



UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE EXATAS E ENGENHARIA
CURSO DE ENGENHARIA CIVIL

EDUARDO VICENSI DE BASTIANI

DESENVOLVIMENTO DE MODELO PREDITIVO PARA A
AVALIAÇÃO IMOBILIÁRIA E ESTUDO DE CASO
EM CAXIAS DO SUL – RS

CAXIAS DO SUL

2018

UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE EXATAS E ENGENHARIA
CURSO DE ENGENHARIA CIVIL

EDUARDO VICENSI DE BASTIANI

**DESENVOLVIMENTO DE MODELO PREDITIVO PARA A
AVALIAÇÃO IMOBILIÁRIA E ESTUDO DE CASO
EM CAXIAS DO SUL – RS**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia Civil da Universidade de Caxias do Sul, como requisito parcial para obtenção do título de Engenheiro Civil.

Orientadora: Prof. Karen Fiuza, PhD.

CAXIAS DO SUL

2018

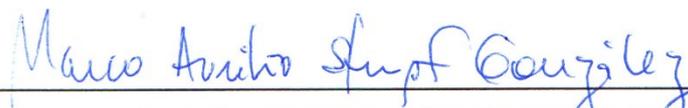
EDUARDO VICENSI DE BASTIANI

**DESENVOLVIMENTO DE MODELO PREDITIVO PARA A
AVALIAÇÃO IMOBILIÁRIA E ESTUDO DE CASO EM
CAXIAS DO SUL – RS**

Trabalho de Conclusão de Curso
apresentado ao curso de Engenharia
Civil da Universidade de Caxias do Sul,
como requisito parcial para obtenção
do título de Engenheiro Civil.

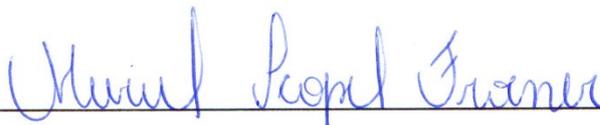
Aprovado em 19/06/2018

BANCA EXAMINADORA



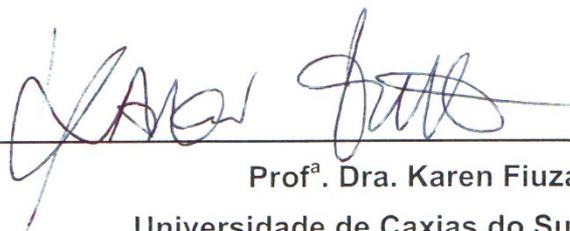
Prof. Dr. Marco Aurélio Stumpf González

Universidade do Vale do Rio dos Sinos - UNISINOS



Prof^a. Ms. Muriel Scopel Froener

Universidade de Caxias do Sul



Prof^a. Dra. Karen Fiuza

Universidade de Caxias do Sul

Dedico este trabalho a meus pais,
pelo amor, apoio e incentivo
em todos os momentos.

AGRADECIMENTOS

Sou muito feliz por compartilhar esta jornada com tantas pessoas fantásticas – cada uma evoca imagens, momentos e histórias tão marcantes na completude deste trabalho. Em um autêntico gesto de amor, mais do que uma mera ação formal, gostaria de agradecer:

Aos meus pais, Luís Carlos e Lourdes Marili, e meu irmão, Gustavo, que sempre quiseram o meu melhor. Esta conquista é pura consequência de todo o amor e educação que sempre me deram, do irrestrito apoio e suporte; do incentivo para estudar, buscar além e ser eternamente curioso. Igualmente à toda minha família, com quem sempre pude contar.

À Prof^ª. Dra. Karen Fiuza, que além da orientação e de compartilhar conhecimentos tão essenciais para a elaboração deste trabalho, gentilmente acreditou na ideia desde o início, ofereceu motivação, oportunidades e inspiração em todos os momentos.

Aos excelentes professores que contribuíram para minha formação em tantos aspectos. Aos professores que acrescentaram ao trabalho – Profs. Adriano Costa, Givanildo Garlet, Matheus Nogueira e Vinício Ceconello. Aos professores que participaram da avaliação – Prof^ª. Ms. Muriel Scopel Froener, que contribuiu desde a pré-banca, e Prof. Dr. Marco Aurélio Stumpf González, que ofereceu sua extensa experiência em engenharia de avaliações (e como presente, seu livro).

Ao Ms. Gustavo Sperotto, representando a GIHAB/CAIXA. Além dos dados de mercado que viabilizaram o modelo, ofereceu enorme receptividade, ensinamentos e suporte ao trabalho.

Às pessoas iluminadas que conheci na University of Illinois at Urbana-Champaign, com quem vivi momentos únicos e descobri que o mundo é grande e cheio de possibilidades. Inspiraram este trabalho: Isaac Howenstine – por acreditar nas ideias entusiasticamente; Pedro Tremacoldi Rossi – por mostrar o quão interessante é a economia; Jaime Gaya Fuertes, Felipe Avileis, Michael Parisotto – pelas incessantes conversas, companheirismo, excelência.

Às pessoas da Biossplena: Dani Caon, Giovana Ulian, Maurício D’Agostini, Maria Mezzomo, Miguel Angel – por acreditarem no meu trabalho, pelos exemplos de competência profissional, conhecimentos compartilhados, ideias inspiradoras, oportunidades oferecidas.

Aos amigos que sempre estiveram juntos (Entony, Gabriel, Matheus, Rodolfo, Túlio). Aos amigos da Universidade de Caxias do Sul e Engenharia Civil por compartilharem tantos desafios e momentos felizes, fazendo a experiência na universidade ainda mais intensa e gratificante. A todos que cruzaram o caminho comigo e fizeram a diferença em minha vida. Aos que pensam diferente, aos que querem um mundo melhor, aos que tornam isso real. *The world is our oyster.*

SUMÁRIO

RESUMO.....	8
ABSTRACT	9
LISTA DE FIGURAS.....	10
LISTA DE TABELAS.....	12
CAPÍTULO 1 – INTRODUÇÃO	14
APRESENTAÇÃO.....	14
1.1 HIPÓTESE DE PESQUISA	18
1.2 ESTRUTURA DE PESQUISA	18
1.2.1 OBJETIVO GERAL.....	18
1.2.3 LIMITAÇÕES DE PESQUISA.....	19
1.3 ORGANIZAÇÃO DO TRABALHO	19
CAPÍTULO 2 – MERCADO IMOBILIÁRIO.....	20
INTRODUÇÃO.....	20
2.1 AVALIAÇÃO DE BENS.....	21
2.2 MODELOS DE PREÇOS HEDÔNICOS	23
2.3 MODELOS DE APOIO	27
2.3.1 MODELOS DE REGRESSÃO ESPACIAL	27
2.3.2 MODELOS COMPOSTOS.....	30
CAPÍTULO 3 – CIÊNCIA DE DADOS	32
INTRODUÇÃO.....	32
3.1 ETAPAS DE CIÊNCIA DE DADOS	33
3.1.1 DEFINIR O PROBLEMA.....	34
3.1.2 ORGANIZAR OS DADOS.....	34
3.1.3 ANÁLISE EXPLORATÓRIA	35
3.1.4 APRESENTAÇÃO DE RESULTADOS	35
3.1.5 TOMAR DECISÕES.....	35
3.2 REGRESSÃO LINEAR MÚLTIPLA	36
3.2.1 COEFICIENTE MÚLTIPLO DE DETERMINAÇÃO - R^2	38
3.2.2 COEFICIENTE DE SIGNIFICÂNCIA - TESTE F	38
3.2.3 COEFICIENTE DE SIGNIFICÂNCIA - TESTE T.....	39
3.2.4 COEFICIENTE DE CORRELAÇÃO - ρ	40
3.2.5 CONSIDERAÇÕES	40

3.3 TÉCNICAS DE APOIO	41
3.3.1 REGRESSÃO ESPACIAL.....	41
3.3.2 ANÁLISE DE CLUSTERS.....	41
3.3.3 NETWORK SCIENCE.....	43
CAPÍTULO 4 – DESENVOLVIMENTO DE MODELO PREDITIVO.....	44
INTRODUÇÃO.....	44
4.1 COLETA E ORGANIZAÇÃO DE DADOS.....	46
4.1.1 OFERTA IMOBILIÁRIA – CAIXA ECONÔMICA FEDERAL	47
4.1.2 INFORMAÇÕES SOCIOECONÔMICAS – CENSO DO IBGE.....	49
4.1.3 PONTOS DE INTERESSE – GOOGLE MAPS API	50
4.2 ANÁLISE EXPLORATÓRIA	51
5.2.1 ANÁLISE DE PREÇOS POR METRO QUADRADO	51
4.2.2 SELEÇÃO DE VARIÁVEIS	53
4.2.3 CORRELAÇÃO ENTRE VARIÁVEIS.....	55
4.2.4 GRÁFICOS DE DISPERSÃO	57
4.3 MODELO PREDITIVO	59
4.3.1 REGRESSÃO LINEAR MÚLTIPLA SIMPLES.....	59
4.3.2 REGRESSÃO LINEAR MÚLTIPLA COM VARIÁVEIS ESPACIAIS	62
4.3.3 ANÁLISE DOS COEFICIENTES	64
CAPÍTULO 5 – ESTUDO DE CASO	67
INTRODUÇÃO.....	67
5.1 INFLUÊNCIA DA LOCALIZAÇÃO	68
5.2 SIMULAÇÃO DE VALORES PARA UNIDADES PADRÃO	72
5.3 ANÁLISE DE VALORES E URBANISMO	74
5.3.1 COMPLEXIDADE URBANA.....	75
5.3.2 SEGREGAÇÃO ESPACIAL	76
5.3.3 CORRELAÇÃO ENTRE VALORES E URBANISMO	77
CONCLUSÃO.....	79
SUGESTÕES PARA FUTUROS TRABALHOS.....	80
REFERÊNCIAS BIBLIOGRÁFICAS	81
APÊNDICE A – INFLUÊNCIA DA LOCALIZAÇÃO	86
APÊNDICE B – COMPLEXIDADE URBANA	87
APÊNDICE C – SEGREGAÇÃO ESPACIAL.....	88

RESUMO

DE BASTIANI, Eduardo Vicensi. “**Desenvolvimento de modelo preditivo para a avaliação imobiliária e estudo de caso em Caxias do Sul – RS**”. *Monografia*. Área do Conhecimento de Ciências Exatas e Engenharias, Universidade de Caxias do Sul, Caxias do Sul – RS, 2018.

Conhecer o valor de mercado de imóveis é essencial na tomada de decisões por construtores, engenheiros, consumidores e múltiplos agentes associados ao mercado imobiliário. Diante das dificuldades e da necessidade de encontrar o valor de maneira rápida e assertiva, este trabalho propõe um modelo preditivo para estimar valores de mercado de imóveis. Exploram-se por meio de uma pesquisa bibliográfica possibilidades e conexões entre o mercado imobiliário (engenharia de avaliações, preços hedônicos, economia urbana) e ciência de dados (conceitos, regressão linear múltipla, técnicas espaciais complementares). O método consiste em três etapas: organização da base de dados, desenvolvimento do modelo preditivo e estudo de caso em Caxias do Sul - RS. Os dados utilizados nesta investigação incluem transações da Caixa Econômica Federal, censo demográfico do IBGE e pontos de interesse do Google Maps API. O conjunto abrange valores de apartamentos residenciais em uma área de 47 bairros da cidade. As análises exploratórias incluem valores médios por m², caracterização de variáveis, análises de correlação e gráficos de dispersão. O modelo preditivo é elaborado utilizando a técnica da regressão linear múltipla para determinar os preços hedônicos, obtendo um coeficiente $R^2 = 0,82$. Verifica-se a presença de autocorrelação espacial nos resíduos da regressão. Incluem-se proxies espaciais que diferenciam localizações geográficas, obtendo um coeficiente $R^2 = 0,87$. O estudo de caso na cidade de Caxias do Sul – RS inclui três componentes. A influência da localização é um mapa de calor mostrando a valorização imobiliária no território. A simulação de valores demonstra o funcionamento do modelo preditivo. A análise de aspectos urbanísticos verifica os indicadores de complexidade urbana e segregação espacial e seus impactos sobre a valorização imobiliária do território. Conclui-se que os modelos preditivos associados ao mercado imobiliário proporcionam novas perspectivas a problemas tradicionais, oferecendo inúmeras respostas aos agentes interessados em imóveis.

Palavras-chave: avaliações imobiliárias; ciência de dados, economia urbana.

ABSTRACT

DE BASTIANI, Eduardo Vicensi. “**Development of a predictive model for real estate appraisals – Case study in Caxias do Sul (Brazil)**”. *Senior thesis*. College of Exact Sciences and Engineering, Universidade de Caxias do Sul, Caxias do Sul – Brazil, 2018.

Being aware of the market value of real estate assets is essential in decision-making by builders, engineers, consumers and several agents associated with real estate markets. Facing the difficulties as well as the need to assess values quickly and assertively, this study aims to propose a predictive model to estimate real estate prices. The study explores possibilities, through a bibliographical research, and existing connections among real estate markets (valuation engineering, hedonic prices, urban economics) and data science (concepts, multiple linear regression, and further spatial techniques). The method has three steps: database organization, development of the predictive model and a case study in Caxias do Sul - Brazil. The data sources used in this investigation include transactions of the bank Caixa Econômica Federal, the Brazilian demographic census and points of interest from Google Maps API. The dataset covers residential apartments in an area of 47 city neighborhoods. The exploratory analysis includes average values per square meter, characterization of variables, evaluation of correlations and scatter plots. The predictive model is employed using the multiple linear regression technique to resolve the hedonic prices, obtaining the coefficient $R^2 = 0.82$. The presence of spatial autocorrelation in the regression residuals is verified. Spatial proxies are included to discern geographic locations, achieving the coefficient $R^2 = 0.87$. The case study in the city of Caxias do Sul - Brazil includes three major components. Influence of location is a heat map representing real estate appreciation on territory. The simulation of values demonstrates the operation of the predictive model. The analysis of urban factors verifies the urban complexity and spatial segregation indicators and their impacts on the spatial real estate appreciation. The main takeaway of this research is that predictive models combined to real estate markets provide new perspectives to solve traditional problems, offering plentiful answers to real estate agents.

Keywords: real estate appraisals; data science; urban economics.

LISTA DE FIGURAS

Figura 1 - Complexidade do valor de imóveis	14
Figura 2 - Preço hedônico como uma composição de características do bem	25
Figura 3 - Distância de imóvel até um ponto fixo	27
Figura 4 - Apresentação de resíduos contendo dependência espacial	28
Figura 5 - Imóveis avaliados pela Zillow em uma área de Champaign, Illinois	30
Figura 6 - Ciência de dados pelo diagrama Venn de Drew Conway's	32
Figura 7 - Processo tradicional de ciência de dados	33
Figura 8 - Regressão linear múltipla.....	37
Figura 9 - Distância euclidiana.....	41
Figura 10 - Classificação por clusters a partir do algoritmo <i>k-means</i>	42
Figura 11 - Representação de uma rede com elementos conectados.....	43
Figura 12 - Desenvolvimento de pesquisa.....	45
Figura 13 - Procedimentos de pesquisa	45
Figura 14 - Distribuição espacial dos dados de mercado - imóveis	48
Figura 15 - Setores censitários em Caxias do Sul, RS.....	49
Figura 16 - Distribuição espacial dos pontos de interesse.....	50
Figura 17 - Valor médio por m ² – por bairro – na amostra de dados de mercado.....	52
Figura 18 - Correlações entre variáveis	56
Figura 19 - Gráfico de dispersão entre valor do imóvel e área privativa	57
Figura 20 - Gráficos de dispersão de atributos e valor de mercado	58
Figura 21 - Construção do modelo de regressão linear múltipla.....	59
Figura 22 - Distribuição espacial dos resíduos da regressão	60
Figura 23 - Resultados de Moran I.....	61
Figura 24 - Comparação entre valores observados e valores estimados	63
Figura 25 - Mapa dos valores de mercado da unidade por localização	69
Figura 26 - Distribuição de valores de mercado para imóveis por localização.....	70
Figura 27 - Distribuição de renda da população brasileira x Distribuição de valores.....	71
Figura 28 - Pontos selecionados para simulação do valor de mercado	72
Figura 29 - Simulação de valores pra unidades e localizações.....	73
Figura 30 - Indicadores territoriais selecionados.....	74
Figura 31 - Indicador de complexidade urbana	75
Figura 32 - Indicador de segregação espacial.....	76
Figura 33 - Resultado da correlação entre indicadores e valor de mercado	77

LISTA DE QUADROS

Quadro 1 – Benefícios do modelo preditivo para o mercado imobiliário	16
Quadro 2 – Metodologia de pesquisa	44
Quadro 3 – Definição dos dados empregados	46
Quadro 4 – Caracterização de variáveis selecionadas	54
Quadro 5 – Componentes para o estudo de caso	68

LISTA DE TABELAS

Tabela 1 – Validação de dados do trabalho em relação à publicação ZAP-Viva Real	53
Tabela 2 – Coeficientes do modelo	65
Tabela 3 – Variável defasada espacialmente - Bairro	66

As pessoas veem as coisas como elas são e perguntam *por quê?*
Eu vejo as coisas como elas poderiam ser e pergunto *por que não?*

George Bernard Shaw

CAPÍTULO 1 – INTRODUÇÃO

A verdadeira viagem do descobrimento não consiste em procurar novas paisagens, e sim em ter novos olhos.

Marcel Proust

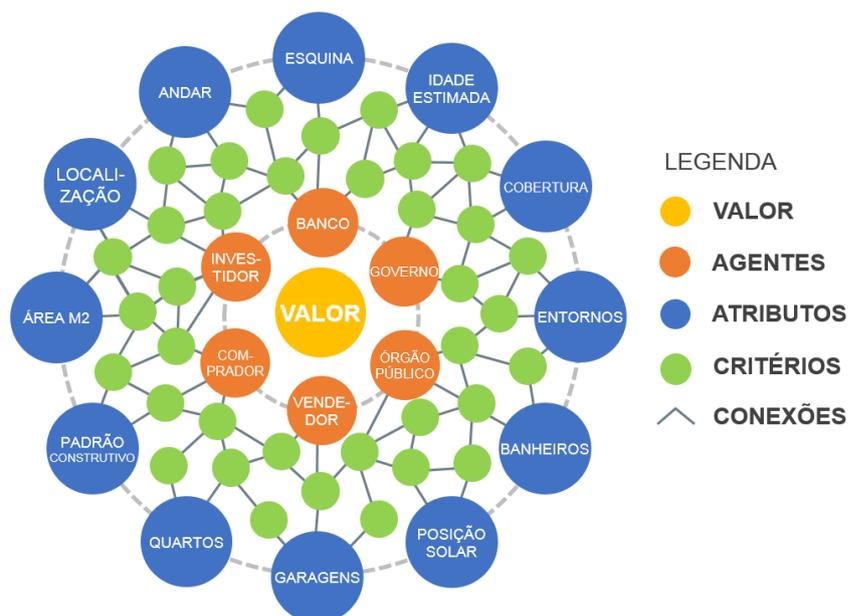
APRESENTAÇÃO

Qual é o valor de mercado de um determinado imóvel?

Esta questão simples (porém, nunca simplória) é um norte para grandes decisões. Para o construtor, indica a viabilidade de um empreendimento. Para o engenheiro civil, é critério indispensável em um projeto. Para o consumidor, influencia na escolha de onde e como viver, e sobre a aquisição, usualmente, do maior patrimônio e investimento de sua vida.

O valor de mercado surge de interações entre os indivíduos. A *mão invisível* do mercado, conceituada por Adam Smith em *The Wealth of Nations* (1776), ajusta a oferta e demanda dos produtos. Desta forma, o valor de mercado é determinado unicamente pelo preço que compradores e vendedores negociam e consideram justo para ambas as partes. Conhecer os preços é imprescindível para decidir, pois “toda escolha que fazemos é moldada pelos preços das opções que se apresentam diante de nós [...] o que calculamos como sendo seus custos relativos – pesadas em relação aos seus benefícios” (PORTER, 2011).

Figura 1 - Complexidade do valor de imóveis



Fonte: O Autor (2018), adaptado de conceito de complexidade (BIOSSPLENA, 2017).

Determinar valores é complexo, como demonstra a Figura 1 – cada agente percebe o valor distintamente. Atributos e critérios como características físicas, localização, hábitos de consumo, momentos econômicos e outras variáveis (situação familiar, expectativas culturais, laços emocionais) impactam os agentes e suas percepções no valor de imóveis (ZILLOW, 2016).

Em busca de determinismo neste mundo estocástico, as avaliações imobiliárias determinam por meio da inferência estatística o *valor mais próximo ao de mercado* a um bem. Ainda assim, as avaliações possuem inconvenientes, podendo ser subjetivas¹ e possuir um elevado custo², inviabilizando diversas aplicações – o que indica em muitos casos a necessidade de *complementos*.

Quando um construtor, um engenheiro, um consumidor, ou qualquer parte interessada no imóvel precisa de uma estimativa (para múltiplas finalidades) rápida, assertiva e acessível sobre o valor do mesmo, existem poucas alternativas.

O mais lógico e intuitivo é prospectar unidades *semelhantes*. Por exemplo, uma pessoa deseja estimar o valor de um determinado apartamento (três quartos, 120m² de área construída, padrão médio, no centro da cidade). O procedimento clássico é buscar o preço de apartamentos semelhantes. Mas e se o primeiro possuir um quarto a menos? E se o segundo possuir menor área, porém melhor padrão construtivo? E se o terceiro, que possui exatamente a mesma configuração, estiver localizado na esquina de um bairro tranquilo e afastado do centro?

O cérebro é excelente em interpretar relações, porém possui grande desvantagem para calcular numericamente dados extensos e complexos (KAHNEMAN, 2012). Embora possível imaginar um valor adequado para este imóvel, estas conclusões seriam subjetivas, pouco assertivas (precisão e margem desconhecidas) e provavelmente distantes da melhor alternativa matemática. Em muitas vezes, são estas as conclusões que embasam decisões significativas.

¹ Marco Aurélio González menciona que as avaliações muitas vezes são realizadas de forma pontual, usando heurísticas, sem o apoio anterior e consolidação de modelos posteriores. Não é comum a análise quantitativa do mercado, avaliando tendências de forma consistente, mas apenas a análise qualitativa, utilizando muitas vezes amostras pequenas (menores que 50 dados, em alguns casos apenas três) para *modelar* o mercado. Questões como a escolha do modelo, seleção de casos e variáveis também são resolvidas subjetivamente (GONZÁLEZ, 2002).

² A avaliação imobiliária demanda o trabalho de um profissional especializado. Atualmente, a Caixa Econômica Federal cobra uma tarifa de R\$2.200,00 por imóvel avaliado (FOLHA, 2015).

Modelo preditivo

Um modelo preditivo é aquele que utiliza estatística para prever resultados. Ou seja, por meio de *dados informativos* (por exemplo, localização e características do imóvel), estima um valor desconhecido de interesse: o *alvo* (neste caso, o preço do imóvel) (SIEGEL, 2013).

O maior benefício de um modelo preditivo é apoiar a tomada de decisões, reduzindo incertezas com informações mais assertivas, baseadas em dados, não apenas na intuição (SIEGEL, 2013).

Quanto mais orientado por dados, mais produtiva é uma empresa. O Massachusetts Institute of Technology (MIT) e a Penn's Wharton School desenvolveram um índice de decisão orientadas por dados nas empresas. Mesmo controlando uma vasta gama de fatores de confusão, um desvio padrão acima na escala de decisões orientadas por dados indica um aumento de 4%-6% na produtividade da empresa, bem como aumentos no retorno sobre ativos, retorno sobre o patrimônio líquido, utilização de ativos e valor de mercado (PROVOST & FAWCETT, 2013).

A estimativa eficiente dos valores de imóveis pode auxiliar diferentes agentes a tomar decisões: construtores, engenheiros, consumidores e demais partes interessadas (investidores, incorporadoras, imobiliárias, corretores, instituições bancárias, governos, órgãos públicos...). O Quadro 1 exemplifica benefícios derivados desta estimativa:

Quadro 1 – Benefícios do modelo preditivo para o mercado imobiliário

Agente	Benefício
Engenheiros, construtores, incorporadoras.	Otimizar a construção de empreendimentos, em termos de localização geográfica e características ótimas, comparando o orçamento (terreno, insumos, mão-de-obra...) com o valor de mercado potencial da configuração, como indicador de rentabilidade.
Público em geral.	Estimar o valor de mercado para imóveis, possibilitando prospecções simples ou melhor conhecimento do mercado.
Engenheiros, imobiliárias, corretores.	Agilizar avaliações imobiliárias, obtendo imóveis adequados para os cálculos, de acordo com os parâmetros especificados em norma.
Investidores.	Conhecer unidades mais valorizadas pelo mercado, para construir estratégias de investimentos que maximizem a rentabilidade.

(continua)

(conclusão)

Agente	Benefício
Instituições bancárias.	Elaborar um índice de valorização imobiliária análogo ao CUB, porém em vez de avaliar custos, indica potencial de mercado para diferentes padrões em múltiplas esferas (municipal, estadual...). Desenvolver instrumentos financeiros, como títulos e derivativos, atrelados a um índice de preço de imóveis gerais ou específicos.
Governos, órgãos públicos.	<i>Secretarias da Fazenda</i> - Aumentar a eficiência e justiça tributária na cobrança de impostos como o IPTU (sobre o valor do imóvel) ou ITBI (sobre transações imobiliárias). <i>Secretarias de Planejamento</i> – Orientar políticas públicas e estudos de impacto de vizinhança sobre indicadores de desenvolvimento, agregando a valorização estimada de imóveis no território; <i>Receita Federal</i> - Detectar indícios e investigar atividades ilícitas no setor da construção, como sonegação ou lavagem de dinheiro.

Fonte: Elaboração do Autor (2017)

A empresa americana Zillow desenvolveu em 2006 um modelo preditivo: a estimativa Zestimate avalia diariamente 110 milhões de imóveis nos EUA, utilizando dados públicos e ofertas de mercado. O site permite consultar uma estimativa de preço com elevada precisão³ para qualquer imóvel sobre o mapa, além de características, oscilações históricas, ofertas semelhantes, estimativas de aluguel e financiamentos. Cerca de 50 milhões de visitantes acessam estes serviços mensalmente, de forma gratuita, e a renda da Zillow provém de publicidade e serviços pagos. (ZILLOW, 2016).

O Brasil ainda não possui ferramentas semelhantes. Os desafios incluem a elaboração de modelos matemáticos e estatísticos que capturem as complexidades e nuances do mercado imobiliário; bem como o acesso aos dados⁴ (públicos e/ou privados) que permitem o treinamento e aperfeiçoamento destes modelos. Estes desafios e benefícios são ampliados para a escala nacional em uma ferramenta como o Zestimate.

³ Possui uma precisão de $\pm 5\%$ na avaliação de 50% dos imóveis a nível nacional (ZILLOW, 2016).

⁴ O acesso aos dados é um esforço mais político que técnico. Prefeituras e secretarias, órgãos públicos e instituições bancárias possuem múltiplos dados de imóveis; porém muitas vezes há preocupações em compartilhar dados devido aos potenciais ganhos estratégicos; a obtenção usualmente envolve negociações e/ou acordos de uso.

