

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA URBANA

ESTUDO DOS CONDICIONANTES ESPACIAIS PARA
AVALIAÇÃO IMOBILIÁRIA UTILIZANDO TÉCNICAS DE
INTELIGÊNCIA ARTIFICIAL – SÃO PAULO/SP

FELIPE PEREIRA HONDA

São Carlos

2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA URBANA

**ESTUDO DOS CONDICIONANTES ESPACIAIS PARA
AVALIAÇÃO IMOBILIÁRIA UTILIZANDO TÉCNICAS DE
INTELIGÊNCIA ARTIFICIAL – SÃO PAULO/SP**

FELIPE PEREIRA HONDA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Urbana da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Urbana.

Orientação: Prof. Dr. Edson Augusto Melanda

São Carlos
2021



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Engenharia Urbana

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Felipe Pereira Honda, realizada em 01/04/2021.

Comissão Julgadora:


Prof. Dr. Edson Augusto Melanda (UFSCar)

p/ Prof. Dr. Ricardo Augusto Souza Fernandes (UFSCar)

p/ Profa. Dra. Mayra Cristina Prado de Moraes (SENAC)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Engenharia Urbana.

AGRADECIMENTOS

Nesses anos de mestrado, de muito estudo, dedicação e aprendizado, gostaria de agradecer a algumas pessoas que me acompanharam e se tornaram fundamentais para a realização de mais este sonho. Primeiramente agradeço à Deus por sempre colocar pessoas incríveis em meu caminho, as quais me fazem acreditar em um mundo melhor e me encorajam a prosseguir. Aos meus pais, Marcus e Constance, que sempre me direcionaram para ensinar o caminho do bem, e me apoiaram em todas as etapas da minha vida. Em especial à minha mãe, sou eternamente grato por tudo que sou, por tudo que consegui conquistar e pela felicidade que tenho hoje. À minha irmã Anna Carolina, por estar sempre ao meu lado, me apoiando e me motivando em todos os momentos. À minha família, por acreditar no meu sonho e sempre me motivar para seguir em frente. Ao meu orientador, Professor Edson Augusto Melanda, pela oportunidade de realizar essa pesquisa. Agradeço pela confiança, pelos ensinamentos compartilhados de forma admirável, e por me guiar nos primeiros passos da pós-graduação. Aos professores do programa de pós-graduação em Engenharia Urbana, pelos ensinamentos que transcendem os limites da Universidade. Aos amigos Maria Eugenia, Alex, Thamiris, Thais e Fabio, os quais me ajudaram em todos os momentos de dificuldade na pós-graduação. A todos que de alguma forma contribuíram para a realização deste trabalho, o meu muito obrigado!

RESUMO

O aumento da oferta de imóveis nos últimos anos influenciou no aumento do crédito, e, contudo, elevou os preços de venda. Todavia, a mensuração dos valores de imóveis ainda é um processo bastante complexo, pois é influenciado por inúmeros fatores, dentre eles físicos, geolocalização e político-econômicos. Deste modo, o objetivo desta pesquisa foi apresentar quais variáveis são relevantes na formação do valor de imóveis urbanos utilizando técnica de inteligência artificial. Optou-se por utilizar a técnica de árvore de decisão para mineração e análise estatística, criada a partir do algoritmo Random Forest. Através da mineração foi possível notar que existe uma forte correlação entre as categorias e subcategorias analisadas, e, sobretudo, a influência que exercem na formação do valor de imóveis urbanos nas 32 regiões administrativas do município. Dentre os fatores que apresentaram maior influência no valor dos imóveis foram a proximidade à Shopping Centers, postos policiais, áreas de ocorrência de riscos geológicos, favelas, consulados, corpo de bombeiros e estações de trem. Verificou-se também que os diferentes aspectos ambientais, positivos e negativos, influenciam na valorização imobiliária e que a metodologia adotada se mostra eficiente na avaliação de grandes bancos de dados, assim como, uma importante ferramenta para tomada de decisão e para o planejamento urbano.

Palavras-chave: Mineração de dados; Inteligência artificial; Valorização imobiliária.

ABSTRACT

The increase in the supply of real estate in recent years has influenced the increase in credit, and, however, has raised sales prices. However, the measurement of real estate values is still a very complex process, as it is influenced by numerous factors, among them physical, geolocation and political-economic. Thus, the objective of this research is to present which variables are relevant in the formation of the value of urban properties using artificial intelligence technique. It was decided to use the decision tree technique for mining and statistical analysis, created from the Random Forest algorithm. Through mining it was possible to notice that there is a strong correlation between the categories and subcategories analyzed, and, however, the influence they have on the formation of the value of urban properties in the 32 administrative regions of the municipality. Among the factors that had the greatest influence on the value of the properties were the proximity to Shopping Centers, police stations, areas where geological hazards occur, slums, consulates, fire departments and train stations. It was also found that the different environmental aspects, positive and negative, influence the real estate valuation and that the adopted methodology is efficient in the evaluation of large databases, as well as an important tool for decision making and for urban planning.

