, p. 422-445, 2010.
- BRASIL. Ministério da Fazenda. Receita Federal. CNAE, Apresentação. 09 Dez, 2014. <Disponível em: <http://idg.receita.fazenda.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/classificacao-nacional-de-atividades-economicas-2013-cnae/apresentacao>. Acesso em: 25 Dez. 2017.
- BUCKLIN, Randolph E.; SIDDARTH, Sivaramakrishnan; SILVA-RISSO, Jorge M. Distribution intensity and new car choice. *Journal of Marketing Research*, v. 45, n. 4, p. 473-486, 2008.
- CHAN, Tat Y.; PADMANABHAN, V.; SEETHARAMAN, P. B. An econometric model of location and pricing in the gasoline market. **Journal of Marketing Research**, v. 44, n. 4, p. 622-635, 2007.
- CHENG, Eddie WL; LI, Heng. Exploring quantitative methods for project location selection. **Building and Environment**, v. 39, n. 12, p. 1467-1476, 2004.
- CHENG, Eddie WL; LI, Heng; YU, Ling. A GIS approach to shopping mall location selection. **Building and Environment**, v. 42, n. 2, p. 884-892, 2007.
- DENNIS, Charles; MARSLAND, David; COCKETT, Tony. Central place practice: shopping centre attractiveness measures, hinterland boundaries and the UK retail hierarchy. **Journal of Retailing and Consumer Services**, v. 9, n. 4, p. 185-199, 2002.
- DOLEGA, Les; PAVLIS, Michalis; SINGLETON, Alex. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. **Journal of Retailing and Consumer Services**, v. 28, p. 78-90, 2016.
- FINN, Adam; LOUVIERE, Jordan J. Shopping center image, consideration, and choice: anchor store contribution. **Journal of business research**, v. 35, n. 3, p. 241-251, 1996.
- HUFF, David L. Parameter estimation in the Huff model. **ESRI, ArcUser**, p. 34-36, 2003.

HUFF, David et al. Calibrating the huff model using ArcGIS business analyst. **ESRI White Paper**, 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Censo Demográfico de 2010: características da população e dos domicílios**. Rio de Janeiro: IBGE, 2011.

PATEL, Ashish; FIK, Timothy J.; THRALL, Grant I. **Trade area definition and calculation**. University of Florida, 2007.

REIMERS, Vaughan; CLULOW, Val. Retail concentration: a comparison of spatial convenience in shopping strips and shopping centres. **Journal of Retailing and Consumer services**, v. 11, n. 4, p. 207-221, 2004.

TELLER, Christoph; REUTTERER, Thomas. The evolving concept of retail attractiveness: what makes retail agglomerations attractive when customers shop at them?. **Journal of Retailing and Consumer Services**, v. 15, n. 3, p. 127-143, 2008.

TITTON, Luiz Antonio; MERLO, Edgard Monforte; GONÇALVES, Marilson Alves; PIRES, Luiz Guilherme Ravacci. Decisão de localização. **PMKT - Revista Brasileira de Pesquisa de Marketing, Opinião e Mídia**, São Paulo, 1 2AD, 2010.

APÊNDICES

Programação utilizada no RStudio

```

#GEORREFERENCIAMENTO DAS LOJAS PELAS COORDENADAS E FUSÃO COM DADOS DO
IBGE DOS SETORES

#Carregando pacotes úteis no geoprocessamento
# Observação: é necessário que os pacotes abaixo já estejam instalados

# rodar as linhas 10 à 13 apenas uma vez para instalar os pacotes

install.packages ('rgdal')
install.packages ('sp')
install.packages ('maptools')
install.packages ('ggmap')

library(rgdal)
library(sp)
library(maptools)
library(ggmap)

# Verificando os limites de memória
memory.limit(10000000)

# Especificar o caminho onde estão os shapefiles e o arquivo de input
con lat long em csv
setwd("C:\\Users\\Desktop\\arquivos para rodar cod distrito por lat
long")

mapa_setores <- readOGR(".", "setores_nse")
# o objeto 'mapa_setores' conterà o conjunto de polígonos de todos os
setores do Brasil
# observação: o objeto 'mapa_setores' deverá ter 316.574 observações

#Verificando variáveis do objeto 'mapa_setores'
names(mapa_setores)
head(mapa_setores@data,3)

# ler arquivo csv com endereços e lat long das lojas
geocoded<-read.csv('Dados_geocod.csv')

#Convertendo os dados de latitude e longitude num objeto adequado
llxy <- data.frame(x=geocoded[,3], y=geocoded[,2])
coordinates(llxy) = ~x+y
llxy@proj4string <- mapa_setores@proj4string
class(llxy)

#encontrando os polígonos do mapa de setores que contêm as latitude e
longitudes
o <- over(llxy, mapa_setores)

# Levando o código de setor censitário e nse_setor no arquivo das
lojas de entrada
geocoded$Cod_setor<-o$CD_GEOCODI