A motivação para este tema de pesquisa surgiu do desejo de desenvolver um trabalho multidisciplinar entre a engenharia civil, a matemática e a economia, possibilitando expandir e conectar conhecimentos nestas áreas. A temática da ciência de dados é interessante para resolver um amplo espectro de problemas, com um futuro promissor⁵. Por fim, a modelagem de valores imobiliários é um assunto contemporâneo, e um estudo de caso em Caxias do Sul (RS) é inédito, proporcionando desafios e oportunidades únicas de aprendizagem.

1.1 HIPÓTESE DE PESQUISA

Diante da necessidade de estimar o valor de imóveis de maneira rápida, assertiva e acessível no setor da construção civil (construtores, engenheiros, consumidores...) e observando os fundamentos econômicos, as avaliações imobiliárias, os métodos estatísticos clássicos e técnicas modernas da ciência de dados, a hipótese de pesquisa constitui que:

“É possível construir um modelo preditivo para estimar o valor de imóveis; bem como empregar este modelo para estudar um mercado imobiliário em uma cidade na região da Serra Gaúcha – RS – buscando informações estratégicas para auxiliar na tomada de decisões”.

1.2 ESTRUTURA DE PESQUISA

1.2.1 OBJETIVO GERAL

O objetivo desta investigação é elaborar um modelo preditivo para estimar o valor de mercado de imóveis; e realizar um estudo de caso em um mercado imobiliário da região.

1.2.2 OBJETIVOS ESPECÍFICOS

1. Investigar modelos de avaliações de imóveis e seus métodos estatísticos.
2. Desenvolver um modelo preditivo para estimar o valor de mercado dos imóveis.
3. Elaborar um estudo de caso apresentando resultados do modelo preditivo, por meio de componentes que permitam alavancar o conhecimento e adquirir informações estratégicas sobre um mercado imobiliário regional.

⁵ Diferentes profissionais (engenheiros, economistas, matemáticos, estatísticos, físicos, programadores, cientistas sociais, médicos...) executam a ciência de dados em diferentes indústrias. Devido ao crescimento acelerado, aos enormes benefícios para diferentes indústrias (sobretudo empresas globais), à multidisciplinaridade envolvida e à dificuldade de encontrar profissionais, a revista Harvard Business Review concedeu ao *data scientist* (cientista de dados) o título de *Sexiest Job of the 21st Century* (HBR, 2012). O portal Glassdoor avaliou número de vagas abertas, salário e oportunidades de carreira e classificou o cientista de dados em 1º lugar para melhores profissões de 2016 (GLASSDOOR, 2016).

1.2.3 LIMITAÇÕES DE PESQUISA

- O método utilizado será a regressão linear múltipla, agregando fundamentos básicos de outros métodos, como os proxies espaciais (regressão espacial);
- O estudo será realizado em uma região da cidade de Caxias do Sul – RS;
- Os imóveis avaliados são exclusivamente da categoria “apartamentos residenciais”.

1.3 ORGANIZAÇÃO DO TRABALHO

Além desta introdução, o trabalho divide-se em seis capítulos. O capítulo 2, denominado “Mercado imobiliário”, aborda aspectos fundamentalistas de imóveis: engenharia de avaliações e norma, economia e econometria, preços hedônicos e aplicações no mercado imobiliário de cidades brasileiras, análise de técnicas de regressão espacial e modelos compostos.

O capítulo 3 conceitua a “Ciência de dados” – apresenta o processo tradicional para a transformação dos dados em informações estratégicas e inclui técnicas estatísticas para análise de dados: a regressão linear múltipla (de forma mais específica) e uma introdução às técnicas de regressão espacial, análise de *clusters* e *networks*.

O capítulo 4 apresenta o “Desenvolvimento do modelo preditivo”. Primeiramente, realiza-se a coleta e organização de dados relativos ao mercado imobiliário. Em seguida, análises exploratórias para conhecer e interpretar os dados. Feito isso, elabora-se o modelo preditivo para estimar o valor de mercado dos imóveis, interpretando seus resultados.

O capítulo 5 demonstra o “Estudo de caso”, visando demonstrar a aplicação do modelo em situações reais dentro da cidade de Caxias do Sul – RS, por meio da análise de três componentes distintos: influência da localização; simulação de valores para unidades; análise de valores e urbanismo. Ao final, incluem-se conclusão, referências e apêndices.

CAPÍTULO 2 – MERCADO IMOBILIÁRIO

O que nos causa problemas não é o que *não* sabemos.
É aquilo que sabemos com toda a certeza.

Mark Twain (1835-1910),

citado em *The Big Short*, referindo-se
à crise imobiliária de 2008.

INTRODUÇÃO

Este capítulo introduz conceitos econômicos e do mercado imobiliário (como preço, valor, mercado, preços hedônicos, regressão, engenharia de avaliações, modelos compostos...), que nortearão as proposições para elaborar um modelo preditivo que estime o valor de imóveis.

O *preço* é a quantia “pela qual se *efetua*, ou se *propõe* efetuar, uma transação”; enquanto o *valor* de mercado é a quantia “mais *provável* pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente” (NBR 14653-1:2001, *itálicos adicionados*)⁶. Um aforismo popular do investidor bilionário Warren Buffet define pragmaticamente: “Preço é o que você paga, valor é o que você recebe”.

Sobre o valor de *mercado* surge a avaliação de bens, sendo a “análise técnica, realizada por um engenheiro de avaliações, para identificar o valor de um bem [...] a uma determinada finalidade, situação e data” (ABNT, 2001). O mercado, onde todas as transações acontecem, é composto por três elementos - os bens levados ao mercado, os vendedores e os compradores (DANTAS, 1999).

Engenharia de Avaliações

A engenharia de avaliações é uma especialidade multidisciplinar⁷ da engenharia, abrangendo ciências sociais, exatas e da natureza, cujo objetivo é “determinar tecnicamente o valor de um bem, seus direitos, frutos e custos de produção” (DANTAS, 1999).

⁶ Estas definições sintetizam uma série de questões econômicas, visando a concisão e o foco da investigação. Embora não seja o cerne dessa discussão, as definições econômicas podem ser afetadas por questões de mercado, como a concorrência perfeita e imperfeita, monopólios e oligopólios, liquidez, homogeneidade dos bens, racionalidade e conhecimentos dos participantes, por exemplo.

⁷ A engenharia de avaliações é uma área considerada complexa, pois integra diversas áreas do conhecimento tais como: arquitetura, psicologia, análise de investimentos, análise de balanços, estatística básica, estatística inferencial, amostragem, matemática aplicada, matemática financeira, micro e macroeconomia, engenharia econômica, economia urbana, planejamento urbano, filosofia, sociologia urbana, econometria, teoria das

Os primeiros trabalhos publicados em engenharia de avaliações no Brasil datam de 1918 e 1919. Em 1941, Luiz Carlos Berrini publicou o livro “Avaliações de Terreno” e em seguida “Avaliações de Imóveis”, que difundiram a área no Brasil e na América Latina. Em 1952, o departamento de Engenharia Civil da Caixa Econômica Federal publicou a primeira norma para avaliação de imóveis. A área ganhou popularidade a partir da década de 70.

Em 1997, Marco Aurélio Stump González publicou o livro “A Engenharia de Avaliações na Visão Inferencial” (GONZÁLEZ, 1997), unificando e divulgando métodos e técnicas estatísticas da engenharia de avaliações. Em 1999, Rubem Alves Dantas publicou o livro “Engenharia de Avaliações” (DANTAS, 1999), em conjunto com a Caixa Econômica Federal.

O exercício profissional da Engenharia de Avaliações é regulamentado desde 24 de dezembro de 1966, pela Lei nº 5.194 em seu Artigo 7º, parágrafo c: “As atividades e atribuições profissionais do engenheiro, do arquiteto e do engenheiro-agrônomo consistem em: c) estudos, projetos, análises, *avaliações*⁸, *vistorias*, *perícias*, pareceres e divulgações técnicas” (BRASIL, 1966, *itálicos adicionados*).

2.1 AVALIAÇÃO DE BENS

Em 2001, a Associação Brasileira de Normas Técnicas (ABNT) publicou a norma “NBR 14653 - Avaliação de Bens”, consolidando conceitos, métodos e procedimentos gerais para a avaliação de bens. A NBR 14653-1:2001 é a norma diretriz para outras seis normas associadas que compreendem avaliações de imóveis urbanos, imóveis rurais, empreendimentos, máquinas e equipamentos, recursos naturais e patrimônios históricos. À medida em que essas seis normas são produzidas, as normas individuais antigas são canceladas e substituídas⁹.

A metodologia de uma avaliação, segundo a norma diretriz NBR 14653-1:2001, consiste em seis etapas: “1. Conhecimento do objeto de pesquisa – o bem avaliado; 2. Preparação de pesquisa - planejamento; 3. Trabalho de campo (coleta de dados); 4.

probabilidades, teoria das decisões, pesquisa científica, pesquisa operacional, álgebra linear, direito imobiliário, marketing, mercado de capitais etc. (DANTAS, 1999).

⁸ Essa lei garante a estes profissionais a exclusividade de avaliar um imóvel. Entretanto, segundo a lei que regulamenta a profissão de corretor de imóveis, Lei Federal nº 6530 (BRASIL, 1978), Art. 3º: “Compete ao Corretor de Imóveis exercer a intermediação na compra, venda, permuta e locação de Imóveis e *opinar* quanto à comercialização imobiliária”. Marisa Baptistella lembra que o valor de imóvel opinado por um corretor de imóveis, apesar de muitas vezes chamado de “avaliação”, não deve ser considerada como tal (BAPTISTELLA, 2005).

⁹ Por exemplo, a publicação da NBR 14653-2:2011 – Avaliação de Bens – Parte 2: Imóveis Urbanos substituiu a norma anterior NBR 5676:1990 – Avaliação de Imóveis Urbanos.

Processamento e análise de dados – análise exploratória; 5. Interpretação e explicação dos resultados – o modelo; 6. Redação do relatório da pesquisa – laudo de avaliação”.

Ainda de acordo com a NBR 14653-1:2001 (ABNT, 2001), os procedimentos avaliatórios usuais com a finalidade de identificar o valor de um bem, seus frutos e direitos são:

Método comparativo direto de dados de mercado: identifica o valor de mercado do bem, por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra.

Método involutivo: identifica o valor de mercado do bem, alicerçado no seu aproveitamento eficiente, baseado em modelo de estudo de viabilidade técnico econômica, mediante hipotético empreendimento compatível com as características do bem e com as condições de mercado com o qual está inserido, considerando-se cenários viáveis para execução e comercialização do produto.

Método evolutivo: identifica o valor do bem, pelo somatório dos valores de seus componentes. Para o valor de mercado, deve ser considerado o fator de comercialização.

Método da capitalização de renda: identifica o valor do bem, com base na capitalização presente da sua renda líquida prevista, considerando-se cenários viáveis. (NBR 14653-1:2001)

Por exemplo, deseja-se avaliar o valor de uma casa. O avaliador utiliza o método comparativo direto de dados de mercado avalia esta casa comparando-a com outras casas negociadas na região. O método de renda avalia com base no valor do aluguel (ou outra renda) que pode ser obtido. O método involutivo a avalia considerando os possíveis recebimentos que esta casa pode gerar caso seja transformada em um imóvel comercial e alugada a preço melhor, ou uma escola, ou um mercado, ou qualquer outro empreendimento. O método evolutivo avalia calculando seus custos de produção, como por exemplo, o valor do terreno (este avaliado pelo método comparativo), materiais, mão-de-obra e lucro do incorporador (BAPTISTELLA, 2005).

A norma especifica que o avaliador pode escolher o método que julgar mais adequado, de acordo com “a natureza do bem avaliado, a finalidade da avaliação e a disponibilidade, qualidade e quantidade de informações colhidas no mercado” (ABNT, 2001). Também prevê que em situações atípicas o engenheiro ainda pode empregar novos procedimentos, desde que justificados. A liberdade possui dois lados: por um lado proporciona maior precisão, mas também pode ser subjetiva ou questionável¹⁰.

¹⁰ Sobre subjetividade - verificar nota de rodapé 1. Sobre os questionamentos, um engenheiro de avaliações afirma que ocasionalmente é possível justificar para um mesmo imóvel diferentes métodos avaliatórios ou o emprego de diferentes dados. Altera-se o valor avaliado, mesmo dentro dos parâmetros exigidos em norma. Cabe empregar a ética profissional, bem como os novos procedimentos para reduzir a subjetividade.

2.2 MODELOS DE PREÇOS HEDÔNICOS

A econometria consiste na aplicação de métodos matemáticos e estatísticos para a resolução de problemas de economia ou explicação de relações econômicas (HOFFMAN, 2016). Um modelo matemático é uma “representação técnica da realidade” (ABNT, 2011). Os modelos voltados à compreensão do mercado imobiliário são segmentados em dois grupos: macroeconômicos, que verificam comportamentos das habitações de forma ampla (juros, política monetária, crédito), e microeconômicos, que direcionam o estudo para o comportamento individual e escolhas do consumidor (DANTAS, 2003).

Os modelos de preços hedônicos, empregados na economia urbana, possuem equações do modelo microeconômico de formação de preços – indicando as parcelas implícitas de contribuição de cada característica (GONZÁLEZ, 1997). As primeiras análises de preços hedônicos surgiram na economia agrícola em 1929, quando Fedrick V. Waugh realizou regressões para verificar a influência das características de cor, tamanho da haste e uniformidade dos brotos no preço dos aspargos em Boston, determinando a valorização relativa que os consumidores atribuíam a estas características. Andrew Court descreveu o método empírico em 1939, posteriormente utilizado em muitos outros setores da economia; como a indústria automobilística, a indústria de computadores e o mercado imobiliário (FERREIRA, 1999).

A abordagem dos preços hedônicos provém da hipótese de que os atributos de uma determinada mercadoria servem para propiciar valor a quem os consome (LANCASTER, 1966). Assim, os preços hedônicos estabelecem que um preço pode ser composto a partir de um vetor de características diretamente mensuráveis em um produto (ROSEN, 1974). Uma análise empírica dos preços hedônicos responde a duas perguntas: “1) quais são as características relevantes em um produto?”; e “2) como estas características se relacionam com o preço do produto?” (GRILICHES, 1971). Griliches (1971) relacionou estas questões e preço com o modelo de regressão múltipla e ajuste semilogarítmico, conforme mostra a Equação 1:

$$\log pit = a_0 + a_1x_{1it} + a_2x_{2it} + \dots + uit \quad (1)$$

A avaliação coletiva de imóveis (BAPTISTELLA, 2005) é uma das aplicações de preços hedônicos, buscando valores de todos os imóveis situados numa área urbana, sendo base de cálculo para tributos imobiliários (imposto predial – IPTU e imposto de transmissão – ITBI)¹¹.

¹¹ Calcular estes preços com exatidão é importante para garantir a equidade fiscal e a justiça tributária. Muitos municípios calculam o valor venal (estimativa do Poder Público de valor do imóvel) simplesmente multiplicando o valor médio do m² construído em cada região da cidade (Planta Genérica de Valores) pela área do imóvel.

Aplicações dos preços hedônicos em cidades brasileiras

De forma ampla, é possível encontrar trabalhos brasileiros com preços hedônicos aplicados nas cidades do Rio de Janeiro – RJ (FERREIRA, 2002), Porto Alegre – RS (GONZÁLEZ, 2002), Recife – PE (DANTAS, 2003), Guarapuava – PR (BAPTISTELLA, 2005), Juiz de Fora – MG (MATTA, 2007), São Paulo – SP (ALVES, 2011), Uberlândia - MG (CARMO, 2014), Belo Horizonte – MG (PAIXÃO, 2015), Juazeiro do Norte - CE (GONÇALVES, 2016) e Balneário Camboriú - SC (HORNBERG & HOCHEIM, 2017).

Conforme estas publicações percebe-se que os preços hedônicos são úteis para diferentes finalidades: comparar com a planta de avaliação de imóveis das prefeituras (MATTA, 2007), estimar o valor de apartamentos novos financiados (GONÇALVES, 2016), avaliar o impacto de amenidades urbanas nos imóveis (FERREIRA, 2002; CARMO, 2014), criar índices de preços imobiliários (PAIXÃO, 2015), computar a relação de variáveis macroeconômicas nos preços dos imóveis (ALVES, 2011), estimar a demanda econômica por diferentes unidades (DANTAS, 2003), dentre outras.

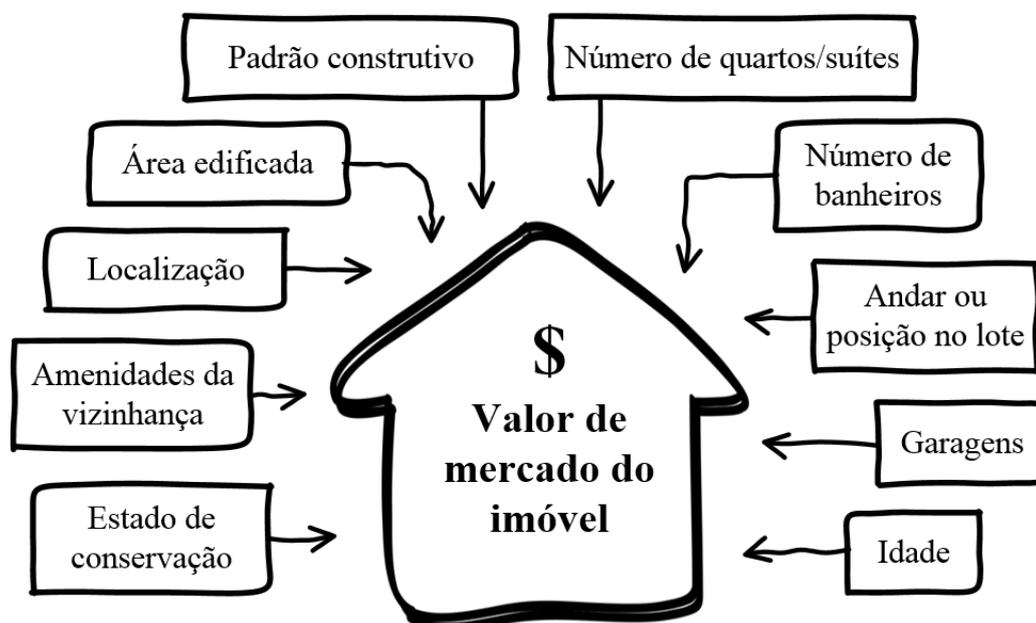
A economia urbana é um ramo da microeconomia que estuda todas as atividades econômicas das áreas urbanas, como moradia, finanças públicas, crime, uso do solo e trânsito. Desta maneira, incorpora a técnica dos preços hedônicos e abrange os estudos citados.

O cálculo dos preços hedônicos exige dados de mercado, sendo associado ao método comparativo de dados de mercado, presente na norma NBR 14653-1:2001 (ABNT, 2001). A fonte dos dados destes trabalhos é majoritariamente das prefeituras, no conjunto do Cadastro Imobiliário e transações da ITBI; ou então transações financiadas pela Caixa Econômica Federal, o maior agente financiador do País. Um dos trabalhos obtém ofertas diretamente de imobiliárias e classificados. Os trabalhos utilizam em média 100 a 200 imóveis para treinar os modelos, com ajustes em valores declarados (subestimados) ou ofertados (superestimados).

Os trabalhos utilizam dados estruturais dos imóveis, e verifica-se que em todos, sem exceção, em maior ou menor grau as características mais relevantes são: “área edificada”, “padrão construtivo”, “número de quartos/suítes” e “distância a um ponto de referência”. A Figura 2 demonstra conceitualmente variáveis compondo os preços hedônicos.

Devido à diversidade de situações e locais, cada trabalho também possui peculiaridades, como por exemplo – os que incluem casas verificam a importância da “posição do lote” (frente ou fundos) e a variável mais relevante em Balneário Camboriú – SC é a “distância à praia”.

Figura 2 - Preço hedônico como uma composição de características do bem



Fonte: O Autor (2017)

Além das variáveis estruturais e de localização, é possível também cruzar variáveis externas para complementar os dados dos imóveis. Ganham-se mais informações; a possibilidade de analisar fatores externos e caso relevantes, maior precisão nas previsões.

Dentre as variáveis complementares incluídas para caracterizar imóveis, encontram-se: dados de renda do Censo do IBGE (DANTAS, 2003), quantidade de equipamentos de educação, saúde e lazer próximos (FERREIRA, 2002), a hierarquia social (PAIXÃO, 2015), ou variáveis mais exóticas, como a quantidade de furtos no bairro (CARMO, 2014).

Todos os trabalhos exceto o dos índices de preços (PAIXÃO, 2015) segmentam a amostra para imóveis residenciais, buscando aumentar a precisão e simultaneamente abranger a maior parte dos casos. A maioria destes ainda utiliza apenas o segmento dos apartamentos, visto que esta é a tipologia mais comum presente nos dados de mercado e nos financiamentos; em especial quando os dados são obtidos na Caixa Econômica Federal.

Em Caxias do Sul, Sara Andrade de Oliveira comparou a evolução dos preços dos imóveis com índices construtivos. Foram utilizados 912 dados de apartamentos da cidade financiados pela Caixa Econômica Federal, abrangendo 18 variáveis, no espaço de 10 anos. Os dados também foram segmentados para apartamentos residenciais (OLIVEIRA, 2017).

A norma também fornece segmentação de imóveis, sendo quanto ao uso: “residencial; comercial; industrial; institucional; misto” e quanto ao tipo de imóvel: “terreno (lote ou gleba); apartamento; casa; escritório; loja; galpão; vaga de garagem; misto; hotéis e motéis; hospitais; escolas; cinemas e teatros; clubes recreativos; prédios industriais” (NBR 14653-2:2011).

O método da regressão linear múltipla é o método preponderante para o cálculo dos preços hedônicos. Este é o método preferido dos avaliadores, especialmente pela sua eficiência. Todos os trabalhos que fazem uso da regressão linear múltipla a resolvem utilizando os mínimos quadrados, e parte considerável ajusta a equação para a forma semilogarítmica.

Dois trabalhos (BAPTISTELLA, 2005; GONZÁLEZ, 2002) calculam os preços hedônicos utilizando redes neurais. A rede neural é um algoritmo que combina camadas de neurônios de maneira não-linear e por vezes alcança maior precisão em relação à regressão linear. O ponto negativo é a possibilidade de sobreajuste¹² nos dados e a impossibilidade de interpretar diretamente o modelo (como na regressão linear), devido a complexidade de suas estruturas de neurônios. A rede neural, sendo um algoritmo computacional não supervisionado, é extremamente útil em casos como *machine learning*. Porém, no mercado imobiliário, impossibilita parcialmente a curiosidade intelectual e interpretações diretas.

Percebe-se que a maioria dos modelos de preços no mercado imobiliário fundamentam-se no método de preços hedônicos. O método dos preços hedônicos proporciona eficiência, versatilidade, consistência, elevado poder explicativo e a possibilidade de revelar a utilidade marginal de cada característica ao consumidor.

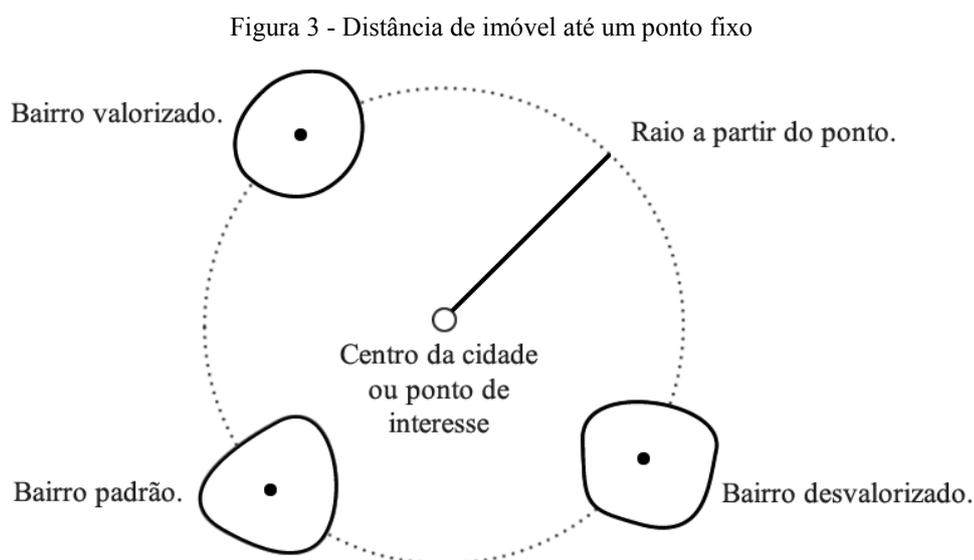
¹² O sobreajuste (*overfitting*), é o mesmo problema que acontece por exemplo, ao ajustar uma equação com um grau polinomial superior à quantidade de dados observados. A equação se ajusta extremamente bem ao conjunto analisado, mas por outro lado, torna-se pouco eficiente para prever novos resultados.

2.3 MODELOS DE APOIO

2.3.1 MODELOS DE REGRESSÃO ESPACIAL

Parte significativa das diferenças de valor estimado nos preços hedônicos não são de ordem técnica, mas conceitual. Os trabalhos citados anteriormente utilizam a econometria tradicional, tratando a questão espacial como uma variável que mede a distância do imóvel a um ponto de influência ou um setor, como o centro da cidade.

O centro da cidade, por exemplo, nem sempre é o polo mais valorizado de uma cidade, podendo ser inclusive o oposto. Também é possível que um local a uma determinada distância do centro seja um bairro de classe alta, enquanto outro local estando à mesma distância do centro (em direção diferente) seja uma região de classe baixa, conforme ilustra a Figura 3:



Fonte: O Autor (2017)

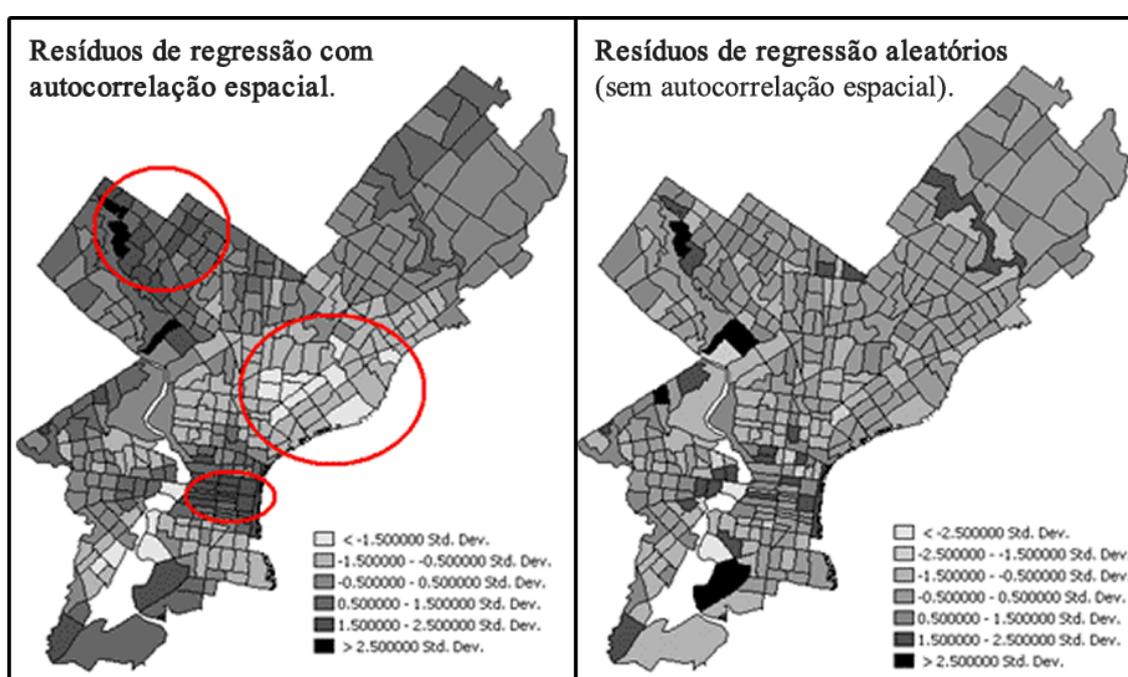
Portanto, observar a cidade como monocêntrica e tratar a questão espacial em função da distância do imóvel a determinados pontos de influência pode não explicar fidedignamente o comportamento do mercado imobiliário.