Keywords: Data Mining, Artificial intelligence, Real estate valuation.

LISTA DE FIGURAS

Figura 1: Etapas do processo de mineração de dados.....	17
Figura 2: Processo de KDD.....	19
Figura 3: Localização das trinta e duas regiões administradas (Prefeituras Regionais), na Escala 1:280.000, Sistema de Coordenadas WGS84/Sirgas.....	26
Figura 4: Localização dos Parques Municipais e Unidades de Conservação em São Paulo-SP.....	29
Figura 5: Localização das Reservas de Mata Atlântica em São Paulo-SP.....	30
Figura 6: Densidade da arborização urbana em São Paulo-SP.....	31
Figura 7: Verde e Recursos Naturais em São Paulo-SP.....	32
Figura 8: Dados de densidade demográfica e localização das favelas nas Prefeituras Regionais de São Paulo-SP.....	36
Figura 9: Processo metodológico.....	41

LISTA DE QUADROS

Quadro 1: Densidade Populacional por Prefeitura Regional, com suas respectivas áreas.	33
Quadro 2: Número de registros referentes à 20 atributos principais e 74 atributos secundários das 32 Prefeituras Regionais de São Paulo-SP.	42

LISTA DE SIGLAS E ABREVIATURAS

AI	Artificial Intelligence
DCBD	Descoberta de Conhecimento em Banco de Dados
DCBDG	Descoberta de Conhecimento em Bancos de Dados Geográficos
EDA	Exploratory Data Analyses
GPS	Global Positioning System
IPCA	Índice Nacional de Preços ao Consumidor Amplo
IA	Inteligência Artificial
KDD	Knowledge Discovery in Databases
RMSP	Região Metropolitana de São Paulo
SGBD	Sistemas de Gerência de Banco de Dados
UTM	Universal Transversa de Mercator

Sumário

1. INTRODUÇÃO	10
1.1. Contextualização.....	10
1.2. Justificativa	13
1.3. Objetivos	14
1.3.1. Objetivos Específicos.....	14
2. BASES CONCEITUAIS	15
2.1. Inteligência Artificial (IA)	15
2.2. Mineração de dados ou “ <i>Data mining</i> ”	16
2.2.1. Árvores de decisão	22
2.3. Método Comparativo de Dados de Mercado	24
2.4. Valores hedônicos de imóveis.....	24
2.5. Descrição da Área de Estudo	25
2.5.1. Características ambientais.....	27
2.5.2. Aspectos Sociais.....	33
2.5.3. Características Econômicas.....	37
2.5.4. Infraestrutura Urbana	38
3. METODOLOGIA.....	40
3.1. Etapas para tratamento dos dados	42
3.2. Mineração dos dados.....	45
4. RESULTADOS E DISCUSSÃO.....	49
5. CONCLUSÕES.....	52
7. REFERÊNCIAS BIBLIOGRÁFICAS	54

1. INTRODUÇÃO

1.1. Contextualização

O período de industrialização na cidade de São Paulo, ocorrida no século XX, foi decisivo e impulsionou o crescimento, assim como refletiu no desenvolvimento econômico e na expansão da urbanização (MARSON, 2014).

Devido ao rápido crescimento da metrópole, em condição tardia e sustentada por baixos salários, o acúmulo de capital resultante ficou concentrado nas mãos de poucos. Desta forma, a massa trabalhadora foi cada vez mais alocada às regiões periféricas da metrópole, em decorrência tanto da valorização do solo urbano na região central quanto do planejamento de políticas públicas segregacionistas (VÉRAS, 2001).

O resultado deste processo é o deslocamento de grande parcela da população para zonas perimetrais, que, em muitos casos, vivendo em condições precárias e distante da região central onde se concentram seus postos de trabalho e grande parte dos equipamentos públicos e comunitários.

A expansão urbana, assim como a aglomeração de homens e suas atividades afetou enormemente a qualidade do meio ambiente nas cidades, e, contudo, vem ocorrendo grande degradação ambiental no que se refere à poluição da água, ar, solo, sonora, dentre outras, que provocam efeitos nocivos e custos à sociedade.