```

```

geocoded$nse_setor<-o$nse_setor

# ler arquivo com setores por distrito

setores<-read.csv('Basico_TotalBrasil_ano2010.csv')
fread('')

# juntando o arquivo de entrada com setores e os distritos

final <- merge(geocoded, distritos, by = c('Cod_setor'))

# Exportando os dados com código de setor e as lojas

write.csv(final, file = "setores.csv",row.names=TRUE)

#UNINDO AS BASES DE LANÇAMENTOS, LOJAS E IBGE POR DISTRITO

setwd("C:/Users/Desktop/cdigosregressaoeboxplot")

library (data.table)
library (ggplot2)
library (ggthemes)
library (dplyr)
library (corrplot)
library (MASS)

#Lendo Base com dados de SP
dfsp<-fread('DFSP.csv')

#Excluindo os CNAEs
dfsp<-dfsp[! ( ( dfsp$CNAE==4741500) | ( dfsp$CNAE==4679601) ) ]
IDH <-fread('IDH.csv')

#agregar variáveis das lojas por distrito
dfdistributo <- dfsp[,list(grande=sum(grande),cadeia=mean(chain_equity),
nlojas=sum(nlojas),somafx=sum(cod_fx),
valor_fx=sum(valor_fx),mediafx=mean(cod_fx),somafunc=sum(Func),mediafunc=mean(Func)),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]

#agregar variáveis censo por distrito
ibge <- fread('Basico_TotalBrasil_ano2010.csv')
ibge <- na.omit(ibge)

ibgesp <- ibge[Cod_municipio == 3550308]

ibgespsoma <-
ibgesp[,list(somamoradores=sum(V002),mediamorad=mean(V002),
mediadomic=mean(V001),somadomicilios=sum(V001),mediamorad.dom=mean(V003),somarendimento=sum(V006),mediarendimento=mean(V005)),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]

#agregar lançamentos residenciais

lanresid <- fread('Lançamentos Residenciais.csv')

```

```

lancredidsp <- lancredid[Cod_municipio == 3550308]
lancredidsp soma <-
lancredidsp[,list(residencial=.N),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]

# agregar lançamentos comerciais

lanccomerc <- fread('Lançamentos Comerciais.csv')
lanccomercsp <- lanccomerc[Cod_municipio == 3550308]
lanccomercsp soma <-
lanccomercsp[,list(comercial=.N),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]
merge <-
lancredidsp soma[lanccomercsp soma,on=c("Cod_distrito","Nome_do_distrito")]
#unindo lançamentos
lançamentos <-
merge[,list(lançamentos=sum(residencial+comercial)),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]

#AGREGAR CNAES

dfcnaes <-
dfsp[,list(V4671100=sum(V4671100),V4672900=sum(V4672900),V4673700=sum(V4673700),V4674500=sum(V4674500),V4679602=sum(V4679602),V4679604=sum(V4679604),V4679699=sum(V4679699),V4742300=sum(V4742300),V4744001=sum(V4744001),V4744002=sum(V4744002),V4744003=sum(V4744003),V4744004=sum(V4744004),V4744005=sum(V4744005),V4744099=sum(V4744099)),by=list(Cod_distrito=Cod_distrito,Nome_do_distrito=Nome_do_distrito)]

# agregando todas as bases
dfmerge <- dfcnaes[dfdistrito,
on=c("Cod_distrito","Nome_do_distrito")]
dfmerge2 <-
ibgespsoma[dfmerge,on=c("Cod_distrito","Nome_do_distrito")]
dfmerge3 <-
lançamentos[dfmerge2,on=c("Cod_distrito","Nome_do_distrito")]
dfmerge4 <- IDH[dfmerge3,on=c("Cod_distrito","Nome_do_distrito")]
dfmerge5 <-
energiadistrito[dfmerge4,on=c("Cod_distrito","Nome_do_distrito")]
dfmerge6 <-
dfmissing2[dfmerge5,on=c("Cod_distrito","Nome_do_distrito")]
dfmerge6[is.na(dfmerge6)] <- 0

write.csv(dfmerge6,"DISTRITOSMEDIA.CSV")

dffinal <-fread("DISTRITOSMEDIA.CSV")

####regressão múltipla final

reg.final<- lm(valor_fx ~
grande+lançamentos+mediarendimento+somafunc+nlojas+somadomicilios,
data = dffinal)

summary(reg.final)

#stepwise

```

```
regAIC <- stepAIC(reg.final)

#modelo final
reg.final2<- lm(valor_fx ~
lancamentos+grande+somafunc+nlojas+somadomicilios, data = dffinal)
predict(reg.final2)

dffinal$modelo <- predict(reg.final2)

# exportando previsão do modelo
write.csv(dffinal, "previsao.csv")

summary(reg.final2)

modelo <- predict(reg.final2)
dffinal$modelo
```