Marisa Baptistella (2005) conclui suas avaliações com redes neurais e afirma que “outras variáveis adicionais como, acessibilidade, distância do centro comercial, acesso a escolas, hospitais, supermercados também são importantes tanto na determinação da aquisição de um imóvel quanto para a sua valorização e, desta forma, estas variáveis mesmas poderiam também ser levadas em consideração no modelo também melhorando, ainda mais, a previsão dos valores venais de imóveis” (BAPTISTELLA, 2005).

Seria extremamente complexo descobrir todas as amenidades urbanas que valorizam ou desvalorizam imóveis. Podem ser, por exemplo, o impacto de um bar ou parque na esquina, a construção de uma nova avenida ou shopping center, uma boa escola na vizinhança.

Uma solução sofisticada é proceder com análises de econometria espacial. Enquanto a regressão tradicional assume a independência espacial das variáveis e dos resíduos da regressão, a regressão espacial considera os efeitos de dependência e heterogeneidade espacial que os dados espaciais (por exemplo, imóveis) adquirem do território, conforme mostra a Figura 4:

Figura 4 - Apresentação de resíduos contendo dependência espacial



Fonte: Adaptado de BRUSILOVSKIY (2009)

A dependência espacial resulta principalmente da complexidade da interação dos dados com o espaço; enquanto a heterogeneidade espacial é relacionada com a variação das formas funcionais dos dados ao longo de diferentes localizações. Os modelos de regressão espacial permitem trabalhar estatisticamente com estas peculiaridades espaciais (ANSELIN, 1988).

No Brasil, duas pesquisas utilizam a econometria espacial no mercado imobiliário: O modelo de defasagem espacial (DANTAS, 2003), obtendo uma equação mais precisa para avaliar imóveis e estimar sua elasticidade econômica; e a *krigagem bayesiana* (HORNBERG & HOICHEIM, 2017) para determinar relações probabilística de causalidade entre o valor de imóveis e o espaço urbano.

O modelo de defasagem espacial (DANTAS, 2003) considera uma variável defasada espacialmente que captura as interações do território, como um *proxy*¹³ para variáveis locais não consideradas explicitamente. Neste método calculam-se os preços hedônicos e posteriormente realiza-se o ajuste espacial conjugando a metodologia da *kriging* espacial e a metodologia de Anselin (1988), estimando uma matriz de pesos espaciais.

Dantas (2003) demonstrou que a dependência espacial afeta as variáveis econômicas. Cada edifício atua como um micropolo de influência sobre seus vizinhos, estendendo esta influência em até 3 quilômetros. Uma transação elevada valoriza também os imóveis vizinhos, numa espécie de “contágio” espacial. A autocorrelação positiva existe inclusive na densidade habitacional: bairros com densidade populacional semelhantes se encontram próximos uns aos outros, e a verticalização¹⁴ num bairro gera verticalização no bairro seguinte.

De fato, o Lorde Harold Samuel possuía fundamentos quando expressou as prioridades para seus investimentos imobiliários – “*location, location, location*”. O índice de preços de uma unidade de apartamento padrão por bairro em Recife mostra que a localização pode compor até 72% do preço total de um imóvel (DANTAS, 2003). Em virtude da importância da localização, desconsiderar os efeitos espaciais pode levar o modelo a conclusões equivocadas.

Um artigo da University of Connecticut analisou diferentes modelos de regressão espacial e verificou que estes podem aumentar a precisão das estimativas, reduzir os erros autocorrelacionados e servir como *proxies* efetivos para variáveis omitidas correlacionadas com a localização. Concluiu poeticamente – “*real estate and spatial statistics naturally complement each other*” (KELLY et al, 1998).

¹³ Um *proxy* é uma variável não diretamente relevante em um modelo, mas que assume o lugar de outras variáveis “não observadas ou de difícil mensuração” (ABNT, 2001). Uma variável atuando como *proxy* de localização pode capturar, por exemplo, o efeito de amenidades públicas, entornos ou o *status* de um local.

¹⁴ O processo urbanístico da construção de grandes edifícios, que costuma aumentar a densidade populacional e alterar a dinâmica das cidades.

2.3.2 MODELOS COMPOSTOS

No estado-da-arte em avaliações automatizadas, a Zillow avalia diariamente 110 milhões de imóveis americanos, utilizando dados públicos e ofertas de mercado. A precisão de $\pm 5\%$ para a avaliação de 50% dos imóveis do país (ZILLOW, 2016) possibilitaria metade dos imóveis avaliados constarem na Classe III da NBR 14653-1:2001 (ABNT, 2001) (a classe com maior assertividade, onde o avaliador visita o imóvel, que admite precisão de $\pm 5\%$ no valor). A Figura 5 demonstra a possibilidade de consultar livremente o valor dos imóveis sobre um mapa:

Figura 5 - Imóveis avaliados pela Zillow em uma área de Champaign, Illinois



Fonte: Zillow (2017), acesso online.

Embora a Zillow mantenha seu método em sigilo, é possível supor algumas das técnicas utilizadas no livro *Zillow Talk – New Rules of Real Estate* (RASCOFF e HUMPHRIES, 2015). Os autores afirmam que a Zillow utiliza algoritmos de inteligência artificial, de processamento de imagem e genéticos – sugerindo um algoritmo de vanguarda, composto por múltiplos modelos.

Descobriram que a melhor maneira de aumentar a precisão é empregando múltiplos modelos simples¹⁵; que contabilizam individualmente o tamanho do imóvel, número de quartos, e virtualmente cada elemento que afeta seu valor. A Zillow distribui pesos a estes modelos menores, com base em dados históricos de localização. Assim, em vez de modelos amplos e genéricos, obtém-se modelos que contabilizam as diferenças de rua para rua.

¹⁵ Embora contraintuitivo que modelos simples sejam mais eficientes que um modelo complexo, cita-se uma analogia como exemplo: múltiplas fotos em diversos ângulos de um imóvel proporcionam um entendimento melhor sobre o mesmo que apenas uma foto em alta resolução de sua fachada (RASCOFF e HUMPHRIES, 2015).

O fato de modelos compostos (*ensemble models*) serem mais eficientes que um único modelo pode ser entendido como se cada modelo fosse uma espécie de “especialista”. Em vez de consultar apenas um especialista, pode-se perguntar a um grupo deles. Em seguida, combinar suas estimativas; por meio de uma média ou então com outro modelo dedicado a realizar esta tarefa, como uma árvore de decisão (PROVOST & FAWCETT, 2013).

Se todos estes especialistas soubessem exatamente as mesmas coisas, forneceriam as mesmas previsões e não haveriam vantagens nessa abordagem. Por outro lado, caso um fosse instruído em um aspecto ligeiramente diferente do problema, ele poderia fazer previsões complementares, e o conjunto inteiro forneceria mais informações do que qualquer especialista individual. Ou seja, o ideal é que os especialistas cometam *diferentes* tipos de erros, e que seus erros sejam os mais sem conexão possível, e igualmente, que sejam completamente independentes. Assim, a média dos erros tende a se anular, tornando o conjunto melhor que os especialistas individuais (PROVOST & FAWCETT, 2013).

Economia comportamental

A economia está longe de ser uma ciência exata. O psicólogo vencedor do prêmio Nobel de Economia em 2002, Daniel Kahneman, elucidou este fato apresentando a economia comportamental: explicou que as pessoas nem sempre fazem escolhas lógicas ou racionais¹⁶, sendo fortemente influenciadas por fatores emocionais, culturais e pelo comportamento dos outros (KAHNEMAN 2012).

Richard Thaler recebeu o Nobel da Economia em 2017 (durante o desenvolvimento deste trabalho), por explicar diferentes mecanismos comportamentais para a tomada de decisões econômicas. Eventualmente, conhecendo a presença de múltiplos comportamentos nos agentes, os efeitos da economia comportamental podem ser incorporados e investigados a partir de métodos como análise de agrupamentos (*clusters*) ou teoria dos jogos.

¹⁶ Dentre os experimentos que suportam este argumento, há o famoso jogo do ultimato. O propositos inicia com uma soma de dinheiro e deve propor uma divisão desta soma com o respondedor. Caso o respondedor aceite, a soma é repartida e ambos ganham sua parte acordada. Caso rejeite, nenhum recebe. Não existe negociação nem segunda rodada. Por isso o “ultimato”: o respondedor deve aceitar ou declinar a oferta. Pela lógica, caso o propositos iniciar com R\$100,00 e oferecer R\$1,00 ao respondedor, este aceitaria, pois, R\$1,00 é melhor do que zero. Na prática, as diferentes aplicações do experimento mostram que o respondedor tende a recusar ofertas inferiores a 25% do total em jogo. O senso de justiça fala mais alto que a lógica e a racionalidade. Não ganhar nada é menos doloroso do que ser vítima de uma *injustiça*.

CAPÍTULO 3 – CIÊNCIA DE DADOS

O todo é maior que a soma de suas partes.

Aristóteles (350 a.C.),

sobre a complexidade do mundo

INTRODUÇÃO

A ciência de dados (*data science*), segundo a visão centrada em criar valor para negócios, é “a transformação de dados por meio da matemática e da estatística em informações valiosas, decisões e produtos”. Aplicações eficientes destas técnicas auxiliam empresas a tomar decisões mais assertivas, aumentar receitas e reduzir custos (FOREMAN, 2014).

John Foreman (2014) afirma que dentre as técnicas utilizadas na ciência de dados estão aquelas que surgiram na Segunda Guerra Mundial, tais como: modelos de otimização, modelos de previsão e simulações – assim como técnicas mais recentes: detecção de outliers, análise de clusters, *network science* e inteligência artificial.

Drew Conway (2010) propõe três eixos multidisciplinares para conceituar a ciência de dados, sendo eles: conhecimento da área¹⁷; matemática e estatística; ciências da computação. A Figura 6 representa o diagrama:

Figura 6 - Ciência de dados pelo diagrama Venn de Drew Conway's



Fonte: Adaptado de CONWAY (2010).

¹⁷ Dentre as áreas mais associadas à ciência de dados, no quesito quantidade de aplicações e trabalhos publicados, destacam-se as ciências sociais, economia, engenharia, finanças, direito e medicina.

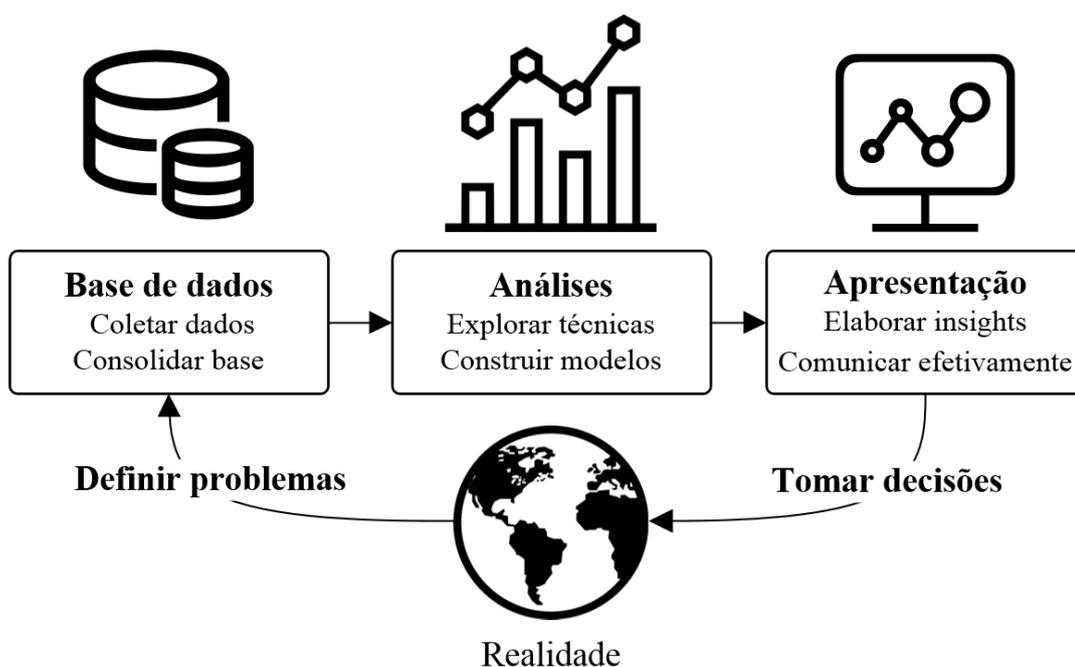
O conhecimento da área proporciona ao pesquisador a capacidade de compreender os problemas específicos do seu campo bem como a sensibilidade para enxergar possíveis soluções. A matemática e estatística fornecem metodologias para extrair informações dos dados. A computação alavanca a execução, com a habilidade de manipular dados, operações vetorizadas, construir algoritmos e pensar sistematicamente (CONWAY, 2010).

Percebem-se outras interações entre os eixos. Quando não se incorpora o conhecimento da área, obtém-se o aprendizado de máquina, em que o computador interpreta dados sem a interferência direta do pesquisador, como nas redes neurais. A pesquisa tradicional inclui a interpretação de dados, sem grande apoio dos métodos computacionais. O desenvolvimento de softwares tradicional não conta com técnicas de matemática e estatística (CONWAY, 2010).

3.1 ETAPAS PARA CIÊNCIA DE DADOS

Diferentes autores (PROVOST & FAWCETT, 2013; FOREMAN, 2014; BANDYOPADHYAY, 2015; SIEGEL, 2016) conceituam o processo da ciência de dados ligeiramente diferente. Embora cada autor possua um foco, é possível representar a transformação dos dados de forma genérica, conforme demonstra a Figura 7:

Figura 7 - Processo tradicional de ciência de dados



Fonte: O Autor, com base nos processos de PROVOST & FAWCETT (2013), FOREMAN (2014), BANDYOPADHYAY (2015) e SIEGEL (2016).

3.1.1 DEFINIR O PROBLEMA

O primeiro passo antes de resolver qualquer problema é delimitar *exatamente* qual é o problema. O conhecimento na área do problema ajuda a entender o *por quê* dos questionamentos e proporciona a sensibilidade para formular perguntas que façam sentido, agreguem valor e possam ser respondidas com os dados. Pode-se perguntar – “O que eu posso fazer agora que não podia fazer antes, ou fazer melhor que antes?”. Ao final deste estágio, é importante possuir o contexto para nortear a resolução do problema, o método e as perguntas a serem respondidas (PROVOST & FAWCETT, 2013).

3.1.2 ORGANIZAR OS DADOS

Definido o problema, é possível determinar e buscar dados que ajudem a obter respostas. Por exemplo, o cálculo dos preços hedônicos exige múltiplos dados do mercado sobre os bens. No caso dos imóveis, a fonte destes dados usualmente é o conjunto do Cadastro Imobiliário com as transações do imposto de transmissão ITBI nas prefeituras; dados de mercado da Caixa Econômica Federal; ou então dados diretamente de imobiliárias e classificados.

As publicações de preços hedônicos empregam em média 100 a 200 dados de imóveis, possibilitando o treinamento adequado de um modelo que considere as variáveis necessárias. Observam também valores de declarações (subestimadas) e ofertas (superestimadas).

Cruzar dados de bases distintas é uma excelente maneira para introduzir novos fatores. Caso sejam expressivos, é possível aumentar a precisão e o poder explicativo do modelo. Por exemplo, O IBGE disponibiliza livremente os dados censitários de 2010. Caxias do Sul possui 742 setores censitário, cada um contendo variáveis potencialmente relevantes, ligadas à população, moradia, renda, educação e infraestrutura (IBGE, 2017).

A consolidação visa unificar todos estes elementos individuais, eventualmente em múltiplos formatos (por exemplo, uma planilha cadastral, pontos espaciais em *shapefiles* ou desenhos no AutoCAD) para uma base única, possibilitando as análises estatísticas e cruzamento de diferentes informações (BIOSSPLENA, 2017).

Por último, a limpeza de dados (*data cleansing*) garante a qualidade dos dados utilizados. Cabe usar critérios e a intuição para detectar valores inconsistentes (como os vazios ou corrompidos), eliminar dados que não fazem sentido (como outliers e duplicados), estimar e corrigir informações quando necessário (BANDYOPADHYAY, 2015).

3.1.3 ANÁLISE EXPLORATÓRIA

Além de encontrar ideias para testar, o maior desafio da análise exploratória consiste em priorizar aquelas que façam sentido e que tragam informações valiosas. Esta análise preliminar busca padrões (previstos ou novos), enxergar possibilidades e priorizar técnicas para análises aprofundadas (SIEGEL, 2016).

A validação do modelo é realizada a partir de critérios pré-estabelecidos. O mais comum é redução quadrática dos erros. Pode-se definir por exemplo, 80% da amostra para treinamento e 20% para validação, possibilitando avaliar técnicas distintas sob a mesma perspectiva.

A análise aprofundada aplica conhecimentos estatísticos, matemáticos e técnicos para alavancar a extração de valor dos dados e transformá-los em *insights* (FOREMAN, 2014). No caso deste trabalho, a proposta é criar um modelo preditivo para avaliar imóveis, e partindo deste elaborar componentes que auxiliem a compreender um mercado imobiliário local.

3.1.4 APRESENTAÇÃO DE RESULTADOS

Apresentar adequadamente os dados é tão importante quanto analisá-los: explicar seu sentido, objetivo, propósitos, priorizar o relevante, enfatizar os pontos mais importantes. As melhores visualizações de dados são aquelas que combinam *insights* qualitativos e análises quantitativas para contar *histórias* que inspirem ação (BANDYOPADHYAY, 2015).

Construir histórias interessantes envolve conectar os dados apresentados com os conhecimentos do interlocutor. Por exemplo, os mesmos resultados podem ser apresentados de maneiras diferentes para um construtor, um engenheiro ou para um consumidor. Cada agente possui conhecimentos e necessidades diferentes, além de tomar diferentes tipos de decisões.

3.1.5 TOMAR DECISÕES

Como resultado final do processo de ciência de dados, buscam-se soluções para os problemas estabelecidos na primeira etapa. O conhecimento adquirido possibilita a tomada de decisões mais assertivas, que proporcionam menores incertezas e custos; e maior produtividade, faturamento e lucratividade em ambientes de negócios (PROVOST & FAWCETT, 2013).

3.2 REGRESSÃO LINEAR MÚLTIPLA

A regressão linear é uma das técnicas mais populares da estatística e da ciência de dados, sendo considerada uma forma básica de inteligência artificial (FOREMAN, 2014). Essencialmente, qualquer área¹⁸ onde exista uma quantidade satisfatória de dados históricos pode ser beneficiada com modelos de regressão linear (PROVOST & FAWCETT, 2013).

A regressão linear múltipla é a principal técnica utilizada na construção deste trabalho, sendo a técnica majoritariamente associada aos preços hedônicos e à econometria. A NBR 14653 (ABNT, 2001) também a indica para o método comparativo direto de dados de mercado.

O modelo de regressão linear simples define que o valor de uma variável dependente é função de uma variável independente, ou explanatória (HOFFMANN, 2016). A regressão linear múltipla admite mais de um coeficiente. A equação da regressão linear múltipla para um problema com k variáveis explanatórias é representada por meio da Equação 2 e Equação 3:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \varepsilon_j, \quad (2)$$

para os dados $j = 1, \dots, n$

$$Y_j = \beta_0 + \sum_{i=1}^k \beta_i X_{ij} + \varepsilon_j \quad (3)$$

Ou em sua forma matricial, como demonstrado na Equação 4 e Equação 5:

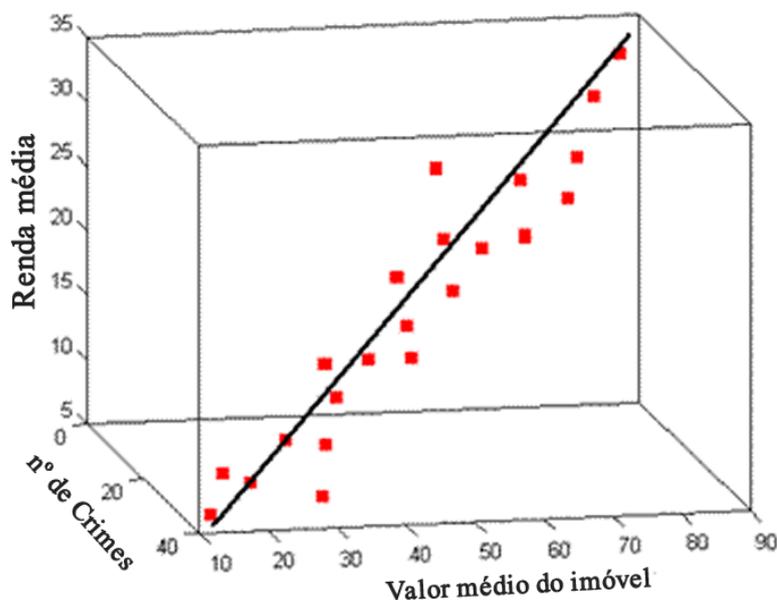
$$Y = X\beta + \varepsilon \quad (4)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (5)$$

Os dados podem ser representados graficamente no espaço. A regressão linear simples representa o eixo x (variável dependente) e eixo y (variável independente). A regressão linear múltipla acrescenta um novo eixo para cada variável independente (ou explanatória). A Figura 8 mostra simultaneamente três eixos, representando uma regressão linear múltipla com uma variável dependente e duas variáveis independentes (BRUSILOVSKIY, 2009).

¹⁸ Modelos de regressão linear são aplicados para resolver problemas de negócios em quase todas as indústrias, como por exemplo bancos para avaliar riscos de empréstimo e identificar transações suspeitas, supermercados para analisar quais ofertas atraem o consumidor, redes de varejo para expandir em localizações mais rentáveis, além de governos e outras indústrias como seguradoras, hospitais, transportadoras, empresas áreas... (FOREMAN, 2014)

Figura 8 - Regressão linear múltipla



Fonte: Adaptado de BRUSILOVSKIY (2009).

É possível utilizar diferentes funções objetivas para realizar o ajuste de uma equação de regressão linear múltipla. Cada procedimento de modelagem utiliza uma escolha particular, conforme for mais apropriado ao problema (HOFFMANN, 2016).

A função objetiva mais comum para resolução de regressão linear (padrão) faz uma escolha conveniente. Sendo o alvo do ajuste uma variável numérica, a função aproxima ao máximo a equação e os dados de treinamento. Observando a diferença entre os dados de treinamento e os dados ajustados, obtém-se o erro absoluto de cada ponto. Assim, o critério para ajustar a curva da regressão é a menor soma possível de todos os erros absolutos (PROVOST & FAWCETT, 2013).

O critério padrão para resolver a regressão linear minimiza a soma dos *quadrados* dos erros – o que dá ao procedimento o nome de regressão dos “quadrados mínimos”. Esta função foi introduzida pelo matemático Carl Friedrich Gauss, e há grandes vantagens em sua escolha. De forma pragmática, prefere-se o ajuste com os quadrados mínimos pois este penaliza fortemente os erros muito grandes (PROVOST & FAWCETT, 2013).

O desvio padrão (σ) auxilia a analisar uma amostra de dados, avaliando a dispersão média desta amostra em relação a sua média amostral, conforme a Equação 6:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N - 1}} \quad (6)$$

Assumindo uma distribuição normal (gaussiana), espera-se que 68% dos valores da amostra estejam compreendidos em $\pm 1\sigma$, 95% dos valores em $\pm 2\sigma$ e 99,7% dos valores em $\pm 3\sigma$. O desvio padrão é particularmente útil na investigação de eventuais pontos anormais da amostra (*outliers*), que podem prejudicar a precisão do modelo da regressão linear.

3.2.1 COEFICIENTE MÚLTIPLO DE DETERMINAÇÃO - R^2

O coeficiente múltiplo de determinação R^2 indica o grau de ajuste da regressão linear múltipla aos dados observados, determinando o poder explicativo deste modelo. O cálculo deste coeficiente mede a proporção entre os erros quadráticos explicados pelo modelo ($SQT-SQR$) em relação aos erros quadráticos totais possíveis (SQT), conforme a Equação 7:

$$R^2 = \frac{SQT - SQR}{SQT} \therefore \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1j} - \dots - \beta_i x_{ij})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Um R^2 mais elevado indica um modelo com maior poder explicativo. Por exemplo, um R^2 de 0,90 expressa que 90% das variações da variável dependente explicam-se pelos regressores presentes na equação (HOFFMANN, 2016).

3.2.2 COEFICIENTE DE SIGNIFICÂNCIA - TESTE F

Embora o coeficiente R^2 determine quão bem o modelo se ajusta aos dados observados, sozinho não determina se o modelo é estatisticamente significativo, ou seja, a possibilidade que o modelo se ajustou a estes dados por coincidência. A prática padrão para rejeitar a hipótese nula, caso seja verdadeira, é obter um ajuste com significância menor que $\alpha = 5\%$. Esta significância, chamada de valor p, é calculada a partir do teste F, na Equação 8:

$$F = \frac{\frac{R^2}{k - 1}}{\frac{1 - R^2}{(N - k)}} \quad (8)$$

O teste F combina informações do modelo: a proporção entre os erros explicados e não explicados ($R^2 / (1-R^2)$), o número de coeficientes do modelo (k), o número de amostras utilizadas (N) e a quantidade de graus de liberdade ($N - k$).

Quanto maior o teste F, menor a probabilidade da existência da hipótese nula. O valor p é encontrado na tabela de distribuições F. Por exemplo, um modelo com $R^2 = 0,90$, com 3 coeficientes (intercepto + 2 variáveis) e 5 amostras possui teste F igual a 9. Assim, o valor de significância p é 2,20% de probabilidade que este ajuste aconteceu por coincidência. Pode-se obter um ajuste melhor empregando mais dados ou utilizando menos coeficientes. Neste mesmo caso, com 50 dados, o teste F resultaria em 211,5, com valor $p = 4 \times 10^{-23} \% < \alpha$, rejeitando a hipótese nula (FOREMAN, 2014).

3.2.3 COEFICIENTE DE SIGNIFICÂNCIA - TESTE T

Porém, enquanto o teste F verifica se a regressão como um todo é significativa, existe a possibilidade de conferir a significância de cada coeficiente individualmente. Assim, consegue-se identificar quais coeficientes contribuem para o sucesso da previsão e eliminar aqueles estatisticamente insignificantes, ou então, investigar a validade da amostra de dados.