De acordo com Léfèbvre (2004), o “direito à cidade”, envolve um conjunto de fatores relacionados não apenas ao acesso à infraestrutura em seus aspectos palpáveis, mas também no que se refere a recursos imateriais ligados ao bem estar da população, da convivência, da participação, da produção social, do acesso à natureza e à cultura, independentemente da condição socioeconômica do indivíduo.

No caso de São Paulo, como grande aglomerado urbano com características excludentes, existe uma relação forte entre a posição geográfica da habitação e o acesso à infraestrutura e o direito à cidade.

A problemática habitacional, desta forma, relaciona-se diretamente com o acesso à infraestrutura urbana, como: saneamento, saúde, educação, cultura, lazer, condições de mobilidade, segurança, meio ambiente, entre outros.

O fato de um terreno ou benfeitoria com características idênticas apresentarem valores discrepantes de acordo com a sua posição geográfica, está relacionado ao que se denomina de

“renda imobiliária” ou “renda fundiária urbana” e está diretamente ligado à disponibilidade de infraestrutura local.

Para Marx (1989, *apud* Botelho, 2008, p.26), a renda fundiária seria elevada pelo rápido e intenso crescimento da população nas grandes cidades, e pela conseqüente necessidade crescente de habitações daí resultante e também pela implementação do capital fixo que se incorporaria à terra (como edifícios, ferrovias, rodovias, armazéns, estabelecimentos fabris e comerciais, docas, etc).

Segundo Pelli Neto (2006), o mercado imobiliário, de acordo com o volume de recursos utilizados em suas transações, implica em um segmento de grande importância para a economia nacional. Para o estudo deste mercado, a base para os cálculos de demanda habitacional voltado às aplicações de recursos financeiros é calculada pelo valor de mercado, cuja estimativa é dada por meio das avaliações em massa.

Devido à mensuração dos valores de imóveis ainda ser um processo bastante complexo, pois é influenciado por inúmeros fatores, dentre eles físicos, ambientais, geolocalização, e fatores político-econômicos, novas técnicas são utilizadas pelos avaliadores de imóveis e propriedades. Utiliza-se os conhecimentos especializados para o desenvolvimento de seleção de atributos semelhantes à outras regiões para fins de avaliação em massa, e integração em técnicas de regressão múltipla, com programas computacionais.

A população atribui um valor ao consumo de amenidades locais, e pagam para usufruírem delas através de maior preço da habitação, portanto, as características locais também influenciam muito o valor da propriedade.

Outra abordagem para calcular a influência de amenidades locais no valor da propriedade é uma combinação de modelagem multinível entre as relações topológicas de propriedade e amenidades locais, por Orford (2002).

Atualmente existem teorias que tratam também sobre a determinação dos custos e benefícios à alteração do meio ambiente, porém, ainda são poucas estimativas e aplicações práticas.

Estudos mostram que existem vários métodos de valoração econômica ambiental no Brasil, porém, além de muito teóricos, apresentam muitas limitações na avaliação dos bens e serviços gerados pela natureza. De acordo com Freitas *et. al.* (2004), a valoração ambiental surge como uma importante ferramenta para a gestão, por oferecer parâmetros para avaliação dos custos relacionados aos recursos ambientais e os serviços, sendo possível atribuir valor

econômico aos ativos ambientais através da medição da variação do bem-estar provocado pela alteração dos recursos ambientais.

Desta forma, as Técnicas de Inteligência Artificial (IA), através da mineração de dados, têm sido utilizadas a fim de suprir as dificuldades para analisar grandes volumes de dados, racionar, perceber e tomar decisões para resolução de problemas complexos.

Para poder estudar uma quantidade tão grande de dados é necessário técnicas e algoritmos para poder organizar, padronizar, e então, obter resultados que contém informação útil, que pode ser estudada e assim obter conhecimento. É nessas técnicas e algoritmos que reside a mineração de dados (*Data Mining*).

O processo de *Data Mining* baseia-se em questões estatísticas. Criam-se modelos para descrever os “dados-alvo”, então verifica-se se o modelo é válido, e se sim, obtém-se padrões que são utilizados para descrever e classificar os dados. Como essas informações são comparadas a um número grande de exemplos, pode-se esperar que os dados que se encaixam aos modelos propostos sigam os mesmos padrões.

A Descoberta de Conhecimento em Bancos de Dados Geográficos (DCBDG) é um tipo particular de descoberta, pois está ligada à extração de características e padrões espaciais interessantes, à identificação de relacionamentos entre dados espaciais e não-espaciais/descritivos/convencionais, restrições entre objetos geográficos e outras características não explicitamente armazenadas nestes bancos de dados. Os dados convencionais descrevem características como: nome e população de um município. Os dados espaciais representam a localização de dados descritivos sobre a superfície da terra e são representados, geralmente, por três primitivas geométricas: ponto, linha e área (BIGOLIN *et al.*, 2003).