O teste de significância para cada coeficiente é chamado de teste t. Similar ao teste F, a hipótese nula supõe que o coeficiente é insignificante e simula sua inexistência. O teste t, conforme a Equação 9, é realizado em função: de cada coeficiente b_i ; dos graus de liberdade ($N - k - 1$); do estimador de erro do modelo S_e^2 ; e por fim, do coeficiente de erro padrão do coeficiente – na regressão linear múltipla obtido nas diagonais da matriz inversa (X^{-1}) da matriz SSCP (*sum of squares and cross products*), sendo esta o produto da matriz da regressão pela sua transposta ($X'X$) (FOREMAN, 2014).

$$t_{(N-k-1)} = \frac{b_i}{\sqrt{S_e^2 * (X'X)^{-1}}} \quad (9)$$

Sendo S_e^2 o estimador de erro do modelo, conforme a Equação 10:

$$S_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1j} - \dots - \beta_i x_{ij})^2}{N - k - 1}} \quad (10)$$

Semelhante ao teste F, a partir do resultado do teste t utiliza-se a tabela de distribuições t para descobrir a significância de cada coeficiente na explicação do modelo.

3.2.4 COEFICIENTE DE CORRELAÇÃO - ρ

A correlação demonstra como duas variáveis relacionam-se, podendo indicar relações de causalidade. Calcula-se por meio do coeficiente de Pearson (ρ), utilizando a Equação 11, onde $\hat{\sigma}_{x,y}$ é a variância (dispersão) das variáveis analisadas (x e y) e $\hat{\sigma}_{xy}$ é a covariância (x e y).

$$\rho = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \therefore \frac{\sum_{i=1}^n [(x_i - \bar{x}) * (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

O coeficiente de correlação de Pearson varia entre -1 e +1. Valores próximos a $\rho=0$ indicam a ausência de correlação. Um valor de $\rho=+1$ indica correlação perfeita entre as variáveis (ambas crescem na mesma magnitude). Um valor de $\rho=-1$ indica correlação *inversa* perfeita entre as variáveis (uma cresce e outra decresce) (HOFFMANN, 2016).

A metodologia de inferência estatística para o mercado imobiliário (GONZÁLEZ, 1997), interpreta a correlação $0 > |\rho| > 0,30$ como correlação fraca; $0,30 > |\rho| > 0,60$ correlação média; $0,60 > |\rho| > 0,90$ correlação forte; $0,90 > |\rho| > 1$ correlação fortíssima e $|\rho|=1$ correlação perfeita.

3.2.5 CONSIDERAÇÕES

A utilidade marginal para de cada uma das características que compõem preço hedônico pode ser encontrada – calculando a primeira derivada parcial de cada componente em relação à equação da regressão dos preços hedônicos (GRILICHES, 1971).

Podem-se realizar “previsões” projetando uma série histórica para um tempo futuro. É preciso lembrar que os modelos são eficientes pois fornecem respostas baseadas em dados, porém são “mágicos”. Embora as tendências sejam úteis na maior parte do tempo, é necessária precaução especialmente em campos como a economia, repletos dos fenômenos conhecidos como “cisnes negros” (TALEB, 2015) – crises, bolhas financeiras, reviravoltas políticas – que tornam incertas ou invalidam tais previsões¹⁹.

¹⁹ Os fenômenos conhecidos como “cisnes negros” (TALEB, 2015) possuem este nome, pois, caso pudessem ser previstos, evidentemente não aconteceriam. Como exemplo, os bancos em Wall Street construíram indicadores de investimentos baseados na premissa que o mercado imobiliário cresceria indefinidamente. A busca por riscos cada vez mais arrojados e a confiança excessiva desencadearam a crise imobiliária de 2008. Diante de tantas previsões econômicas (preço do dólar, mercado de capitais, juros), cabem sábias palavras do antigo filósofo chinês Lao Tzu: “aqueles que têm conhecimento não fazem previsões, e que aqueles que fazem previsões não têm conhecimento...”

3.3 TÉCNICAS DE APOIO

Esta seção introduz brevemente técnicas mais complexas. Embora sejam secundárias ao objetivo principal do trabalho, estas técnicas colaboram no planejamento conceitual do método.

3.3.1 REGRESSÃO ESPACIAL

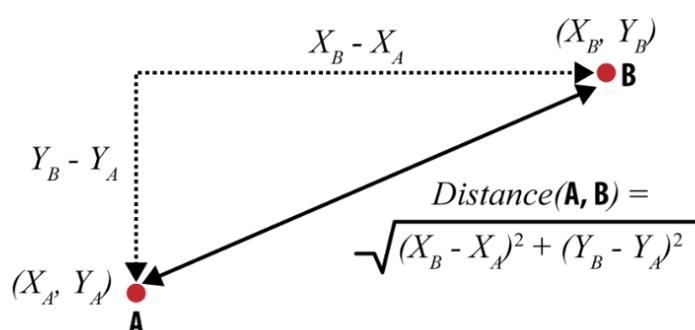
As técnicas de regressão espacial expandem a regressão linear. Vinculam os dados ao território, considerando os efeitos da dependência e heterogeneidade espacial (ANSELIN, 1988). No caso do mercado imobiliário, cada edifício atua como um micro polo de influência, e também é influenciado pelos seus vizinhos (DANTAS, 2003).

Existem múltiplas técnicas para se explorar a regressão espacial, tais como o modelo de *Moran's I*, o *Spatial Lag Model* e o software *GeoDA* (BRUSILOVSKIY, 2009). O modelo de defasagem espacial (DANTAS, 2003) utiliza uma variável para captar as interações espaciais, como um *proxy* para variáveis locais não consideradas explicitamente. Os preços hedônicos são calculados e a partir deles realiza-se o ajuste espacial, conjugando a metodologia de *kriging* espacial da geoestatística e a metodologia de Anselin (1988), para estimar uma matriz de pesos espaciais onde cada ponto corresponde a uma coordenada no território.

3.3.2 ANÁLISE DE CLUSTERS

A similaridade é base para resolver muitos métodos de ciência de dados e problemas de negócios. Parte-se da premissa que dois dados (pessoas, empresas, produtos, imóveis) semelhantes em algum aspecto possivelmente também compartilham outras características em comum. Determinam-se os elementos semelhantes por meio de uma medida de distância entre seus atributos – como a distância euclidiana (FOREMAN, 2014), demonstrada na Figura 9:

Figura 9 - Distância euclidiana



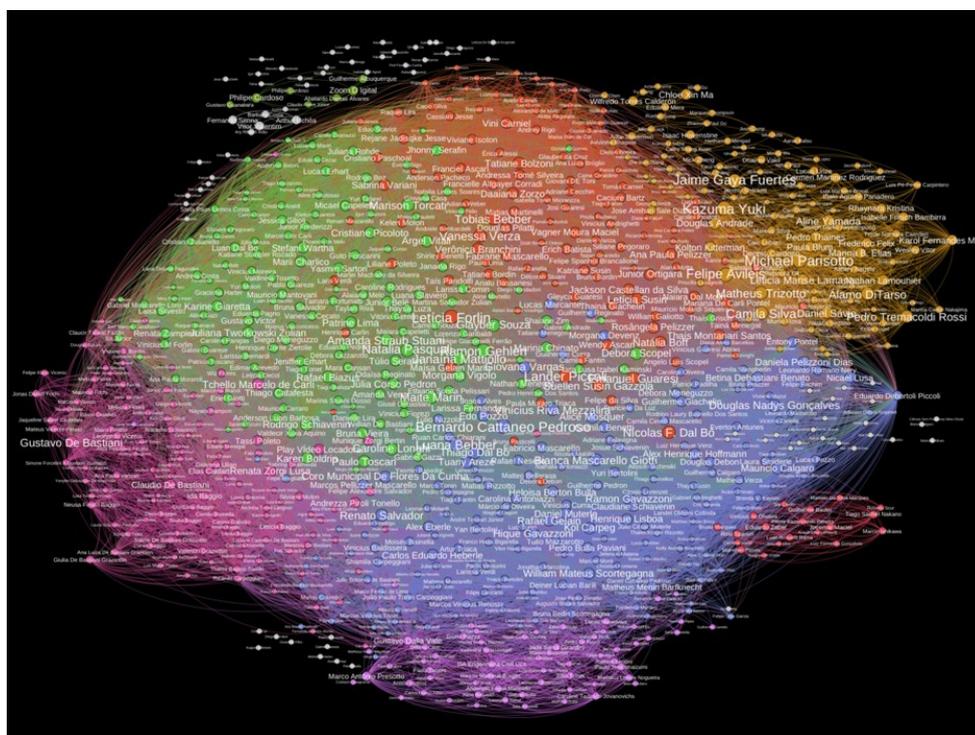
Fonte: PROVOST & FAWCETT (2013)

Uma possibilidade para os imóveis é verificar –se– e –como– unidades semelhantes possuem preços semelhantes; extraíndo assim potenciais informações relevantes, ideias e maneiras para explorar e aprimorar a precisão do modelo preditivo.

A similaridade pode também ser utilizada para classificação. Assim é possível precificar de formas semelhantes imóveis semelhantes. A NBR 14653:2001 recomenda que se “utilize na avaliação uma amostra de imóveis semelhantes ao imóvel avaliado” (ABNT, 2001). González (1997) aponta que a demanda é diferente em diferentes localizações e existem vários “submercados” para cada tipo de imóvel; reforçando a importância de avaliar os semelhantes.

O algoritmo *k-means* (FOREMAN, 2014) classifica dados de modo não supervisionado. Pode-se avaliar, por exemplo, imóveis com características ou localização semelhantes, explorando padrões relativamente ocultos nos dados. A Figura 10 demonstra um exemplo de classificação – numa analogia, o algoritmo *k-means* encontra grupos de relacionamentos (família, escola, universidade, estágio) por meio de conexões entre amigos no *Facebook*.

Figura 10 - Classificação por clusters a partir do algoritmo *k-means*



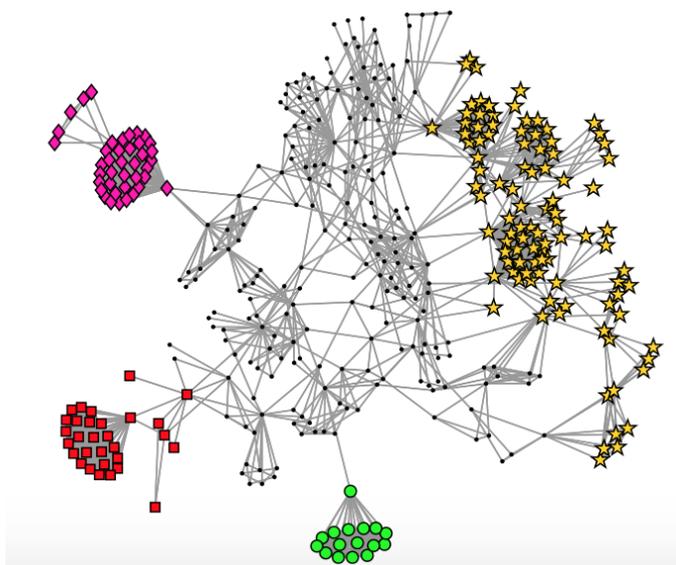
Fonte: O Autor (2016).

3.3.3 NETWORK SCIENCE

A análise de redes (*network science*) (BARABASI, 2014) é uma técnica relativamente nova e promissora: sistemas que representam conexões existentes entre elementos de estruturas. Por exemplo, redes de infraestrutura (sistemas elétricos, hidráulicos e viários de cidade), sociais, econômicas, biológicas, culturais, dentre outras.

As redes são compostas por nós (*nodes*) interligados por conexões (*links*), conforme mostra a Figura 11. Estrutura-se uma rede identificando e conectando seus elementos. As redes são extremamente úteis para descrever e quantificar a conexão e interação entre os elementos, possibilitando compreender comportamentos e padrões emergentes (NETSCIED, 2016).

Figura 11 - Representação de uma rede com elementos conectados



Fonte: NETSCIED (2016).

No caso do mercado imobiliário, existe a possibilidade de explorar conexões entre imóveis para eventualmente descobrir padrões em relacionamentos, aprimorar o método, quantificar questões, como por exemplo – o impacto que alguns imóveis causam sobre outros, a relação entre os imóveis e o território, dentre outras perspectivas.

CAPÍTULO 4 – DESENVOLVIMENTO DE MODELO PREDITIVO

A imaginação é mais importante que o conhecimento.

Albert Einstein

INTRODUÇÃO

Esta investigação consiste em três fases sequenciais, similares à transformação dos dados, descrita no Capítulo 3. A primeira fase consiste na organização da base de dados, a segunda fase no desenvolvimento do modelo preditivo e a terceira fase na elaboração do estudo de caso. A Figura 12 apresenta um fluxograma com a sequência das etapas do método e suas associações com os procedimentos de pesquisa – Figura 13 – e com os objetivos propostos. A metodologia científica (finalidade, método, objetivo, abordagem e procedimentos) (FONTENELLE, 2017) utilizada neste trabalho é classificada conforme o Quadro 2:

Quadro 2 – Metodologia de pesquisa

Finalidade	Método	Objetivo	Abordagem	Procedimentos
Aplicada	Indutivo	Exploratório Descritivo	Quantitativa	<i>(Figura 12)</i>

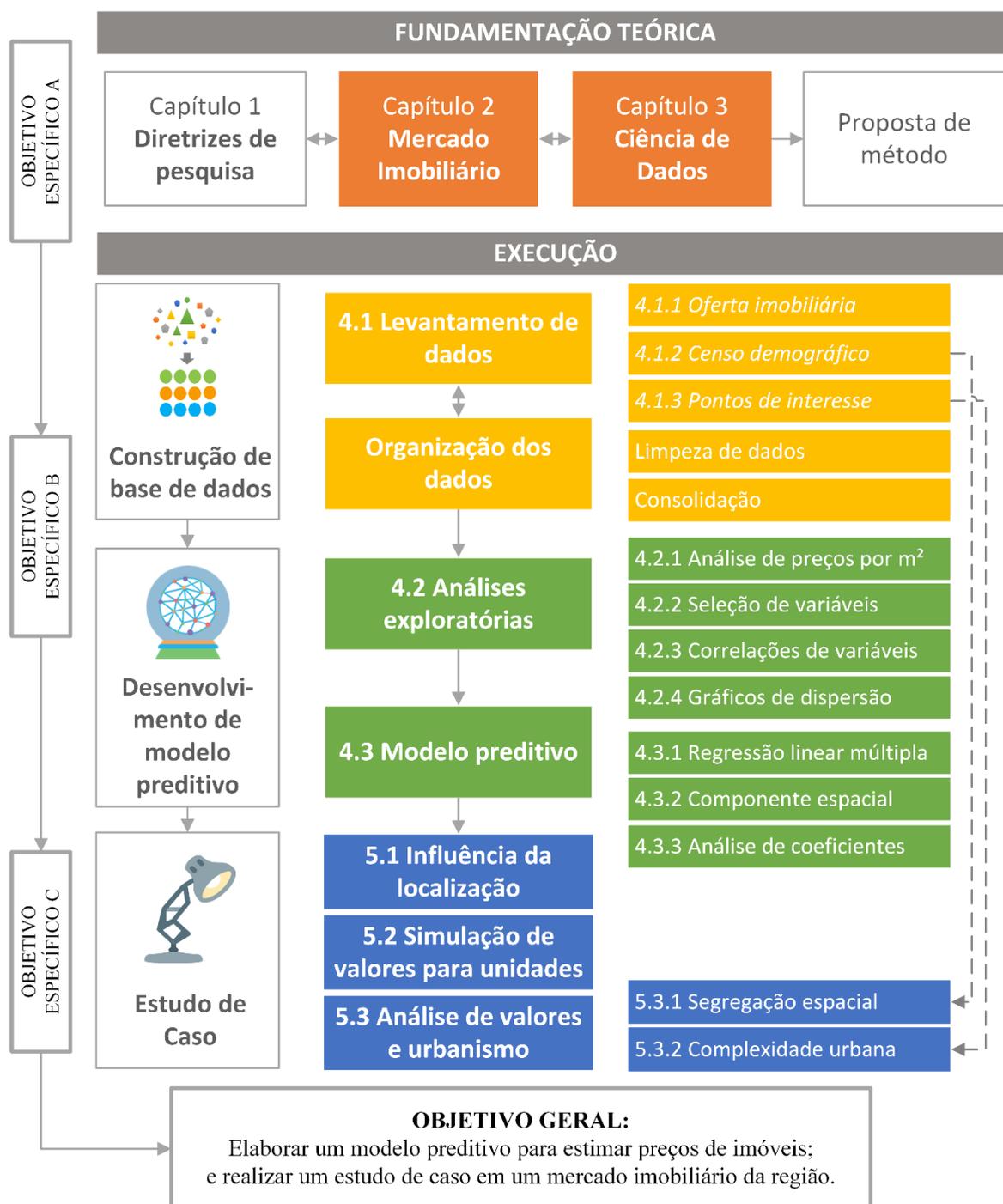
Fonte: O Autor (2017)

A finalidade da pesquisa *aplicada* visa solucionar um problema específico (valores no mercado imobiliário), empregando o método *indutivo*: partir de observações específicas (dados de imóveis) para buscar premissas gerais (mercado imobiliário).

A revisão bibliográfica consistiu em uma pesquisa exploratória, buscando conexões entre o mercado imobiliário e a ciência de dados, visando “a descoberta, o achado e a elucidação de fenômenos” (NETTO, 2006). A execução do trabalho também possui o objetivo *exploratório* (identificar fatos ou fenômenos, tornando-os mais claros, propor problemas ou hipóteses) bem como *descritivo* (expor, classificar e interpretar fatos ou fenômenos).

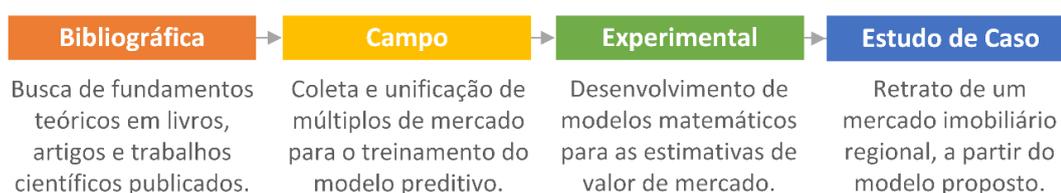
Utiliza-se a abordagem *quantitativa*, caracterizada pelo uso de ferramentas e técnicas estatísticas para análise dos dados, na qual são realizadas observações e constatações e as conclusões são obtidas grau de probabilidade, ao invés de certeza (FONTENELLE, 2017).

Figura 12 - Desenvolvimento de pesquisa



Fonte: O Autor (2018).

Figura 13 - Procedimentos de pesquisa



Fonte: Baseado em FONTANELLE (2017).

4.1 COLETA E ORGANIZAÇÃO DE DADOS

A presença de dados é fundamental para o desenvolvimento de um modelo preditivo. O levantamento de dados compreendeu uma amostra de imóveis localizados na cidade de Caxias do Sul – RS. O Quadro 3 sintetiza as fontes de dados empregadas nesta investigação:

Quadro 3 – Definição dos dados empregados

Classificação	Origem	Abrangência	Período
Oferta imobiliária	Caixa Econômica Federal – GIHAB/CX	1.744 fichas de pesquisa de apartamentos de uso residencial	2016-2018
Informações socioeconômicas	IBGE – Censo Demográfico	682 setores censitários	2010
Pontos de interesse	Google Maps API	3.399 pontos de interesse	2018

Fonte: O Autor (2018).

Discutem-se a seguir os métodos e tratamentos com os quais estes dados foram coletados, organizados, ajustados, unificados e priorizados, permitindo as análises exploratórias e o desenvolvimento do modelo preditivo.

4.1.1 OFERTA IMOBILIÁRIA – CAIXA ECONÔMICA FEDERAL

A Caixa Econômica Federal é o maior agente financiador imobiliário do País e conta com uma série de profissionais e empresas externas executando avaliações imobiliárias – cerca de 60 em Caxias do Sul e aproximadamente 500 no Rio Grande do Sul. Em cada avaliação imobiliária o avaliador preenche uma ficha de pesquisa, cadastrando ao menos um novo dado de mercado ou transação imobiliária no SIMIL – Sistema de Informações do Mercado Imobiliário – um sistema colaborativo em que os agentes enviam e recebem dados para aumentar a eficiência de suas avaliações imobiliárias (GIHAB-CX, 2018).

Por meio da Gerência Executiva de Habitação de Caxias do Sul, a Caixa Econômica Federal disponibilizou um conjunto de dados de mercado imobiliário (GIHAB-CX, 2018) para o desenvolvimento deste trabalho.

Os dados contemplam 1744 fichas de pesquisa de apartamentos de uso residencial, contendo ofertas de imóveis e transações imobiliárias no município de Caxias do Sul - RS. Estas informações são obtidas de imobiliárias, vendedores, corretores e sites. Em grande parte delas o agente realiza uma visita presencial para verificar os atributos do imóvel.

Cada imóvel é registrado com cerca de 250 colunas de informação, compreendendo âmbitos como endereço e localização geográfica, caracterização do entorno, características construtivas da unidade, amenidades disponíveis, data de avaliação e valor de mercado.

Os dados de oferta imobiliária encontram-se relativamente bem estruturados, organizados e classificados, tendo em vista a finalidade de suportar avaliações imobiliárias. Observando que o banco de dados é resultado da colaboração entre centenas de avaliadores, é importante a precaução para verificar possíveis inconsistências. O tratamento de dados foi realizado a partir de critérios, executando as seguintes ações:

1. Filtro de dados – Seleção de transações/ofertas de apartamentos residenciais em Caxias do Sul - RS, em um período entre o início de 2016 e final de 2017 (24 meses);
2. Tratamento de *outliers* – Ajuste ou remoção de pontos excessivamente fora da curva, tais como apartamentos com menos de 20 m² ou mais de 1000 m²;
3. Correção de inconsistências – Retificação de campos com inconsistência na digitação, por exemplo, 20 garagens (acréscimo de zero) ou latitude +29 (sinal inverso);

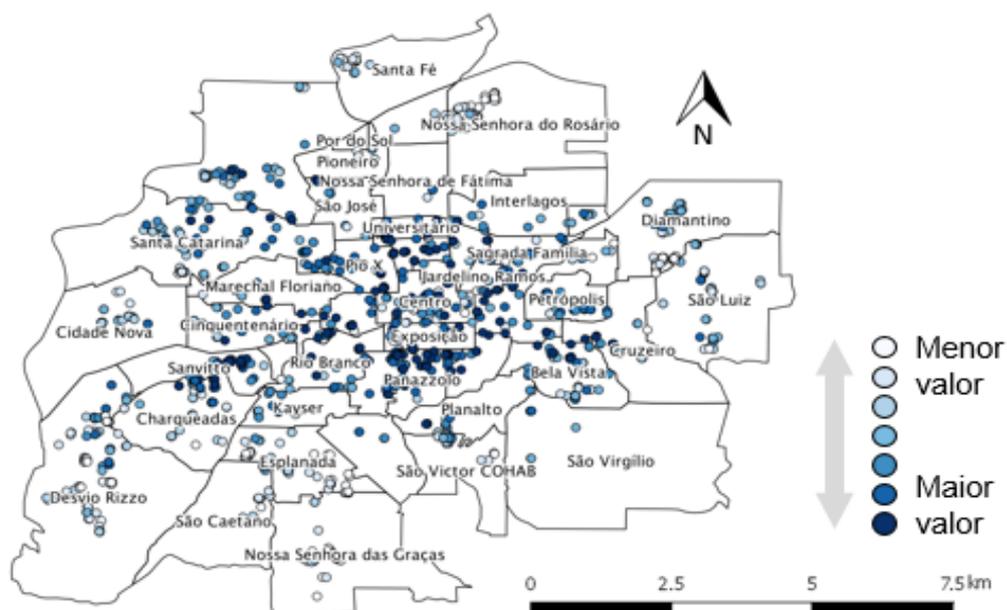
4. Quantificação de campos textuais – Quantificar campos textuais, como por exemplo, “baixo, normal, alto” para “1, 2, 3”, permitindo o cálculo do modelo de regressão;
5. Remoção de duplicados – Remover unidades idênticas, cadastradas mais de uma vez;
6. Estimativa de campos – Na ausência de algum campo secundário, estimar o mesmo por meio de interpolação de unidades semelhantes ou média de categoria próxima;
7. Priorização de informações – Seleção de colunas mais representativas da amostra, abrangentes e relevantes em relação as informações necessárias para estimar o valor.

Executado este tratamento de dados, delimitou-se uma área de Caxias do Sul - RS para a abrangência do estudo. Seleccionaram-se primeiramente os bairros urbanos, formando uma massa territorial única. Em seguida, elegeram-se os bairros contendo ao cinco dados de mercado, para evitar sobreajustes espaciais no modelo.

A área definida para o estudo abrange 47 bairros na cidade, contendo 1.539 dados de imóveis disponíveis para o modelo. Esta área corresponde a 119,1 km² (7,24% do território municipal), e concentra a maior parte da população do município – 5 em cada 6 habitantes de Caxias do Sul residem neste território (355.443 habitantes ou 83,3% do total).

A Figura 14 exibe a área selecionada e a distribuição espacial dos imóveis. A escala em azul mostra os valores de mercado (separados em sete quantis de mesma contagem), entre R\$83.900,00 e R\$1.400.000,00 – média de R\$244.457,50 e desvio padrão de ± R\$156.593,17.

Figura 14 - Distribuição espacial dos dados de mercado - imóveis



Fonte: O Autor (2018)

4.1.2 INFORMAÇÕES SOCIOECONÔMICAS – CENSO DO IBGE

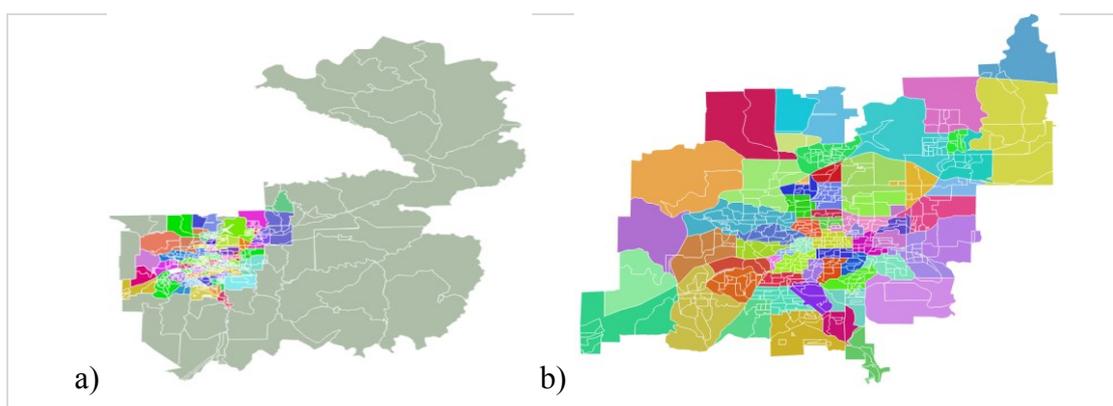
O IBGE – Instituto Brasileiro de Geografia e Estatística – realiza a cada 10 anos o Censo Populacional, contendo informações que elucidam “quem é, quantos são, onde e como vive a população brasileira”. O objetivo é contribuir para a definição de políticas públicas e tomada de decisões – para o setor público, iniciativa privada e o terceiro setor. Os dados produzidos pelo IBGE são de acesso livre e disponibilizados gratuitamente (IBGE, 2010a).

O IBGE divide o território brasileiro em setores censitários, cada um contendo cerca de 500 moradias. O recenseador visita parte dos domicílios para realizar o questionário de pesquisa. O IBGE, posteriormente, faz o tratamento estatístico destas amostras e compila cerca de 1000 variáveis para cada setor censitário existente (IBGE, 2010a).

A cidade de Caxias do Sul - RS é composta por 682 setores censitários (IBGE, 2010). Os setores censitários de Caxias do Sul²⁰ e os microdados agregados por setor censitário²¹ foram obtidos no site do IBGE (2010) e unidos por meio do software de geoprocessamento QGIS (2018), conforme representados na Figura 15. Selecionaram-se as variáveis buscando acrescentar informações para aumentar a eficiência do modelo preditivo:

- V005 – Rendimento mensal médio de responsáveis por domicílios (com e sem renda);
- V011 – Rendimento mensal médio das pessoas de 10 anos ou mais de idade no domicílio (apenas pessoas com renda).

Figura 15 - Setores censitários (682) em Caxias do Sul, RS – a) Área total e b) área urbana



Fonte: Representação do Autor (2017) com dados do Censo do IBGE (2010).

²⁰ Setores censitários disponíveis no formato *shapefile* no endereço <https://ftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios_divisoes_intramunicipais/censo_2010/>

²¹ Resultados do Censo 2010 agregados por setor censitário disponíveis no formato *csv* no endereço <https://downloads.ibge.gov.br/downloads_estatisticas.htm> na pasta /Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios

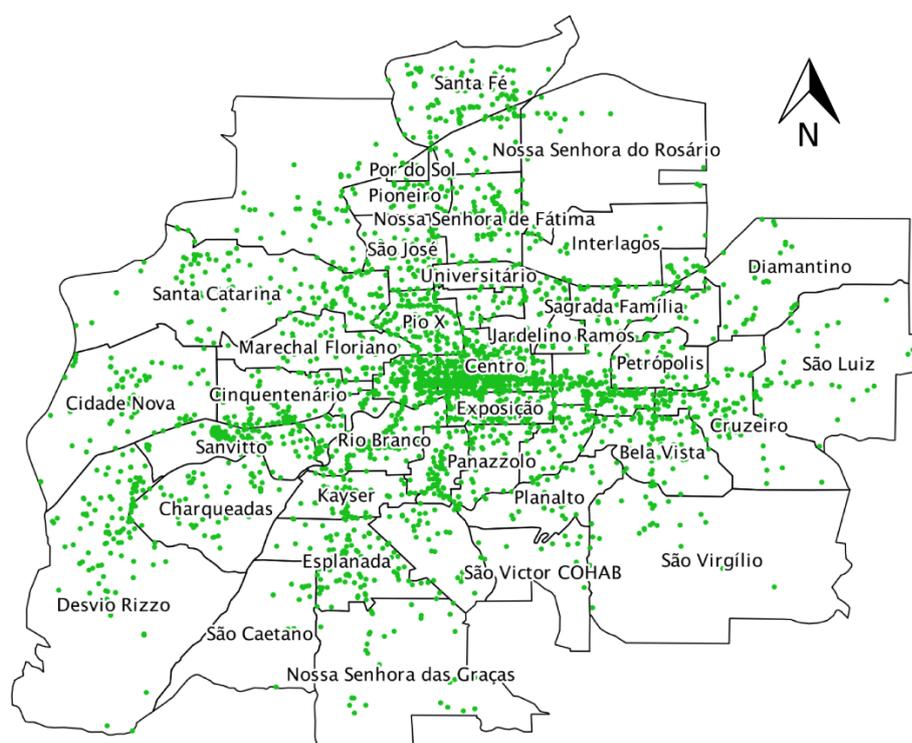
4.1.3 PONTOS DE INTERESSE – GOOGLE MAPS API

Os pontos de interesse selecionados representam comércios, tais como - restaurantes, padarias, farmácias, hotéis, lojas – da área selecionada na cidade de Caxias do Sul - RS. Com esta informação é possível realizar um estudo de caso para validar o modelo e investigar a complexidade urbana da cidade, por meio de uma análise cruzada de informações.

O *Google* utiliza a informação de pontos de interesse em seus serviços de buscas e mapas e também disponibiliza consultas por meio da *Google Maps API*, com informações de 100 milhões de locais ao redor do mundo (GOOGLE, 2018). A pesquisa foi automatizada com a construção de um robô para coletar todos os pontos de interesse na área abrangida e suas respectivas informações. As requisições foram realizadas na modalidade *Radar Search*, na qual se fornece uma coordenada e um raio para a coleta de pontos. Tendo em vista o limite de 200 locais por requisição, o território foi dividido em 24 requisições, com raios menores no centro.

Foram encontrados 3.399 pontos de interesse na área dos 47 bairros selecionados em Caxias do Sul – RS. Estes pontos correspondem a 84,4% do total de pontos na cidade (aproximadamente 5 em cada 6 pontos). Em seguida, os dados foram posicionados no mapa, conforme mostra a Figura 16, para posterior integração com o estudo de caso.

Figura 16 - Distribuição espacial dos pontos de interesse



Fonte: O Autor (2018)

4.2 ANÁLISE EXPLORATÓRIA

A análise exploratória busca investigar e compreender os dados analisados, encontrar padrões (previstos ou novos), explorar possibilidades e técnicas, visando proporcionar uma melhor explicação destes dados.

Dentre as análises, realizou-se a comparação entre valor médio do m² da amostra em relação a outros estudos publicados; seleção de variáveis para o uso no modelo; avaliação das correlações presentes entre as variáveis selecionadas e interpretação de gráficos de dispersão.

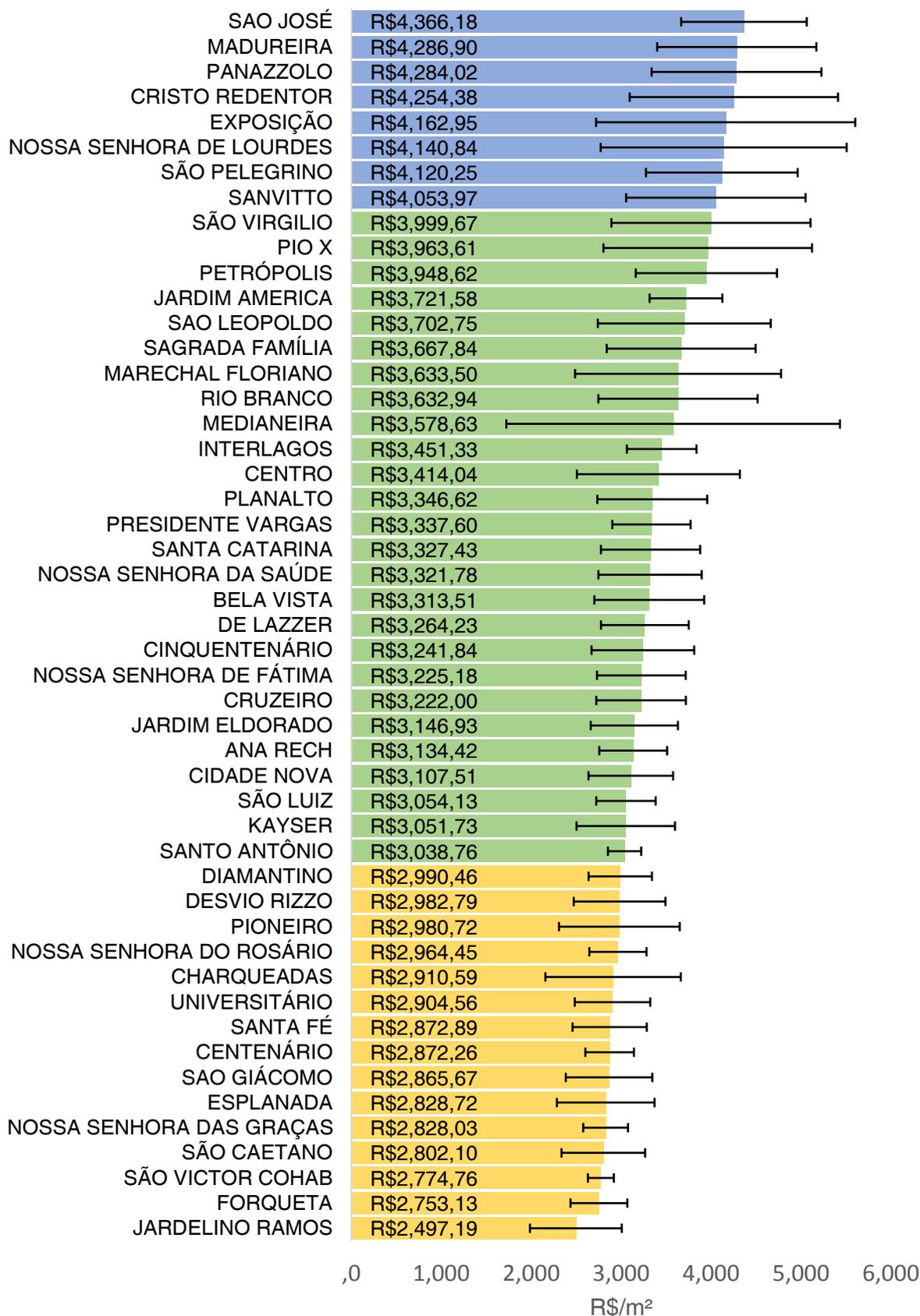
5.2.1 ANÁLISE DE PREÇOS POR METRO QUADRADO

Esta análise exploratória inicial visou analisar a acuracidade dos dados de mercado existentes na amostra utilizada neste trabalho. Primeiramente, verificou-se a média do preço do m² de cada imóvel agregados em relação ao preço do m² em cada bairro. Em seguida, compararam-se com os dados dos bairros mais valorizados de Caxias do Sul, avaliados em estudo do ZAP-Viva Real publicado no Jornal Pioneiro em julho de 2016 (PIONEIRO, 2016).

É importante ressaltar que estes dados partem diretamente do mercado, sem tratamentos ou homogeneizações posteriores. Portanto, indicam o valor médio do metro quadrado *existente* em cada bairro. Cada bairro possui padrões de construção que o diferenciam do metro quadrado existente em outros bairros. Ou seja, além das localizações, estas diferenças incorporam também outras características, como o padrão construtivo dos imóveis presentes naquela região.

Posteriormente o estudo de caso “Influência da Localização” simula o *mesmo* imóvel em diferentes localizações, possibilitando isolar e compreender a real influência do território sobre o valor de mercado dos imóveis, suprimindo vieses de diferentes construções.

O gráfico da Figura 17 demonstra o preço médio do metro quadrado na amostra, para cada um dos bairros de Caxias do Sul. As barras em preto mostram a variação de \pm um desvio padrão em cada bairro, mostrando bairros com amostras mais consistentes (valores médios próximos) e bairros com amostras mais dispersas (maiores oscilações).

Figura 17 - Valor médio do m² – por bairro – na amostra de dados de mercado

Fonte: O Autor (2018).

Realizou-se uma comparação entre o valor médio do metro quadrado presente na amostra (Figura 17) com os cinco valores médios do metro quadrado avaliados e publicados no estudo ZAP-Viva Real (PIONEIRO, 2016), conforme mostra a Tabela 1:

Tabela 1 – Validação de dados do trabalho em relação à publicação ZAP-Viva Real

Bairro	Valor publicado ZAP	Média da amostra	Δ Variação
Exposição	R\$ 5.272,00 / m ²	R\$ 4.162,95 / m ²	-26,64%
Madureira	R\$ 4.246,00 / m ²	R\$ 4.286,90 / m ²	0,95%
Panazzolo	R\$ 4.209,00 / m ²	R\$ 4.284,02 / m ²	1,75%
N. Senhora de Lourdes	R\$ 4.094,00 / m ²	R\$ 4.140,84 / m ²	1,13%
São Pelegrino	R\$ 4.091,00 / m ²	R\$ 4.120,25 / m ²	0,71%

Fonte: O Autor (2018).

Em 80% dos casos (Madureira, Panazzolo, Lourdes e São Pelegrino), observam-se variações de 0,71% a 1,75%. Tendo em vista que os dados do trabalho (GIHAB-CX, 2018) são cerca de dois anos mais recentes que os dados da pesquisa ZAP-Vivareal (PIONEIRO, 2016), provavelmente a diferença represente a valorização dos imóveis neste período.

Observando esta elevada consistência, mesmo antes tratamentos de preços hedônicos, percebe-se uma alta qualidade da amostra de dados empregada. O bairro Exposição é a única exceção (variação de -26,64%). Supõem-se que existam ofertas de imóveis de maior valor, ou que imóveis de maior valor sejam menos financiados – situações que causam diferenças negativas.

4.2.2 SELEÇÃO DE VARIÁVEIS

Dentre as colunas de informação contidas na amostra de dados sobre os imóveis, selecionaram-se aquelas com informações mais relevantes, para otimizar o desenvolvimento do modelo. Os seguintes critérios foram avaliados conjuntamente para a seleção das variáveis:

- Variáveis presentes em modelos semelhantes;
- Correlação da variável com o valor de mercado da unidade;
- Disponibilidade de informações para a variável.

O Quadro 4 apresenta as variáveis selecionadas, contendo abrangência de menor e maior valor registrados, unidade mediana (centro da amostra), média e desvio padrão dos valores:

Quadro 4 – Caracterização de variáveis selecionadas

Descrição	Abrangência	Unidade mediana	Média	Desvio padrão
1 - Área privativa total da unidade (m ²)	23,73 m ² a 309,03 m ²	58,02 m ²	69,79 m ²	± 33,20 m ²
2 - Total de vagas cobertas	0 a 6 vagas	1 vaga	0,95 vagas	± 0,81
3 - Quarto(s) (dormitórios)	1 a 4 quartos	2 quartos	2,15 quartos	± 0,48
4 - Padrão de acabamento da unidade	1 - Mínimo a 6 - Alto	4 – Normal (forte predominância)	3,88	± 0,83
5 - Estado de conservação da unidade	1 – Regular/reparos 4 – Bom/novo	4 – Normal (forte predominância)	3,40	± 0,75
6 - Idade estimada da edificação como um todo	1 – Na planta a 5 – > 20 anos	2 - < 5 anos	2,41	± 0,90
7 - Padrão do empreendimento	1 - Mínimo a 7 – Superior/luxo	4 – Normal (forte predominância)	3,88	± 0,84
8 - Padrão de acabamento do prédio	1 - Mínimo a 7 – Superior/luxo	4 – Normal (forte predominância)	3,89	± 0,85
9 – Apts. por andar *Derivada de: número de unidades do prédio / número de pavimentos.	1 a 16 apartamentos por andar	3 apartamentos por andar	3,73	± 1,84
10 - Andar da unidade privativa	1º a 19º	3º	3,51	± 2,78
11 - Orientação solar da unidade	1 – Desfavorável a 3 – Favorável	2 – S/ influência	2,26	± 0,52
12 - Na planta (0/1) * Derivada de idade estimada.	0 – Não 1 – Sim	0 – Não	0,06	± 0,23
13 - Esquina (0/1) * Derivada de posição na quadra.	0 – Não 1 – Sim	0 – Não	0,22	± 0,42
14 - Cobertura (0/1)	0 – Não 1 – Sim	0 – Não	0,01	± 0,11

(continua)

(conclusão)

Descrição	Abrangência	Unidade mediana	Média	Desvio padrão
15 - Estado de Conservação do Empreendimento	1 – Regular/reparos 4 – Bom/novo	4 – Bom/novo	3,49	± 0,72
16 - Posição física da unidade privativa	1 – Fundos/Meio a 3 – Frente/Meio/ Canto	3 - Frente/Meio/ Canto	2,22	± 0,91
17 – Renda média mensal do chefe familiar / setor * Incorporada: IBGE	R\$702,15 a R\$8.013,16	R\$ 1.584,45	1.976,03	± 1.005,65
18 – Renda média mensal de familiares / setor * Incorporada: IBGE	R\$616,00 a R\$4.316,30	R\$ 1.125,07	1.333,56	± 606,87
19 – Bairro / Defasada espacialmente * Incorporada: IBGE	Variável textual – nome do bairro			

Fonte: O Autor (2018).

Selecionaram-se primeiramente as 16 variáveis obtidas e derivadas no banco de dados de mercado (GIHAB-CX, 2018). Em seguida, cada ponto de dado foi unido com seu setor censitário (IBGE, 2010), por meio do software de geoprocessamento QGIS (2018), acrescentando as variáveis 17/18/19 para as análises espaciais posteriores.

4.2.3 CORRELAÇÃO ENTRE VARIÁVEIS

A análise de correlações permite quantificar a influência presente entre as variáveis analisadas, compreender e interpretar seus efeitos. A Figura 18 demonstra o coeficiente de correlação de Pearson (ρ) existente entre cada uma das 18 variáveis utilizadas no modelo.

Figura 18 - Correlações entre variáveis

	01 Área privativa m ²	02 Vagas cobertas	03 Quarto(s)	04 Padrão unidade	05 Conservação und	06. Idade estimada	07 Padrão emp.	08 Padrão prédio	09 Aptos andar	10 Andar da UP	11 Orientação solar	12 Na planta	13 Esquina	14 Cobertura	15 Conserv. emp.	16 Posição física	17 Renda - chefe	18 Renda - família	Valor R\$	
01 Área privativa m ² unid	1,00																			
02 Vagas cobertas	0,59	1,00																		
03 Quarto(s)	0,62	0,42	1,00																	
04 Padrão acab. unidade	0,53	0,51	0,33	1,00																
05 Conservação da unidade	0,11	0,21	0,01	0,29	1,00															
06. Idade estimada	0,04	-0,09	0,07	-0,20	-0,59	1,00														
07 Padrão empreendimento	0,55	0,52	0,35	0,92	0,29	-0,20	1,00													
08 Padrão acab. prédio	0,54	0,51	0,35	0,93	0,29	-0,21	0,97	1,00												
09 Apartamentos por andar	-0,25	-0,23	-0,15	-0,24	-0,22	0,09	-0,23	-0,23	1,00											
10 Andar da UP	0,20	0,23	0,21	0,28	0,04	-0,05	0,30	0,29	-0,03	1,00										
11 Orientação solar	0,11	0,10	0,10	0,10	0,01	0,03	0,11	0,12	-0,04	0,04	1,00									
12 Na planta	0,13	-0,07	-0,02	0,14	0,20	-0,37	0,12	0,12	-0,05	0,07	-0,11	1,00								
13 Esquina	0,05	0,06	0,04	0,04	-0,10	0,00	0,05	0,05	0,06	0,09	0,06	0,05	1,00							
14 Cobertura	0,17	0,08	0,08	0,06	0,02	-0,02	0,05	0,05	-0,05	0,09	0,02	0,04	-0,05	1,00						
15 Conservação empreend.	0,12	0,19	0,01	0,28	0,76	-0,52	0,29	0,29	-0,21	0,06	0,01	0,17	-0,07	-0,01	1,00					
16 Posição física da UP	0,10	0,11	0,08	0,00	0,03	0,00	0,03	0,01	-0,03	-0,06	0,05	-0,01	0,05	-0,01	0,03	1,00				
17 Renda familiar - chefe	0,48	0,39	0,30	0,39	-0,02	0,18	0,40	0,41	-0,13	0,25	0,08	0,06	0,07	0,01	0,01	0,01	1,00			
18 Renda familiar - completa	0,50	0,38	0,31	0,39	-0,03	0,24	0,41	0,41	-0,13	0,23	0,08	0,05	0,05	0,02	-0,01	0,01	0,97	1,00		
Valor R\$	0,85	0,65	0,56	0,62	0,14	-0,05	0,64	0,64	-0,23	0,31	0,12	0,15	0,08	0,09	0,14	0,08	0,57	0,57	1,00	

Fonte: O Autor (2018).

Dentre as correlações, observa-se que:

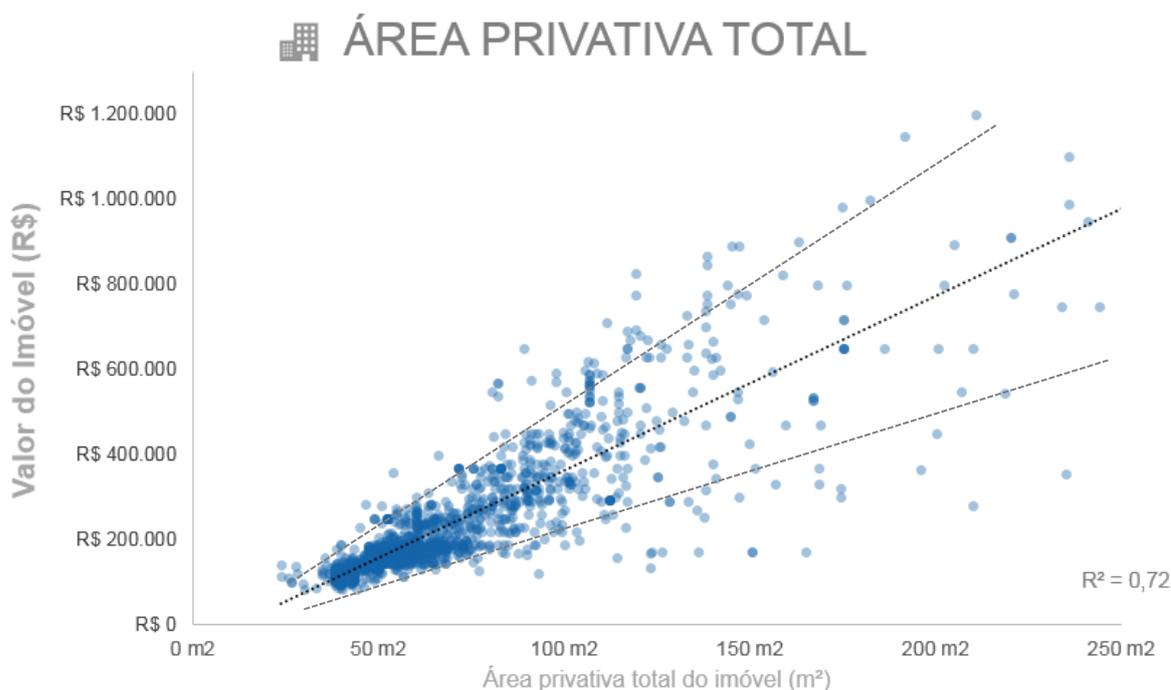
- A área privativa possui uma correlação forte ($\rho=0,85$) com o valor de mercado, sendo uma variável com elevado poder explicativo: 85% da variação do valor de mercado nos imóveis da amostra acontece em conjunto com a variação de sua área.
- Dentre as variáveis que também possuem correlações com o valor de mercado: número de vagas cobertas (correlação forte – $\rho=0,65$), padrão de acabamento da unidade (correlação forte – $\rho=0,62$) e número de quartos (correlação média – $\rho=0,56$).
- Observam-se também correlações inversas – quanto maior a idade estimada do imóvel, menor o estado de conservação desde imóvel (correlação média – $\rho=-0,59$).
- Imóveis maiores possuem mais quartos (correlação forte – $\rho=0,62$), mais vagas cobertas (correlação média – $\rho=0,59$) e melhores padrões de acabamento (correlação média – $\rho=0,53$).
- A renda das famílias de um setor censitário (variável externa incorporada aos dados) possui uma correlação média com o valor de mercado dos imóveis naquele setor censitário ($\rho=0,57$). A correlação entre a renda e área privativa ($\rho=0,50$) indica que localizações com renda maior tendem a possuir imóveis maiores. Estes comportamentos socioeconômicos refletidos podem ser incorporados ao modelo preditivo.

4.2.4 GRÁFICOS DE DISPERSÃO

Os gráficos de dispersão permitem verificar a relação existente entre duas variáveis. Pode-se explorar a relação entre atributos individuais e valores de mercado dos imóveis. Relações fortes entre os atributos e valores (indicadas por meio de coeficientes de determinação múltiplos R^2 mais elevados) são indícios de que é possível construir bons modelo preditivos.

A Figura 19 apresenta a relação entre os valores de mercado dos imóveis e suas áreas privativas totais. O coeficiente de determinação múltiplo ($R^2=0,72$) indica que 72% da variação do valor de mercado dos imóveis pode ser explicada unicamente por meio de sua área privativa. Portanto, a “Área privativa total” é uma variável com elevado poder explicativo.

Figura 19 - Gráfico de dispersão entre valor do imóvel e área privativa

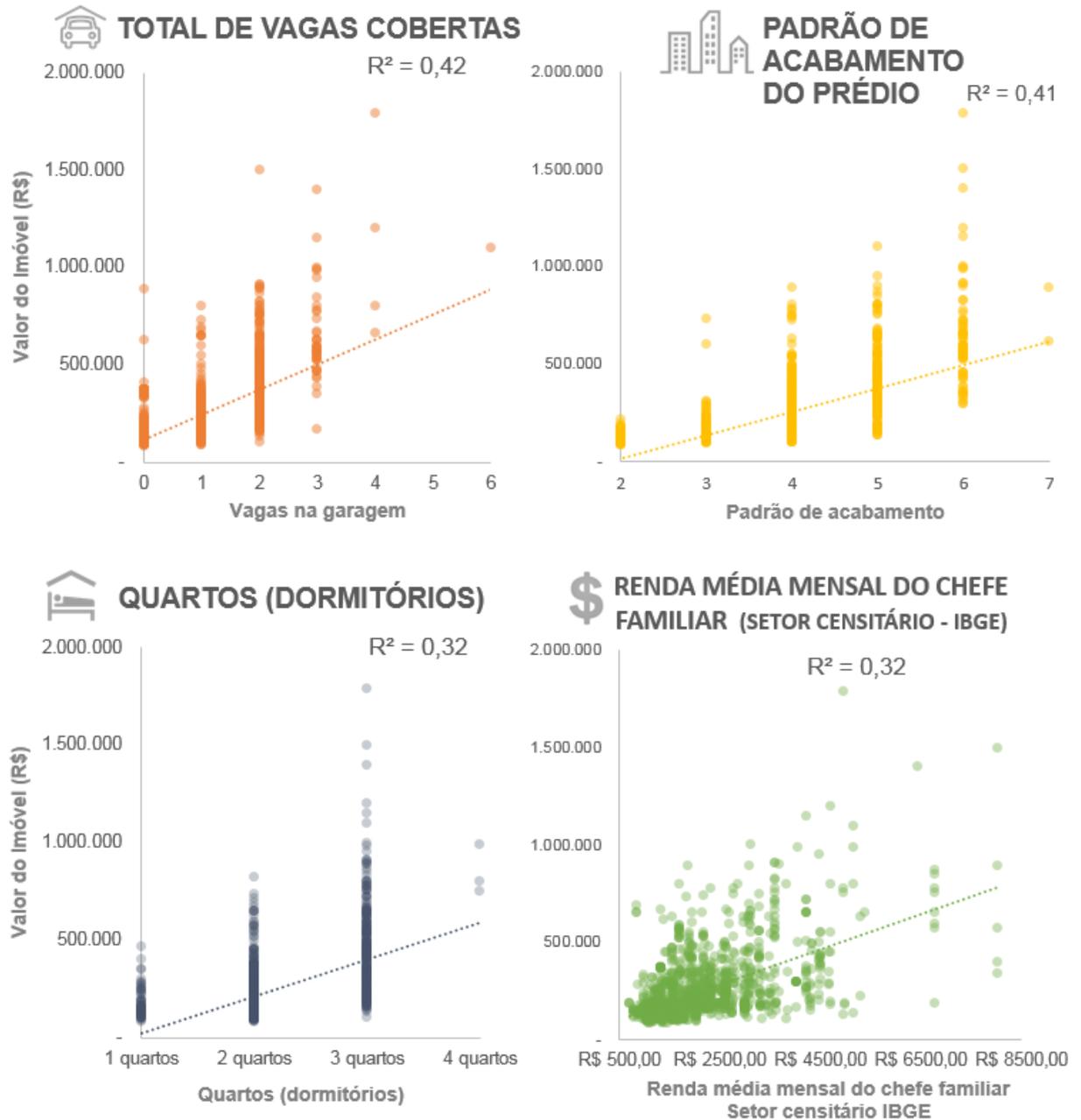


Fonte: O Autor (2018).

Conforme demonstram as duas linhas paralela à regressão, a dispersão da área privativa exhibe um padrão de heterocedasticidade (GONZÁLEZ, 1999) com variância crescente. Este comportamento indica que os imóveis com áreas menores são relativamente parecidos (menor variância nos resíduos); enquanto os imóveis maiores possuem ainda outros fatores importantes a serem contabilizados (maior variância nos resíduos) conjuntamente no modelo.

A Figura 20 mostra gráficos de dispersão de atributos em relação ao valor de mercado do imóvel, apresentando outras variáveis explicativas empregadas no modelo:

Figura 20 - Gráficos de dispersão de atributos e valor de mercado



Fonte: O Autor (2018).

Mesmo contendo coeficientes de correlação múltipla menores, o conjunto de variáveis que se complementam alavanca o poder explicativo do modelo, semelhante ao caso dos modelos compostos (*ensemble models*) (PROVOST & FAWCETT, 2013). Unido, o conjunto de variáveis produz um modelo completo maior do que a soma de suas partes.

4.3 MODELO PREDITIVO

Seguindo o conceito da defasagem espacial (DANTAS, 2003), inicialmente modelam-se os dados por meio da regressão linear múltipla, omitindo as variáveis espaciais. Então, verifica-se a existência da dependência espacial com a presença da autocorrelação espacial dos resíduos, e caso necessário, acrescentam-se as variáveis espaciais.

4.3.1 REGRESSÃO LINEAR MÚLTIPLA SIMPLES

O modelo é composto por três partes, conforme conceitua a Figura 21. Os dados de entrada de imóveis são organizados em uma matriz: cada linha representa um imóvel e cada coluna um atributo. O valor de mercado é o alvo da otimização. As variáveis livres do modelo são os fatores, ajustados com o critério da minimização da soma quadrática dos resíduos.

Figura 21 - Desenvolvimento do modelo de regressão linear múltipla

$$\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

DADOS OBSERVADOS
FATORES
RESÍDUOS
VALORES

▲
▲
▲

DADOS DE ENTRADA
VARIÁVEIS LIVRES
MINIMIZAR

Σ (RESÍDUOS²)

Fonte: Adaptação de FOREMAN (2014) e HOFFMANN (2016).

O modelo foi construído no software Microsoft Excel. Executou-se a resolução por meio do pacote *Solver*. O algoritmo empregado foi o *GRC Non-Linear*, indicado no software para resolver problemas não lineares (quadrático, no caso) com variações relativamente contínuas.

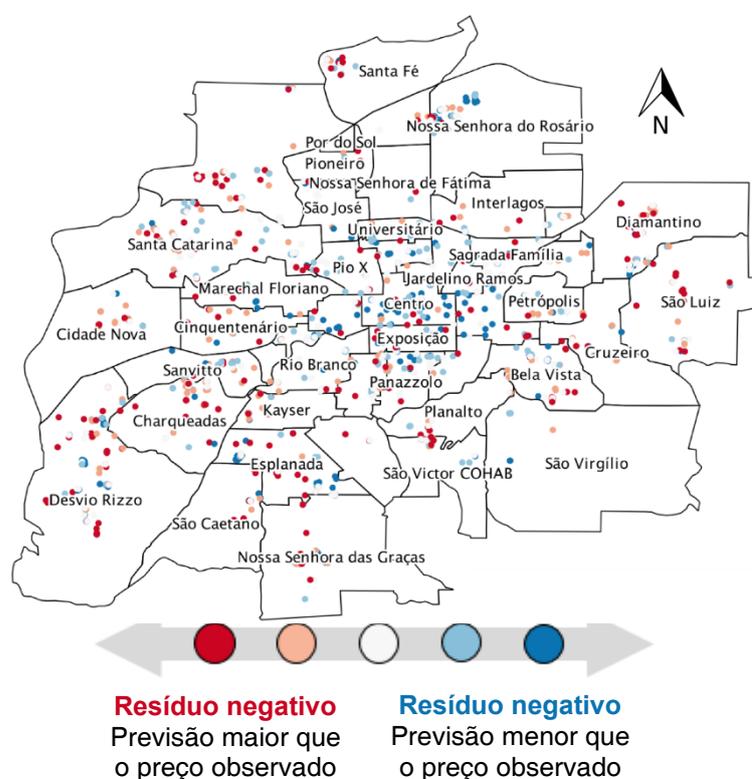
Nesta resolução, desconsiderando efeitos espaciais, obteve-se o coeficiente de determinação múltiplo $R^2 = 0,82$. Procede-se analisando a autocorrelação espacial nos resíduos.

Análise da presença de autocorrelação espacial nos resíduos

Espera-se que regressão linear múltipla omitindo informações espaciais possibilite um entendimento sobre a influência exercida pelo território no valor dos dados espaciais.

O mapa apresentado na Figura 22 demonstra a distribuição espacial dos resíduos da regressão. Cada ponto representa um imóvel. A escala vermelha mostra resíduos negativos, ou seja, quando o modelo previu um valor maior que o valor observado para aquele imóvel. A escala azul mostra resíduos positivos, onde valor previsto foi menor que o valor observado.

Figura 22 - Distribuição espacial dos resíduos da regressão

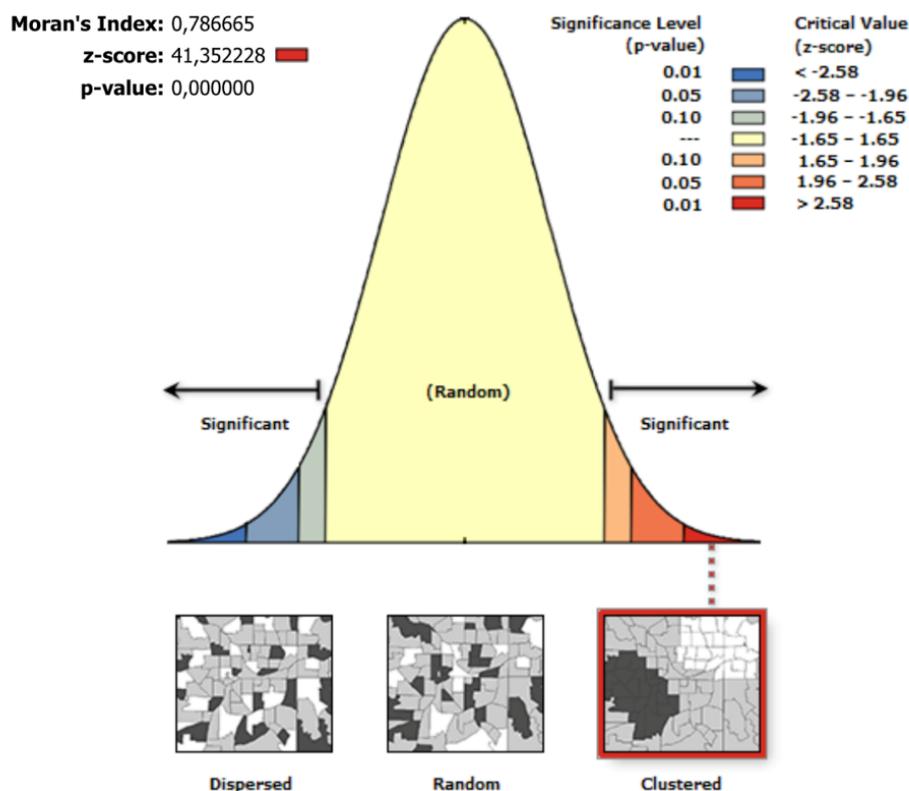


Fonte: O Autor (2018).

Caso não exista influência espacial, os resíduos positivos e negativos devem estar distribuídos aleatoriamente no espaço. Percebe-se na Figura 22 que os pontos possuem uma tendência ao agrupamento – pontos vermelhos e azuis se encontram próximos uns dos outros.

A hipótese a ser verificada é que os pontos se encontram clusterizados, ou agrupados. Assim, entende-se que o território valoriza ou desvaloriza o valor de mercado de certas regiões. Executou-se o teste de Moran I por meio do software ArcGIS (ESRI, 2011), para verificar esta hipótese. Os resultados são demonstrados na Figura 23:

Figura 23 - Resultados de Moran I



Global Moran's I Summary

Moran's Index:	0,786665
Expected Index:	-0,000588
Variance:	0,000362
z-score:	41,352228
p-value:	0,000000

Dataset Information

Input Feature Class:	reg2
Input Field:	DF
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN

Fonte: O Autor (2018).

O índice de Moran I de 0,78 (Figura 23) indica que os dados espaciais (Figura 22) se encontram agrupados (ou clusterizados). O valor de p indica $\approx 100\%$ de significância, indicando a impossibilidade deste agrupamento acontecer por mera casualidade estatística.

Portanto, confirma-se que o território influencia significativamente o valor de mercado dos imóveis. Sendo assim, a maneira mais adequada de proceder com a composição dos preços hedônicos é incluir componentes que capturem estes comportamentos espaciais.

4.3.2 REGRESSÃO LINEAR MÚLTIPLA COM VARIÁVEIS ESPACIAIS

Demonstrada a presença da autocorrelação espacial, buscou-se contabilizar os efeitos espaciais no modelo preditivo. Este trabalho incluiu variáveis *proxies* espaciais, visando demonstrar os propósitos da análise espacial. Os *proxies* são uma forma simples (em relação à densidade da regressão espacial) e efetiva para capturar significativamente os efeitos espaciais.

Acrescentaram-se as variáveis, procedendo então com uma nova resolução do modelo:

- Variável *proxy* de renda – Corresponde ao acréscimo, via setores censitários, da renda média das famílias e renda média do chefe familiar na localização geográfica onde o imóvel se encontra. Presume-se que a renda média seja *proxy* para outras variáveis não consideradas explicitamente, tais como amenidades públicas, entornos e *status* de uma determinada localização.
- Variável defasada espacialmente de bairros – Corresponde a uma matriz contendo uma variável livre para cada bairro presente na amostra. Objetiva-se com isso corrigir parte da discrepância entre renda e valores de imóveis, abrangendo elementos tais como a evolução da cidade em relação ao Censo de 2010 e eventuais tendências de ascensão e declínio no mercado imobiliário em diferentes bairros. Estabeleceu-se um mínimo de cinco dados por bairros, visando reduzir a possibilidade de sobreajustes no modelo.

Resultados

O modelo contendo as variáveis espaciais resultou em um coeficiente de determinação múltiplo de $R^2 = 0,87$, indicando que este modelo consegue explicar 87% da variação dos dados. Este resultado do modelo encontra-se consistente com as análises do mercado imobiliário, cujos coeficientes de determinação normalmente resultam entre 0,65 e 0,95 (GONZÁLEZ, 1997). As variáveis espaciais incluíram a possibilidade para o modelo analisar distintos pontos do território, acrescentando também 5% de poder explicativo aos dados de mercado.

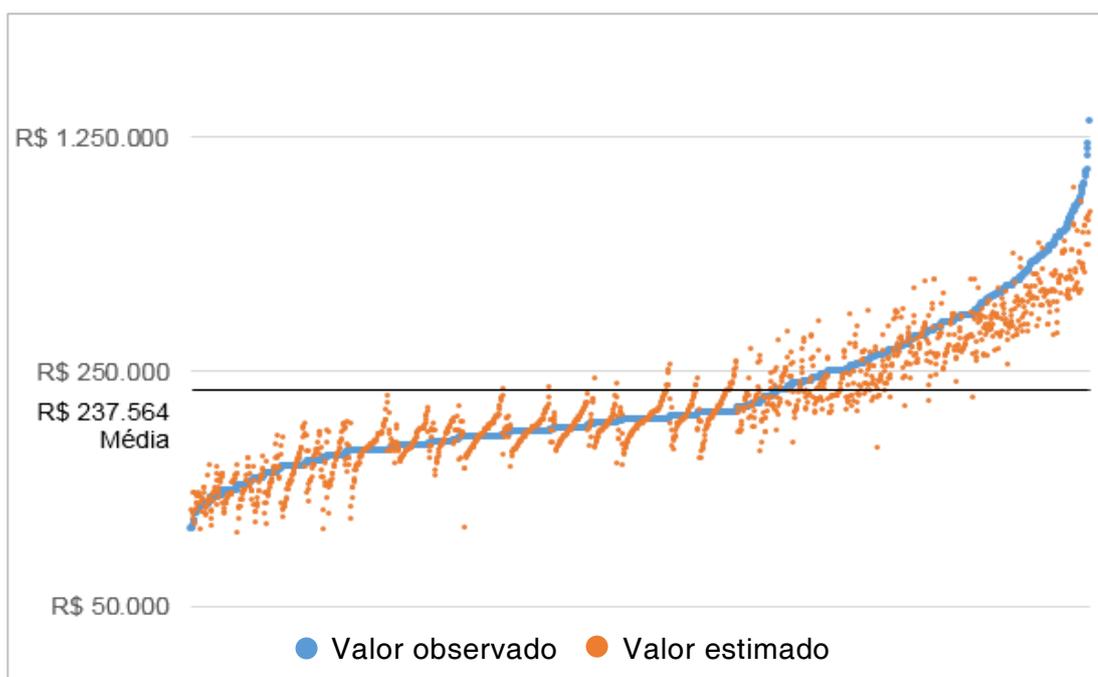
O intervalo de confiança para 50% dos casos estimados é de $\pm 10,2\%$. Este intervalo indica que em 50% dos casos o modelo estima o valor de mercado dentro de uma precisão de $\pm 10,2\%$. Como comparativo (*benchmark*), a empresa americana Zillow possui uma precisão de $\pm 5\%$ em 50% de suas estimativas de valor de mercado.

No intervalo de confiança de 80% dos casos, o modelo possui uma precisão de $\pm 17,4\%$. Como comparativo, a norma de avaliações imobiliárias NBR 14653-1:2001 (ABNT, 2001). estabelece para 80% dos casos o limite de $\pm 15\%$ (ou somatória de 30% para as duas caudas), para a avaliação imobiliária mais simples – Grau I – realizada pelo avaliador.

O teste de significância F resultou no coeficiente de 380, com valor $p \cong 0$. Desta forma, rejeitou-se a hipótese nula de que a distribuição dos dados poderia ter ocorrido por acaso, mostrando que o modelo é matematicamente significativo.

A Figura 24 mostra a comparação entre os valores de mercado observados nos imóveis (em azul) e os valores estimados pelo modelo (em laranja), bem como a média observada dos dados. A correlação entre estes valores é extremamente forte ($\rho=0,925$).

Figura 24 - Comparação entre valores observados e valores estimados



Fonte: O Autor (2018).

Percebe-se a tendência do modelo em estimar com melhor precisão os valores centrais. Este fenômeno pode ocorrer devido à grande heterogeneidade dos dados presentes, supostamente formando *clusters* de mercado, com diferentes comportamentos e valorização de unidades. A tendência em estimar as unidades de maior valor quase sempre abaixo do seu valor observado pode ocorrer pelo fato de que estas unidades (muitas superiores a R\$1.000.000,00) possuírem amenidades especiais (alto padrão) não diretamente contabilizados nas variáveis.

4.3.3 ANÁLISE DOS COEFICIENTES

Pode-se compreender a utilidade econômica determinada para cada atributo do imóvel, por meio dos coeficientes encontrados para a solução do modelo – que são os preços implícitos.

Dentre os preços implícitos, citam-se:

- O acréscimo de cada metro quadrado extra na unidade representa um acréscimo de R\$2.384,75 em seu valor de mercado.
- Uma vaga coberta extra acrescenta R\$14.633,76 no valor de mercado da unidade.
- Um quarto extra representa um acréscimo de R\$7.056,54 no valor de mercado.
- Cada andar superior em que o apartamento se localiza em relação ao prédio representa um acréscimo de R\$1.929,10 em seu valor de mercado.
- Cada nível de idade estimada (na planta; até 2 anos; 2 a 5 anos; 5 a 20 anos, maior que 20 anos) reduz em R\$4.908,96 o valor de mercado da unidade.
- O padrão divide-se entre a unidade, o acabamento do empreendimento e acabamento da unidade. Existem 7 categorias (desde mínimo até superior/luxo). Cada categoria acrescida simultaneamente acresce mais de R\$8.200,00 ao valor de mercado da unidade.
- A valorização da localização é explicada em média 53,2% pela renda do setor censitário, e 46,8% pela variável defasada espacialmente “bairro”. O “Estudo de Caso 1 – Influência da Localização” demonstra detalhes sobre a valorização da localização.

A Tabela 2 lista os coeficientes completos encontrados para a resolução do modelo. A Tabela 3 lista os valores da variável β_{100} . Posteriormente, o Capítulo 5 - “Estudo de Caso” emprega estes coeficientes do modelo preditivo para simular situações reais e responder questões específicas na área em Caxias do Sul – RS.

Tabela 2 – Coeficientes do modelo

Coeficiente	Valor	Variável Associada	Abrangência
β_0	-47080,462	<i>Intercepto</i>	-
β_1	2384,752	Área privativa total da unidade (m ²)	23,73 m ² a 309,03 m ²
β_2	14633,761	Total de vagas cobertas	0 a 6 vagas
β_3	7056,549	Quarto(s) (dormitórios)	1 a 4 quartos
β_4	2194,198	Padrão de acabamento da unidade	1 - Mínimo a 6 - Alto
β_5	-704,670	Estado de conservação da unidade	1 – Regular/reparos 4 – Bom/novo
β_6	-4908,964	Idade estimada da edificação como um todo	1 – Na planta a 5 – > 20 anos
β_7	1298,051	Padrão do empreendimento	1 - Mínimo a 7 – Superior/luxo
β_8	4708,890	Padrão de acabamento do prédio	1 - Mínimo a 7 – Superior/luxo
β_9	102,265	Aptos. por andar	1 a 16 apartamentos por andar
β_{10}	1929,106	Andar da unidade privativa	1º a 19º
β_{11}	1362,620	Orientação solar da unidade	1 – Desfavorável a 3 – Favorável
β_{12}	8435,191	Na planta	0 – Não 1 – Sim
β_{13}	929,925	Esquina	0 – Não 1 – Sim
β_{14}	4403,749	Cobertura	0 – Não 1 – Sim
β_{15}	2378,155	Estado de Conservação do Empreendimento	1 – Regular/reparos 4 – Bom/novo
β_{16}	-1468,257	Posição física da unidade privativa	1 – Fundos/Meio a 3 – Frente/Meio/Canto
β_{100}	56,032	Bairro / Defasada espacialmente	-151,45 a 1398,49
β_{101}	4,138	Renda média mensal do chefe familiar / setor censitário (IBGE)	R\$702,15 a R\$8.013,16
β_{102}	6,962	Renda média mensal dos membros da família / setor censitário (IBGE)	R\$616,00 a R\$4.316,30

Fonte: O Autor (2018).

Tabela 3 – Variável defasada espacialmente - Bairro

Valor	Bairro	Valor	Bairro
148,8461	Ana Rech	857,668	Nossa Senhora de Lourdes
271,8644	Bela Vista	307,5445	Nossa Senhora do Rosário
108,7653	Centenário	1398,488	Panazzolo
697,3368	Centro	667,836	Petrópolis
176,1255	Charqueadas	883,1436	Pio X
286,9211	Cidade Nova	31,6178	Pioneiro
41,0337	Cinquentenário	579,1143	Planalto
920,3152	Cristo Redentor	529,9278	Presidente Vargas
280,7226	Cruzeiro	433,0614	Rio Branco
406,7928	de Lazzer	546,4392	Sagrada Família
226,3183	Desvio Rizzo	-20,0972	Salgado Filho
259,9288	Diamantino	309,3987	Santa Catarina
250,5232	Esplanada	194,6652	Santa Corona
523,7372	Exposição	101,1063	Santa Fé
-151,545	Floresta	312,3017	Santo Antônio
227,6162	Forqueta	1000,973	Sanvitto
151,3126	Interlagos	-18,2864	São Caetano
363,338	Jardelino Ramos	176,5644	São Cristóvão
827,265	Jardim América	156,767	São Giacomo
124,9017	Jardim Eldorado	1433,384	São José
61,12634	Kayser	1067,811	São Leopoldo
1201,416	Madureira	216,3579	São Luiz
329,4322	Marechal Floriano	56,57424	São Peregrino
542,0193	Medianeira	238,4911	São Victor COHAB
118,4542	Nossa Senhora da Saúde	723,5586	São Virgílio
196,2742	Nossa Senhora das Graças	382,7637	Serrano
501,5225	Nossa Senhora de Fátima	1991,549	Universitário

Fonte: O Autor (2018).

CAPÍTULO 5 – ESTUDO DE CASO

Em algum lugar, alguma coisa incrível está esperando para ser descoberta.

Carl Sagan

INTRODUÇÃO

O estudo de caso simula uma aplicação do modelo preditivo em situações próximas das reais, a fim de responder questões específicas em uma região da cidade de Caxias do Sul – RS.

Além de demonstrar a utilização do modelo preditivo, a finalidade do estudo de caso é proporcionar uma melhor compreensão de aspectos do mercado imobiliário nesta região. Para isso, selecionam-se componentes com possíveis aplicações em situações práticas.

Os objetivos destes componentes incluem, mas não se limitam a:

- Auxiliar construtoras e incorporadoras na tomada de decisões mais assertivas quanto à viabilidade econômica e maximização do retorno financeiro dos empreendimentos;
- Proporcionar a consumidores um maior conhecimento sobre o mercado imobiliário, bem como fornecer uma perspectiva econômica sobre diferentes unidades e bairros.

Propõem-se três componentes individuais para o estudo de caso: influência da localização; simulação de valores para unidades padrão; análise de valores e urbanismo.

O Quadro 5 descreve-se cada um destes componentes:

Quadro 5 – Componentes para o estudo de caso

Estudo de Caso	Conteúdo
Influência da localização	Mapa de calor indicando localizações com maior e menor valorização imobiliária, com delimitação de bairros e vias.
Simulação de valores para unidades padrão	Estimativa de valor de unidades medianas próximas a pontos de referência na cidade, utilizando preços hedônicos.
Análise de valores e urbanismo	Análise do impacto de aspectos urbanos em relação aos valores de mercado dos imóveis da área.

Fonte: O Autor (2018)

5.1 INFLUÊNCIA DA LOCALIZAÇÃO

O território exerce uma enorme influência sobre os imóveis. Este primeiro estudo de caso avalia a distribuição desta influência e quantificar o seu impacto. Determinou-se o formato de “mapa de calor” buscando a melhor representação das nuances espaciais.

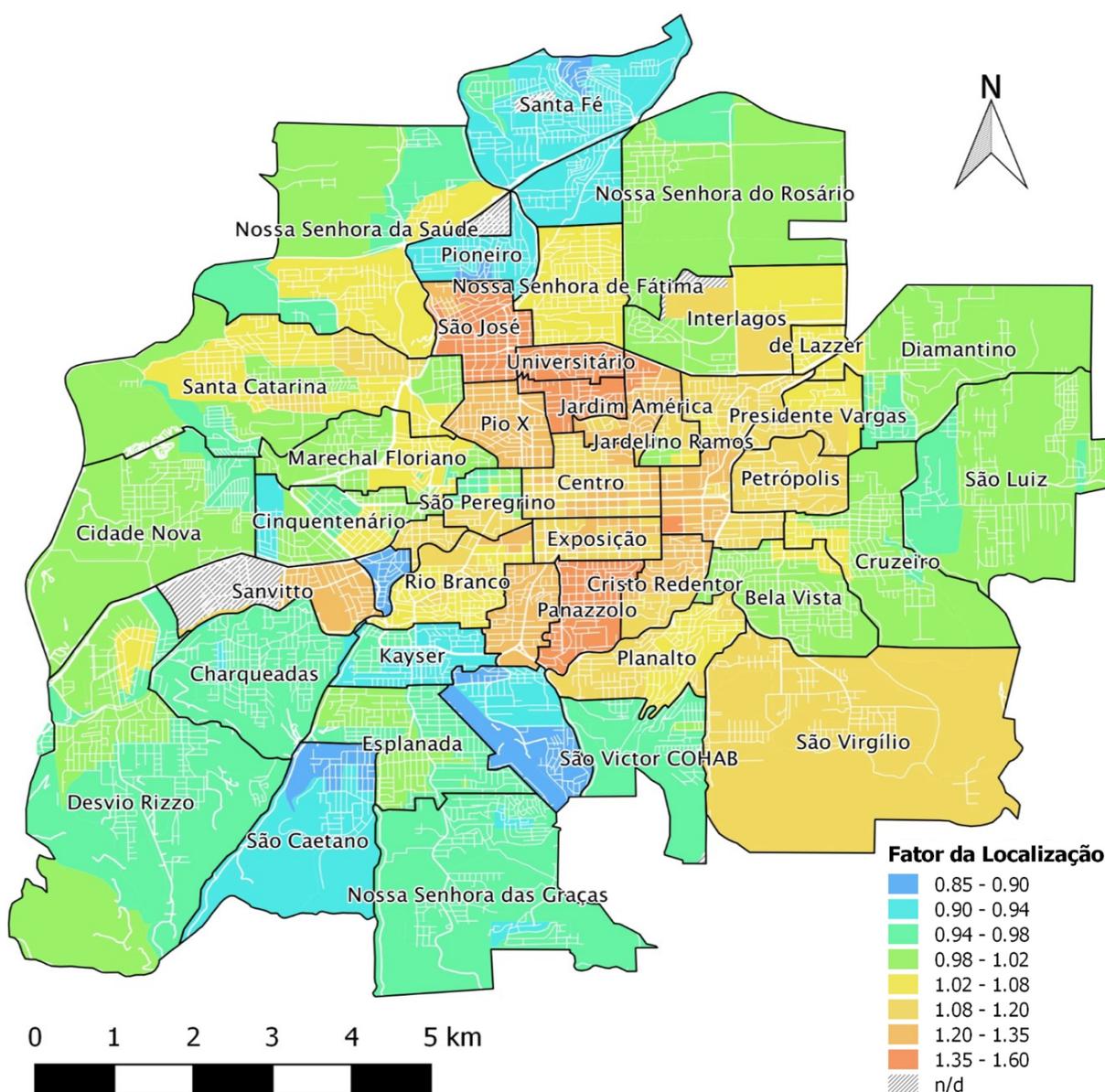
Inicialmente, selecionou-se uma configuração de imóvel com as características medianas da amostra, conforme descrito anteriormente na coluna “unidade mediana” no Quadro 4. Então, simulou-se o valor de mercado desta unidade em cada setor censitário (a menor agregação geográfica disponível).

O valor mediano deste imóvel nos setores censitários foi de R\$181.767,40. Então, todos os valores foram normatizados (divididos por esta base), para corresponder a um multiplicador, estabelecendo assim uma base unitária para comparação.

Por meio do software de geoprocessamento QGIS (2018), estes multiplicadores foram plotados em um mapa de calor, representando a variação de valores de mercado de um mesmo imóvel para diferentes localizações da cidade. O mapa foi complementado com os perímetros dos bairros, derivados dos setores censitários do IBGE (2010) e com o sistema viário urbano, acrescido a partir de da base de mapas abertas Open Street Maps (2018).

A Figura 25 e o Apêndice A apresentam o mapa de calor com os valores de mercado dos imóveis, onde as regiões menos valorizadas estão representadas em tons azuis, regiões medianas em verde e regiões mais valorizadas em tons vermelhos. A escala simboliza o fator “multiplicador” do valor de um imóvel naquela região.

Figura 25 - Mapa dos valores de mercado da unidade por localização



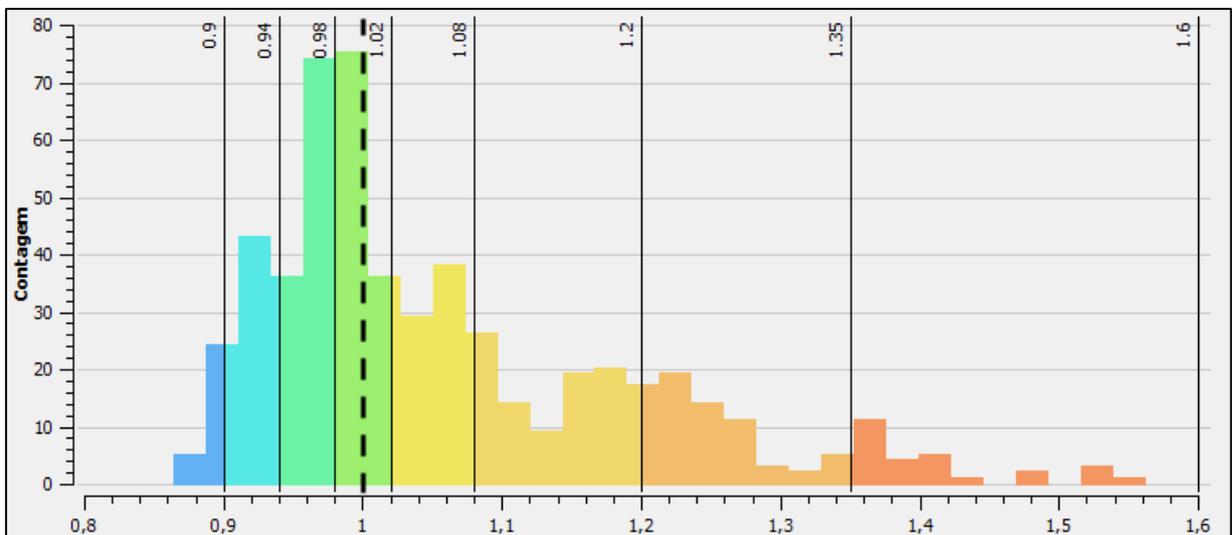
Fonte: O Autor (2018).

De acordo com o mapa, um mesmo imóvel com valor de mercado de R\$181.767,40 na região verde (1x), possui um valor de mercado de R\$154.502,29 em regiões mais azuis (0,85x) ou R\$290.827,84 em regiões mais vermelhas (1,60x). Assim, a localização acresce em até 88% o valor de mercado de um imóvel. Na composição total do valor, a parcela da localização pode representar até 45% do valor de mercado do imóvel.

Os intervalos no mapa foram estabelecidos utilizando o algoritmo de quebras naturais (*jenks*), que busca o melhor agrupamento dos valores, minimizando a variância de cada classe em relação à sua média e maximizando a variância em relação à média das outras classes. Realizou-se um ajuste para manter a classe verde em torno do valor unitário (0,98 – 1,02).

A Figura 26 demonstra o histograma da distribuição e intervalos estabelecidos do fator multiplicador da localização. Percebe-se uma tendência de distribuição normal da valorização em torno da mediana; em simultâneo à tendência de *power law* – na qual uma pequena parcela das localizações se valoriza de maneira excepcional em relação à mediana.

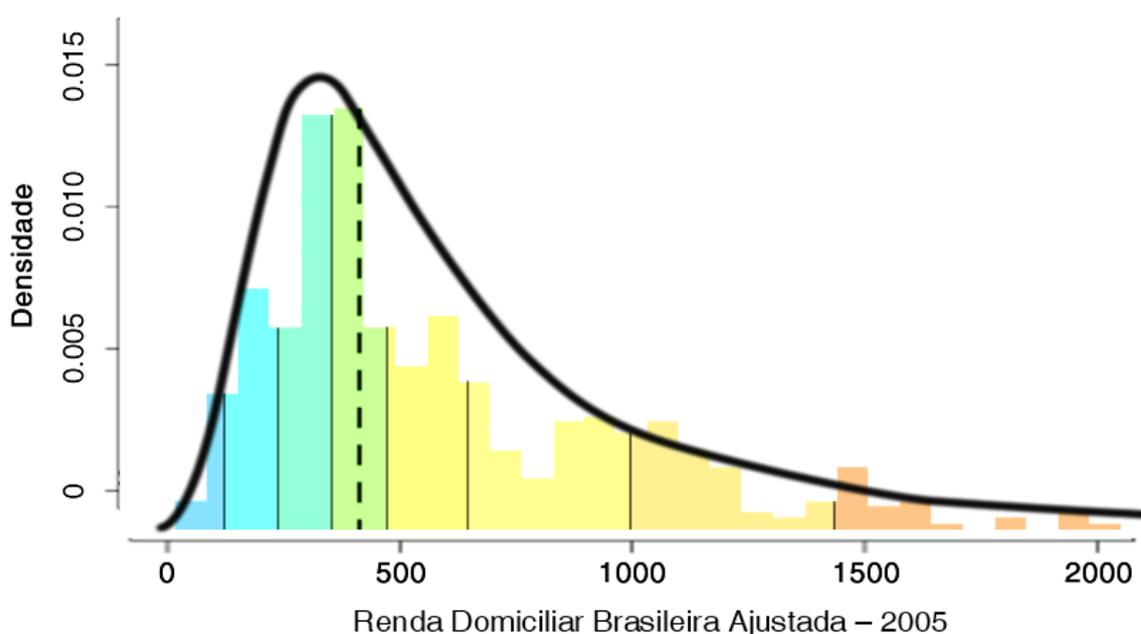
Figura 26 - Distribuição de valores de mercado para imóveis por localização



Fonte: O Autor (2018).

Curiosamente, o comportamento de valorização do território (onde uma pequena parcela dos espaços possui valores excepcionalmente acima da média) é muito semelhante à distribuição de renda entre a população brasileira, conforme mostra a Figura 27. A semelhança entre ambas as distribuições pode indicar que o comportamento de valorização do território imita o padrão socioeconômico da população.

Figura 27 – “Distribuição de renda da população brasileira” x
“Distribuição de valores no mercado imobiliário”



Fonte: FIGUEREDO & ZIEGELMANN (2009), ajustado sobre a Figura 26 pelo Autor (2018)

Percebe-se que o mantra do mercado imobiliário “*location, location, location*” continua verdadeiro e relevante. A localização é parte essencial na composição do valor de mercado de imóveis, e um componente indispensável para estimar precisamente seus valores.

5.2 SIMULAÇÃO DE VALORES PARA UNIDADES PADRÃO

Este segundo estudo de caso busca demonstrar a aplicação do modelo para estimar valores de mercado em diferentes unidades imobiliárias. Foram escolhidas três configurações para atributos de imóveis, e cada uma presente em três diferentes localizações da cidade. Assim, realizaram-se nove estimativas de valores de mercado.

Conforme mostra o mapa da Figura 28, as localizações selecionadas na cidade de Caxias do Sul - RS estão próximas a pontos de referência conhecidos, em regiões de comportamentos imobiliários distintos:

- 1 – No bairro Marechal Floriano, Rua Prof. Marcos Martini, próximo ao Hospital Unimed.
- 2 – No bairro Nossa Senhora de Lourdes, Rua Conselheiro Dantas, próximo à UCS.
- 3 – No bairro Panazollo, Rua Tronca, próximo à Prefeitura Municipal de Caxias do Sul.

Figura 28 - Pontos selecionados para simulação do valor de mercado



Fonte: O Autor (2018).

As unidades foram selecionadas com base nas unidades medianas e quartis da amostra. Os quartis representam os dados de maneiras mais fiéis que a média simples dos valores, especialmente em amostras heterogêneas na presença de diferentes desvios padrão e distribuições. O quartil inferior descreve a posição 25% da amostra; o mediano sua metade e o superior a posição 75% da amostra. Enquanto a média pode sofrer vieses de valores muito altos ou baixos, os quartis representam posições. Por exemplo, o quartil mediano encontra-se na posição onde metade dos valores são mais altos e metade mais baixos que o mesmo.

Selecionaram-se três unidades (\$, \$\$ e \$\$\$) para estimar seus valores de mercado. A unidade mediana (\$\$) é a mesma especificada anteriormente na coluna “unidade mediana” do Quadro 4. Todas as unidades possuem 2 quartos, e cada uma possui os seguintes atributos:

\$. Apartamento de 41,90m². Padrão construtivo normal, com aspectos de baixo. Sem garagens. Idade entre 5 e 10 anos, com aparência de usado e localizado no 2º andar.

As unidades mediana e superior (\$\$ e \$\$\$) possuem padrão construtivo normal, uma vaga na garagem, idade estimada em até 5 anos e aparência de novas –

\$\$. Apartamento de 49,95m², localizado no 3º andar.

\$\$\$. Apartamento de 80,50m², localizado no 4º andar, com boa orientação solar.

Realizou-se a simulação de valores para cada uma das três unidades em três pontos, conforme a Figura 29. As colunas mostram o mesmo apartamento localizado em três diferentes pontos, e as linhas, a mesma localização com três diferentes configurações de apartamentos.

Figura 29 - Simulação de valores pra unidades e localizações

PONTO	REF.	\$ INFERIOR	\$\$ MEDIANA	\$\$\$ SUPERIOR
1		R\$ 140.268,41 -59%	R\$ 187.821,04 -19%	R\$ 244.875,27 +9%
2		R\$ 175.273,38 -27%	R\$ 222.826,01 referência	R\$ 279.880,23 +20%
3		R\$ 224.966,08 +1%	R\$ 272.518,71 +18%	R\$ 329.572,93 +32%

● < R\$200.000 ● R\$200.000 – R\$250.000 ● > R\$250.000

Fonte: O Autor (2018).

Evidencia-se novamente a importância da localização, e a distinta valoração dos atributos. Resultados como estes podem auxiliar na tomada de decisões, especialmente naquelas onde os agentes precisam calcular *trade-offs*, riscos e retornos.

5.3 ANÁLISE DE VALORES E URBANISMO

Uma cidade é um sistema complexo: todos os seus elementos funcionam em conjunto. Este estudo de caso quantifica a relação existente entre componentes distintos: o valor de mercado dos imóveis e aspectos urbanísticos na cidade.

Os indicadores territoriais sintetizam múltiplos aspectos urbanísticos. Estes indicadores são calculados sobre os mapas, e por serem numéricos, permitem comparações, cruzamentos de informações e diagnósticos abrangentes, imparciais e assertivos para a tomada de decisões.

Nesta análise, selecionaram-se dois indicadores para estudo: complexidade urbana (RUEDA, 2009) e segregação espacial (HERMIDA, 2015), conforme mostra a Figura 30. Ambos fizeram parte de um diagnóstico urbano contendo dezenas de indicadores, com objetivo de propor diretrizes ao plano diretor da cidade de Lajeado – RS (BIOSSPLENA, 2017).

Figura 30 - Indicadores territoriais selecionados



Fonte: Adaptado de HERMIDA et al (2015).

Em seguida, estes indicadores foram comparados com o valor de mercado dos imóveis na mesma localização, para verificar a existência e impacto das correlações existentes.

5.3.1 COMPLEXIDADE URBANA

O indicador da complexidade urbana (RUEDA, 2009) mede simultaneamente a diversidade e frequência de usos lucrativos do solo. Assim, permite conhecer o equilíbrio de usos no tecido urbano e avaliar seu grau de desenvolvimento.

A complexidade urbana é calculada em quadrículas de 200 x 200 metros no território, na unidade de bits de informação, proveniente da Teoria da Informação. O valor proposto como ideal é acima de 4 bits por quadrícula. Calcula-se com a fórmula de Shannon, na Equação 12:

$$\text{Complexidade urbana} = - \sum_{i=1}^n Pi \log_2 Pi \quad (12)$$

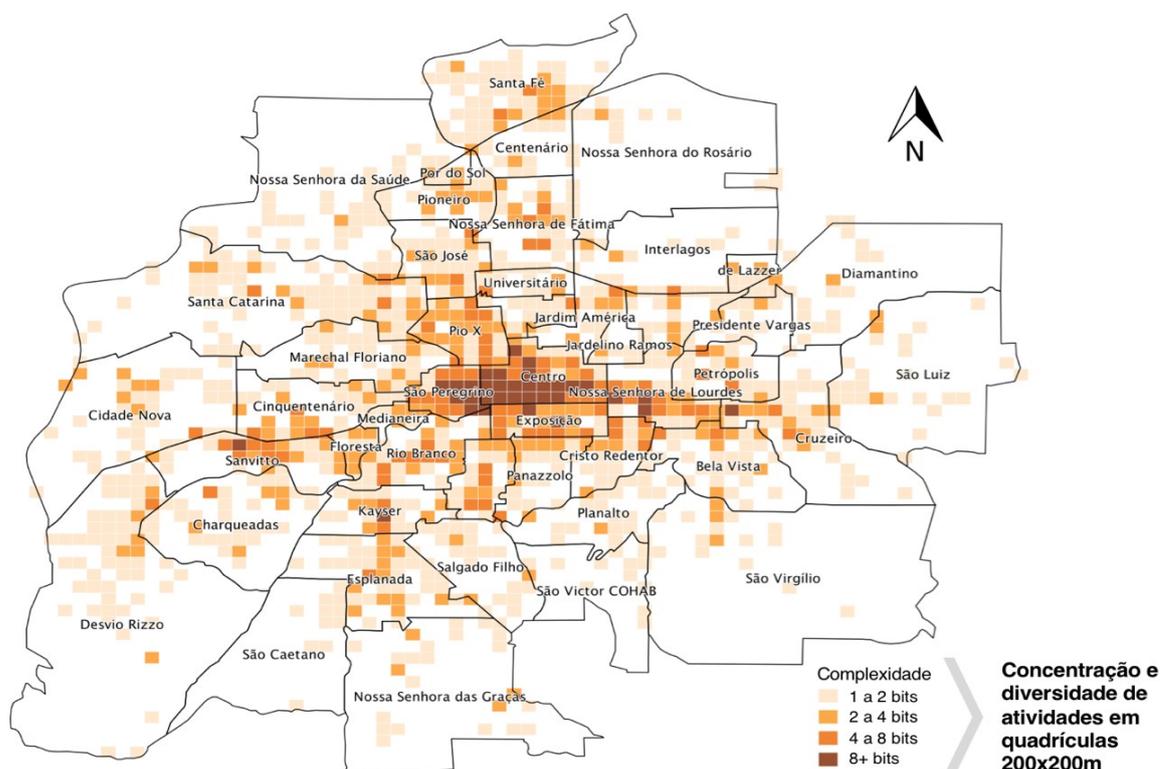
Onde:

N = número de estabelecimentos na quadrícula

Pi = abundância relativa de estabelecimentos desta categoria na quadrícula (Ni/N).

A Figura 31 e Apêndice B apresentam a complexidade urbana em Caxias do Sul – RS, na área do estudo. O mapa demonstra uma grande complexidade no centro da cidade e também pequenas centralidades no território. Ao todo, 134 quadrículas atingem o valor ideal.

Figura 31 - Indicador de complexidade urbana



Fonte: O Autor (2018).

5.3.2 SEGREGAÇÃO ESPACIAL

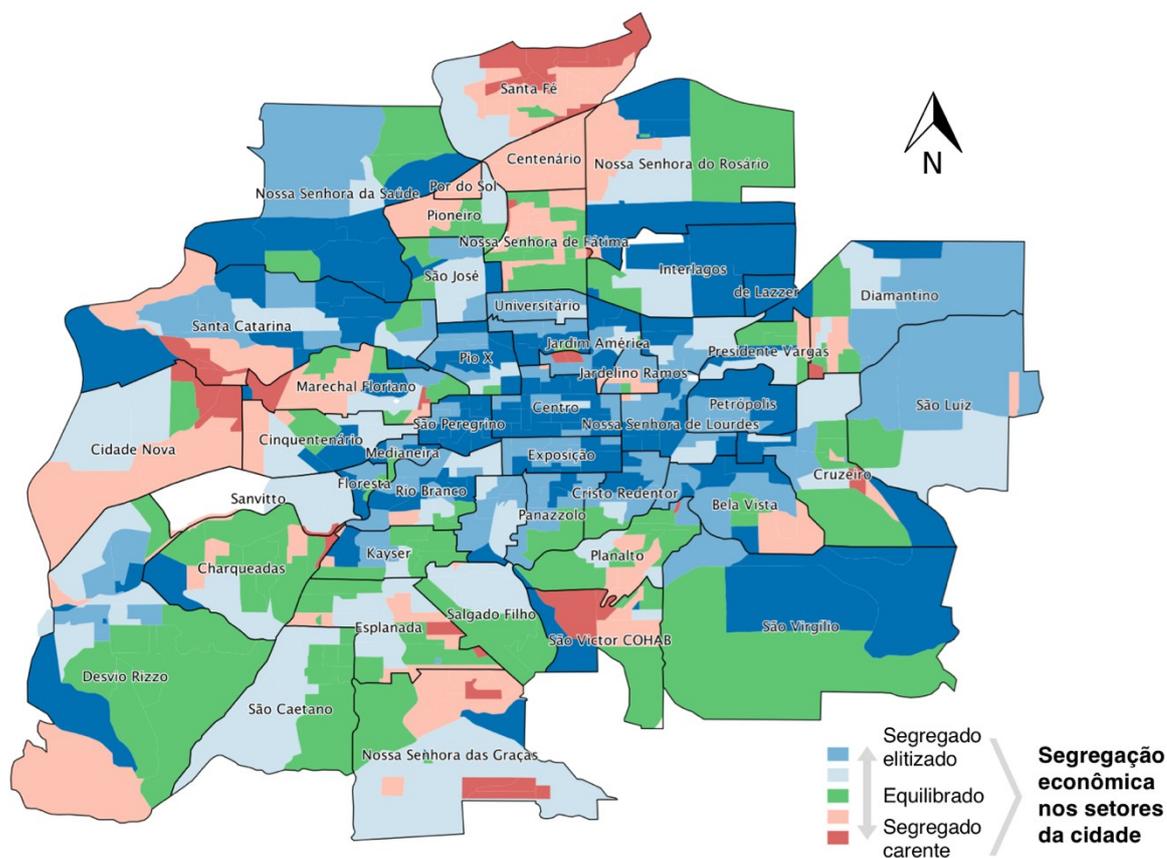
O indicador da segregação espacial (HERMIDA, 2015) mede a proporção de pessoas com menores recursos em relação à população total nos setores da cidade. Permite assim evidenciar situações de segregação e exclusão de populações.

Consideraram-se pessoas carentes aquelas cuja média de renda familiar é de até ¼ do salário mínimo. Calcula-se indicador da segregação espacial por meio da Equação 13:

$$\text{Segregação espacial} = \frac{\frac{\text{População carente no setor}}{\text{População no setor}}}{\frac{\text{População carente no município}}{\text{População no município}}} \quad (13)$$

A Figura 32 e o Apêndice C apresentam um mapa que permite identificar áreas com segregação espacial da população carente, áreas com equilíbrio de populações e áreas com segregação espacial de população com recursos superiores.

Figura 32 - Indicador de segregação espacial



Fonte: O Autor (2018).

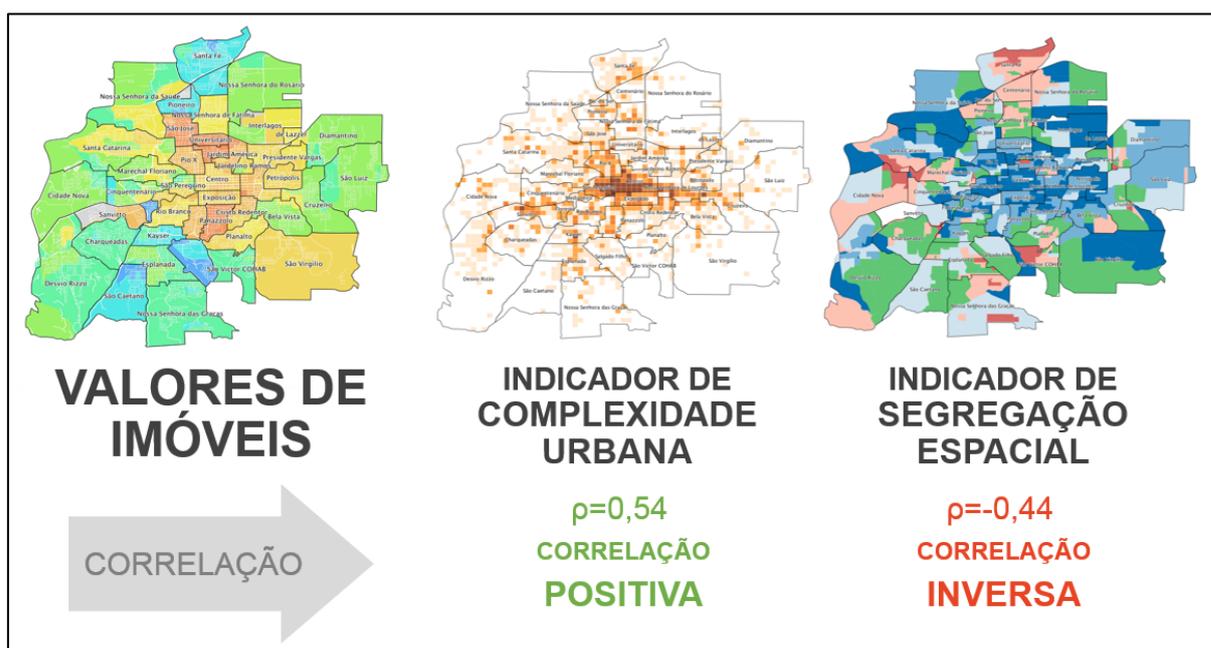
5.3.3 CORRELAÇÃO ENTRE VALORES E URBANISMO

Calculou-se a relação entre o valor de mercado dos imóveis e os aspectos urbanísticos avaliados por meio da correlação existente entre cada um dos indicadores e o mapa da “Influência da Localização”, elaborado no primeiro estudo de caso.

A análise de correlação exige a comparação entre dados com a mesma abrangência. As diferentes escalas (bairros, setores, quadriculas) dos mapas foram compatibilizados, interpolando as informações sempre na menor escala. Por exemplo, na complexidade urbana, avaliou-se a complexidade de cada quadricula em relação ao valor de mercado da mesma.

Encontraram-se as seguintes correlações, conforme demonstradas na Figura 33:

Figura 33 - Resultado da correlação entre indicadores e valor de mercado



Fonte: O Autor (2018).

A complexidade urbana ($\rho=0,54$) possui uma correlação positiva com os valores de mercado dos imóveis. Ou seja, a maior quantidade e diversidade de atividades lucrativas está associada a 54% com a maior valorização imobiliária daquele território.

Por outro lado, a segregação espacial ($\rho=-0,44$) possui uma correlação inversa com o valor de mercado dos imóveis. Desta forma, demonstra-se uma tendência de desvalorização imobiliária nos locais onde a população vive em situação de exclusão social.

Estes resultados demonstram que mesmo informações diferentes de fontes distintas (mercado imobiliário, atividades econômicas e coesão social) conectam-se, indicando a complexidade da cidade e a importância de análises conjuntas, holísticas, observando o todo.

A partir de conceitos como este é possível responder dúvidas pontuais, por exemplo, estimar *trade-offs*²² relacionadas às políticas públicas. Indicadores e cruzamentos de dados permitem calcular o impacto que os incentivos causam e embasar diretrizes otimizadas, capazes de promover um maior desenvolvimento, e conseqüentemente, uma maior qualidade de vida.

²² *Trade-off* é a situação que envolve a decisão entre trocar a redução de um aspecto pelo benefício de outro. Neste contexto, por exemplo: a avaliação de uma política pública para desenvolver centralidades urbanas. Conceitualmente, as análises apresentadas podem ajudar a quantificar o impacto imobiliário que esta política causa, permitindo uma decisão mais assertiva sobre seu custo de oportunidade (em relação às outras ações que poderiam igualmente ser tomadas).

CONCLUSÃO

Ask, and it shall be given to you. Seek, and you shall find. Knock, and it shall be opened to you.

Matthew 7:7

Este trabalho iniciou com uma revisão bibliográfica, explorando conceitos do mercado imobiliário, tais como engenharia de avaliações, preços hedônicos e econometria; e também ciência de dados: a transformação de dados em estratégias, com foco na regressão linear múltipla. Investigar estas ideias possibilitou conceitualizar um modelo preditivo.

O modelo preditivo elaborado permitiu estimar valores de mercado de apartamentos em Caxias do Sul – RS. O desenvolvimento deste modelo proporcionou a oportunidade de percorrer brevemente a jornada do *data science*: coleta e organização de dados, análises exploratórias, modelagem matemática e apresentação de narrativas envolventes – empregando a economia urbana para compreender um contexto do mercado imobiliário.

A “influência da localização” e a “simulação de valores de mercado” apresentaram situações onde agentes interessados em imóveis (construtores, engenheiros, consumidores...) podem explorar resultados de um modelo preditivo para tomar decisões assertivas, capazes de reduzir incertezas e custos, e aumentar a produtividade e lucratividade de organizações.

A “análise de valores e urbanismo” demonstra a possibilidade da descoberta e da pesquisa científica ligadas ao modelo preditivo. Evoca um cenário onde agentes vinculados às políticas públicas (planejadores urbanos, economistas, governantes...) podem embasar e elaborar políticas efetivas, que promovam desenvolvimento e potencializem as cidades.

Diante da construção do modelo e dos estudos de caso, pode-se afirmar que o objetivo “*elaborar um modelo preditivo para estimar valores de imóveis e realizar um estudo de caso*” foi atingido nesta investigação. Executar este trabalho possibilitou a descoberta de conceitos, a aplicação de conhecimentos de engenharia em cenários do mundo real e o incentivo para conhecer mais sobre os assuntos desenvolvidos.

A multidisciplinaridade entre a engenharia civil, matemática e economia permitiu conectar diferentes pontos, contribuindo para expandir conhecimentos e trazer novas perspectivas a problemas tradicionais. Os modelos preditivos oferecem inúmeras respostas aos agentes do mercado imobiliário e a certeza de que há muito potencial a ser explorado.

SUGESTÕES PARA FUTUROS TRABALHOS

Nesta investigação foram explorados métodos e exemplos clássicos. Percebe-se que estes são “uma gota no oceano” em termos de possibilidades de análises no mercado imobiliário. Na linha deste trabalho, como sugestões, podem-se analisar mais profundamente:

- a. Métodos de regressão, como técnicas de regressão não-lineares (exemplo: transformadas exponenciais e logarítmicas) explicando com maior fidelidade variáveis com distintos comportamentos; metodologias para validar sistematicamente os resultados do modelo.
- b. Modelos de regressão espacial, mais densos e complexos, capazes de capturar precisamente a distribuição da influência espacial, explicando assertivamente a influência que o território exerce sobre os dados.
- c. Aplicações práticas distintas, que gerem valor aos múltiplos agentes interessados em imóveis, utilizando modelos preditivos para responder questões tradicionais.
- d. Modelos que contemplem *similaridades* entre imóveis para modelar seus valores, ou seja, que estimem da mesma maneira como os agentes observam. Esta é uma abordagem paralela aos preços hedônicos, podendo capturar nuances de mercado. Substitui a regressão linear múltipla por métodos como *network science* e/ou *clusters*.
- e. Em modelos de similaridade, aproximar a estimativa do valor de mercado ao cerne da economia comportamental, de maneira que permita elucidar o *como* e *porquê* as pessoas decidem, buscando novas percepções a partir destes conhecimentos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, Denisard Cneio de Oliveira. “**Modelagem dos Preços de Imóveis Residenciais Paulistanos**”, Rev. Bras. Finanças, Rio de Janeiro, Vol. 9, No. 2, pp. 167–187, 2011.
- ANSELIN, Luc. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic, 1988. 294p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653: Parte 1 - Avaliação de Bens**. Rio de Janeiro: ABNT, 2001.
- _____. **NBR 14653: Parte 2 - Imóveis Urbanos**. Rio de Janeiro: ABNT, 2011.
- BANDYOPADHYAY, Raj. “**The Data Science Process**” [S.l.]: KDNuggets, 2016. Disponível em: <<https://www.kdnuggets.com/2016/03/data-science-process.html>>. Acesso em: 15 nov 2017.
- BAPTISTELLA, Marisa. “**O uso de redes neurais e regressão linear múltipla na engenharia de avaliações: Determinação dos valores venais de imóveis urbanos**”. *Dissertação de mestrado*. Ciências Exatas e Tecnologias, UFPR – PR, 2005.
- BARABASI, Albert-László. **Linked: How everything is connected to everything else and what it means for business, science and everyday life**. New York: Basic Books, 2014.
- BIOSSPLENA INTELIGÊNCIA URBANA. “**Indicadores de desenvolvimento urbano: Ferramenta para medir a eficiência no uso do território. Município de Lajeado – RS**”. Lajeado, RS. Fórum das Entidades. Apresentação de Giovana Ulian e Miguel Angel Pino Quilodrán em 10 out 2017.
- BRASIL. **Lei n. 5.194, de 24 de dez. de 1966: Do Exercício Profissional da Engenharia, da Arquitetura e da Agronomia**, Brasília, 1966. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L5194.htm>. Acesso em: 06 out 2017.
- _____. **Lei n. 6.530, de 12 de maio de 1978: Do exercício da profissão de Corretor de Imóveis**. Brasília, DF, 1978. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L6530.htm>. Acesso em: 08 out 2017.
- BRUSILOVSKIY, Eugene. “**Spatial Regression: A Brief Introduction**”. Philadelphia: Business Intelligence Solutions, 2009. Disponível em: <<http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Regression.php>>. Acesso em: 04 out 2017.

CARMO, Carlos Roberto Souza. “**Precificação imobiliária baseada em modelagem hedônica e externalidades: um estudo aplicado a terrenos urbanos**” ReFAE – Revista da Faculdade de Administração e Economia, v. 5, n. 2, p. 2-23, 2014.

CONWAY, Drew. “**The Data Science Venn Diagram**” [S.l.]: Drew Conway Data Consulting, 2010. Disponível em: <<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>>. Acesso em: 05 out 2017.

DANTAS, Rubens Alves. *Engenharia de Avaliações: Uma introdução à metodologia científica*. São Paulo: PINI, 1998. 255p.

_____. “**Modelos Espaciais aplicados ao mercado habitacional. Um estudo de caso para a cidade do Recife**”. *Tese de doutorado*. Economia, UFPE – PE, 2003.

ESRI. **ArcGIS Desktop: Release 10**. Redlands, CA: Environmental Systems Research Institute, 2011.

FERREIRA, Eduardo Neto “**Estimação do preço hedônico: Uma aplicação para o mercado imobiliário da cidade do Rio de Janeiro**”. *Dissertação de mestrado*, EPGE/FGV – RJ, 2002.

FIGUEREDO, E. A., & ZIEGELMANN, F. A. “**Mudança na distribuição de renda brasileira: significância estatística e bem-estar econômico**”. *Econ. Apl.* vol.13 no.2 Ribeirão Preto Abr /Jun 2009. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-80502009000200004>. Acesso em 10 jun 2018.

FONTENELLE, André. “**Metodologia científica: Como definir os tipos de pesquisa do seu TCC**”. Professor André Fontenelle, 2017. Disponível em: <<http://www.andrefontenelle.com.br/tipos-de-pesquisa/>>. Acesso em: 05 nov 2017.

FOLHA. “**Caixa sobe em 175% taxa de avaliação de imóvel para financiamento**”, 2015. Disponível em: <<http://www1.folha.uol.com.br/mercado/2015/05/1631722-caixa-sobe-em-175-taxa-de-avaliacao-de-imovel-para-financiamento.shtml>>. Acesso em: 21 nov 2017.

FOREMAN, John. *Data Smart: Using Data Science to Transform Information into Insight*. Indianapolis: John Wiley & Sons Inc., 2014. 409p.

GIHAB-CX - Gerência Executiva de Habitação de Caxias do Sul. **Conjunto de dados de Mercado Imobiliário para Caxias do Sul (RS) – 2016 a 2017** [Conjunto de Dados]. Orientações do Eng. Ms. Gustavo Sperotto, 2018.

GLASSDOOR. **“50 Best Jobs in America for 2016”**, Mill Valley, California: Glassdoor, 2016. Disponível em: <https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm>. Acesso em: 22 nov 2017.

GOOGLE. **“Places | Google Maps Platform | Google Cloud”**. Disponível em: <<https://cloud.google.com/maps-platform/places/>>. Acesso em fev 2018.

GONÇALVES, DANTAS, Ítalo Samuel; LIMA, José Eduardo de Carvalho; CASTRO, Lucas Ferreira. **“Modelagem econométrica para avaliação de imóveis urbanos residenciais novos no município de Juazeiro do Norte/CE”**. Congresso Técnico Científico da Engenharia e da Agronomia CONTECC, 2016.

GONZÁLEZ, Marco Aurélio Stumpf. **A Engenharia de Avaliações na Visão Inferencial**. São Leopoldo: Editora UNISINOS, 1997. 142p.

_____. **“Aplicação de técnicas de descobrimento de conhecimento em bases de dados e de inteligência artificial em avaliação de imóveis”**. *Tese de doutorado*. Programa de Pós Graduação em Engenharia Civil, UFRGS – RS, 2002.

GRILICHES, Zvi. **“Hedonic price indexes for automobiles: An econometric analysis of quality change”**. In: The price statistics of the Federal Government. National Bureau of Economic Research, 1961. New York. (General Series, 73). Cambridge, Harvard University Press, 1971 – p.55-87.

HBR. **“Data Scientist: The Sexiest Job of the 21st Century”**. Boston: Harvard Business Review, 2015. Disponível em: <<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>> Acesso em: 21 nov 2017.

HERMIDA, Augusta et al. **La ciudad es esto: Medición y representación espacial para ciudades compactas y sustentables**. Ecuador: Universidade de Cuenca, 2015. 172p.

HOFFMAN, Rodolfo. **Análise de regressão: uma introdução à econometria**. Piracicaba: Portal de Livros Abertos da USP, 2016. 393p.

HORNBURG, R. A., & HOCHHEIM, N.. **Avaliação em massa de imóveis usando Regressão Espacial e Krigagem Bayesiana: Um estudo de caso em Balneário Camboriú - SC**. Reec – Revista Eletrônica de Engenharia Civil – Volume 13, No 1, 133-155, Jan 2017 – Jun 2017.

IBGE. **Censo Demográfico 2010**: Resultados do universo, agregados por setores censitários. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2010. Disponível em <https://downloads.ibge.gov.br/downloads_estatisticas.htm>. Acesso em 25 ago 2017.

_____. **Censo Demográfico**: O que é. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2010a. Disponível em <<https://www.ibge.gov.br/estatisticas-novoportal/sociais/saude/9662-censo-demografico-2010.html?=&t=o-que-e>>. Acesso em 25 ago 2017.

KAGGLE. “**Zillow Prize: Zillow’s Home Value Prediction (Zestimate)**”. San Francisco: Kaggle, 2017. Disponível em <<https://www.kaggle.com/c/zillow-prize-1>>. Acesso em 2 de jul 2017.

KAHNEMAN, Daniel. *Thinking, Fast and Slow*. New York: Farrar Straus Giroux, 2012. 499p.

KNAFLIC, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. New Jersey: John Wiley & Sons, 2015. 288p.

KELLEY, P. R, Barry. R; Sirmans, C.F. “**Spatial Statistics and Real Estate**” Journal of Real Estate Finance and Economics, Volume 17, Number 1, p. 5-13, 1998.

LANCASTER, Kelvin. “**Consumer Demand: A New Approach**”. New York: Columbia University Press, 1971.

MATTA, Túlio Alves. **Avaliação do valor de imóveis por análise de regressão: um estudo de caso para a cidade de Juiz de Fora**. *Monografia*. Faculdade de Engenharia, Universidade Federal de Juiz de Fora, 2007.

MCCANDLESS, David. *The Visual Miscellaneum*. Honk Kong: Haper Design, 2012. 195p.

NETSCIED. *Network Literacy: Essential Concepts and Core Ideas*. [S.l]:[S.n], 2016.

OLIVEIRA, Sara Andrade. **Comportamento de índices econômicos em relação aos valores de venda de imóveis em Caxias Do Sul/RS**. *Monografia*. Área do conhecimento de Ciências Exatas e Engenharias, Universidade de Caxias do Sul, 2017.

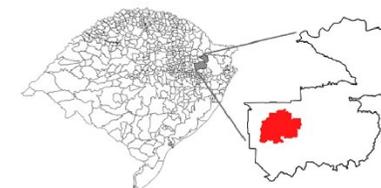
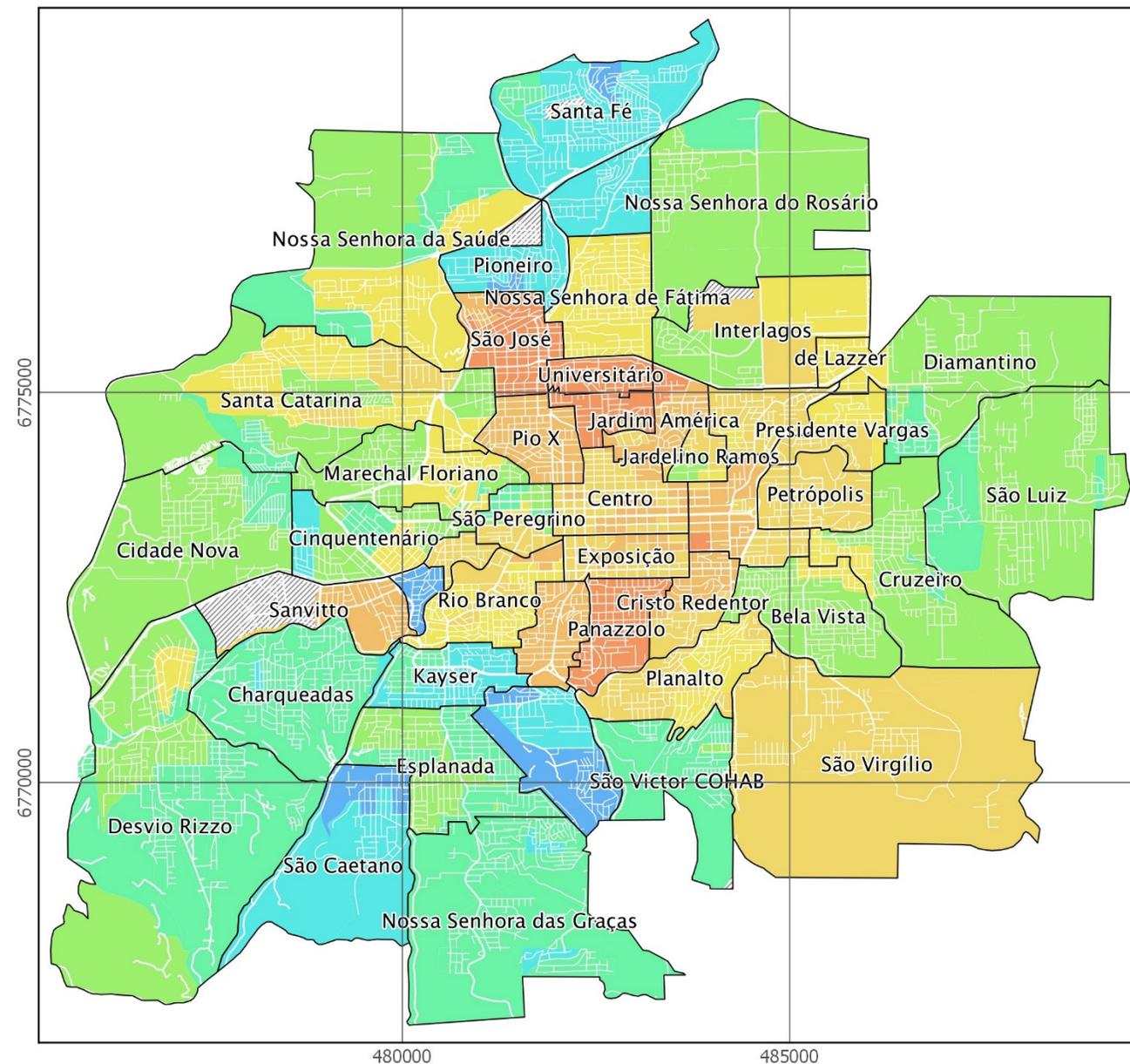
OLIVEIRA NETTO, Alvim Antônio de. *Metodologia da Pesquisa Científica: Guia Prático para Apresentação de trabalhos Acadêmicos*. 2º ed. Florianópolis: Visual Books, 2006.

OPEN STREET MAPS, Contribuidores. “**OSM Data for Caxias do Sul - RS**”. [S.l], 2018. Disponível em: <<https://www.openstreetmap.org/export>>. Acesso em: 30 mar 2018.

PAIXÃO, Luiz André Ribeiro. “**Índice de preços hedônicos para imóveis: uma análise para o município de Belo Horizonte**”. Econ. Apl. vol.19 no.1 Ribeirão Preto Jan./Mar. 2015.

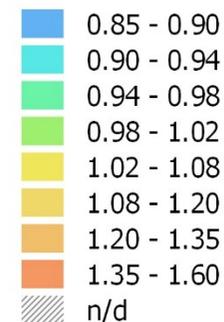
- PIONEIRO. “*Confira quais são os bairros mais caros de Caxias*”. Caxias do Sul, RS, 2016. Disponível em: <<http://pioneiro.clicrbs.com.br/rs/economia/noticia/2016/07/confira-quais-sao-os-bairros-mais-caros-de-caxias-6688102.html>>. Acesso em: 03 mar 2018.
- PORTER, Eduardo. *O Preço de Todas As Coisas*. São Paulo: Objetiva, 2011. 288p.
- PROVOST, Foster; FAWCETT, Tom. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol: O'Reilly Media, 2013. 414p.
- RASCOFF, Spencer; HUMPHRIES, Stan. *Zillow Talks: The New Rules of Real Estate*. New York: Grand Central Publishing, 2015. 310p.
- REVELLE, Charles; WHITLATCH, Earl; WRIGHT, Jeff. *Civil & Environmental Systems*. London: Pearson, 2003. 552p.
- QGIS, Development Team. *QGIS Geographic Information System 2.18.17*. Open Source Geospatial Foundation Project, 2018. Disponível em <<http://qgis.osgeo.org>>. Acesso em mar 2018.
- ROSEN, Sherwin. “**Hedonic Prices and Implicit Markets: Product Differentiation Perfect Competition**”. *Journal of Political Economy* 82, p 34-55, 1974.
- RUEDA, Salvador. *Plan de indicadores de sostenibilidad urbana de Vitoria-Gasteiz*. Agência de Ecologia Urbana de Barcelona: Barcelona, 2010. 477p.
- SIEGEL, Eric. *Predictive Analysis: The power to predict who will click, buy, lie or die*. New Jersey: John Wiley & Sons, Inc., 2016. 332p.
- SMITH, Adam. *The Wealth of Nations, Book I*. Penguin Classics, 1999 (1776).
- TALEB, Nicholas. *A Lógica do Cisne Negro: o impacto do altamente improvável*. 6a edição. Rio de Janeiro: BestSeller, 2012. 458p.
- UCS. *Guia para elaboração de Trabalhos Acadêmicos*. Caxias do Sul: [s.n.], 2016. 88p.
- ZILLOW. [Mapa de preços de imóveis em Champaign, Illinois]. Seattle: Zillow, 2017. Disponível em <https://www.zillow.com/homes/for_sale/globalrelevanceex_sort/40.115104,-88.229192,40.111888,-88.236107_rect/17_zm/>. Acesso em 05 out 2017.
- _____. *What is a Zestimate? Zillow's Zestimate Accuracy*. Seattle: Zillow, 2016. Disponível em <<https://www.zillow.com/zestimate/>>. Acesso em 03 out 2017.

Apêndice A - Influência da Localização



Localização no estado do Rio Grande do Sul e na cidade de Caxias do Sul

Fator da Localização



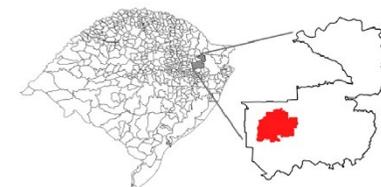
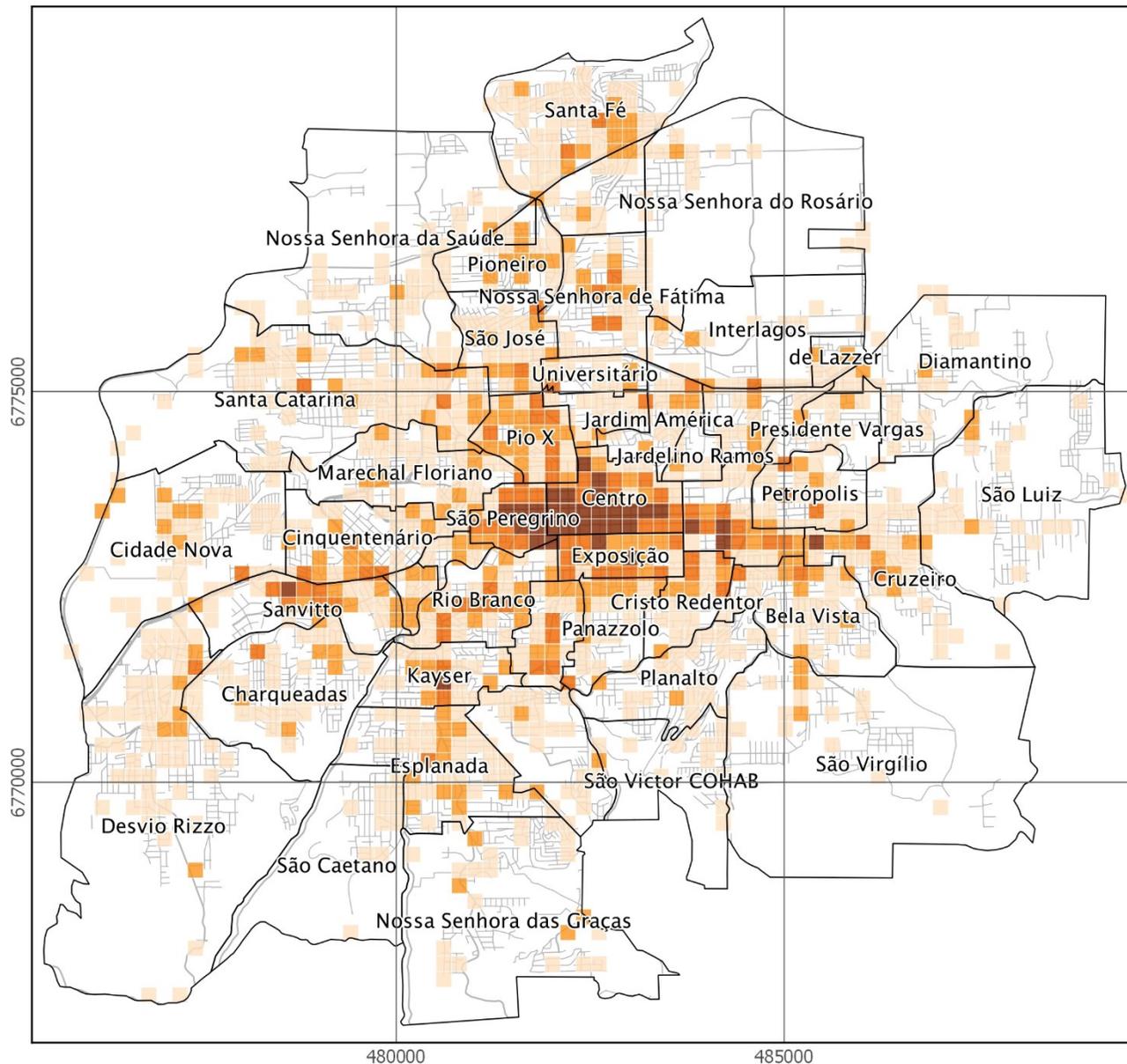
Escala 1:80.000

Sistema de coordenadas UTM Zona 22S
Sistema geodésico: SIRGAS 2000

Elaboração: Eduardo Vicensi De Bastiani
Junho/2018.



Apêndice B - Complexidade Urbana



Localização no estado do Rio Grande do Sul e na cidade de Caxias do Sul

Bits de informação

- 1 - 2 bits
- 2 - 4 bits
- 4 - 8 bits
- 8+ bits



0 1 2 3 4 5 km



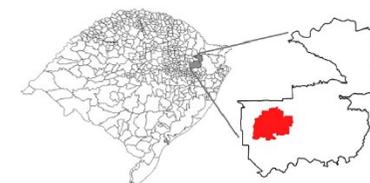
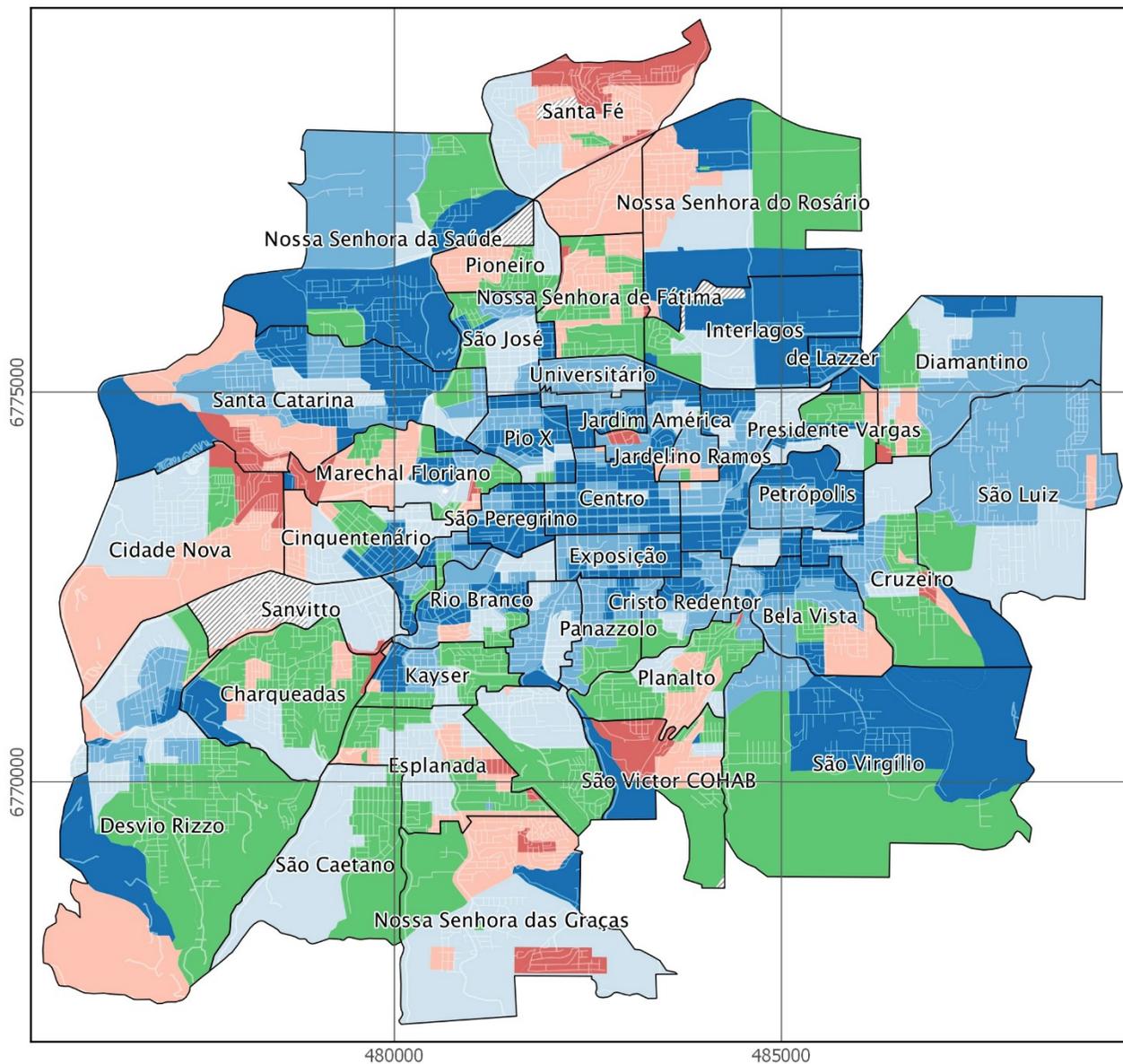
Escala 1:80.000

Sistema de coordenadas UTM Zona 22S
Sistema geodésico: SIRGAS 2000

Elaboração: Eduardo Vicensi De Bastiani
Junho/2018.



Apêndice C - Segregação Espacial



Localização no estado do Rio Grande do Sul e na cidade de Caxias do Sul

Segregação Espacial

- 0.00 - 0.25
- 0.25 - 0.50
- 0.50 - 0.75
- 0.75 - 1.25
- 1.25 - 2.50
- 2.50 - 6.00
- n/d



Escala 1:80.000

Sistema de coordenadas UTM Zona 22S
Sistema geodésico: SIRGAS 2000

Elaboração: Eduardo Vicensi De Bastiani
Junho/2018.