Neste contexto, para solucionar esses problemas, foram desenvolvidos dois enfoques diferentes que se complementam, sendo um pela comunidade de Inteligência Artificial e o outro, pela comunidade de Banco de Dados.

O objetivo da comunidade de Banco de Dados foi desenvolver Sistemas de Gerência de Banco de Dados (SGBD) a fim de oferecer ferramentas que possibilitem o armazenamento e a manipulação de grandes quantidades de informação estruturada e com rápido acesso. O aspecto de modelagem dos dados, as linguagens de consulta e a eficiência na recuperação da informação foi a maior preocupação durante o desenvolvimento desse sistema.

A integração de técnicas de inteligência artificial com as ferramentas computacionais de produção e análise de dados geográficos (Geoprocessamento) é, portanto, muito benéfica,

principalmente em relação à grande quantidade de dados geralmente envolvida nas questões de natureza espacial.

Desta forma, é importante ressaltar a importância de se desenvolver estudos sobre novas técnicas que integrem o capital natural na análise econômica, considerando a valoração ambiental como aspecto relevante, assim como, se obter análise integrada da infraestrutura e equipamentos públicos comunitários disponíveis nas regiões urbanas para implementação de eficientes políticas públicas focadas ao cumprimento da função social da cidade.

1.2. Justificativa

Nos últimos anos, o aumento do número de empreendimentos imobiliários e da facilidade para conseguir crédito para financiamento influenciou no aumento dos preços de imóveis no Brasil. Porém, esse momento pode ser considerado mais uma exceção do que uma regra no mercado imobiliário brasileiro. Em São Paulo, os valores de venda dos imóveis tiveram um aumento de 103% entre 1975 e 2015, equivalente a uma valorização média anual modesta de 1,8% ao longo dos últimos 40 anos, já descontada a inflação do período medida pelo IPCA (FIPE, 2015).

Este estudo é um dos poucos casos de aplicação de técnicas e mineração de dados para avaliação imobiliária com o objetivo de mostrar a influência da proximidade e distribuição espacial dos serviços públicos comunitários na valorização imobiliária, assim como, demonstrar que as variáveis ambientais são importantes na formação do valor de imóveis urbanos no município de São Paulo/SP.

Além da metodologia proposta de estudo de condicionantes espaciais através da Técnicas de Arvore de Decisão para correlação de dados e verificação dos condicionantes mais relevantes, será demonstrada a influência, positiva e negativa, das variáveis ambientais na formação do valor de imóveis urbanos localizados nas diferentes subprefeituras regionais, através de análise de dados obtidos nas bases de dados disponíveis na internet sobre o município de São Paulo (SP).

O estudo permitirá analisar a valoração ambiental de locais existentes, assim como, mostrar que é possível estimar os benefícios auferidos pelas prefeituras, através de um aumento na arrecadação do imposto predial e territorial urbano devido à valorização de imóveis urbanos beneficiados com a implementação de projetos de urbanização que melhorem as condições

ambientais da cidade. Assim, espera-se propor melhorias e soluções mitigadoras para a problemática geográfica da habitação na cidade.

1.3. Objetivos

O objetivo desta pesquisa é mostrar quais equipamentos, serviços públicos e elementos do meio ambiente são relevantes na formação dos valores imobiliários, assim como, aferir se as variáveis ambientais são importantes na formação do valor de imóveis urbanos no município de São Paulo/SP.

1.3.1. Objetivos Específicos

- Verificar quais equipamentos, serviços urbanos e elementos do meio ambiente mais influenciam na valorização imobiliária;
- Verificar se os elementos do meio ambiente próximos à imóveis urbanos possuem aspecto valorativo ou depreciativo no município de São Paulo-SP;
- Contribuir aos estudos de valoração econômica do meio ambiente, através da análise estatística de grande banco de dados utilizando técnica de inteligência artificial;
- Propor melhorias e soluções mitigadoras para a problemática geográfica da habitação na cidade.

2. BASES CONCEITUAIS

2.1. Inteligência Artificial (IA)

As pesquisas na área de inteligência artificial iniciaram nos anos de 1940 com o intuito de buscar funcionalidades para o computador que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas. Após a Segunda Guerra Mundial, buscou-se também o desenvolvimento de tecnologias para impulsionar a indústria bélica.

Após alguns anos passam a surgir vários estudos relacionados à aplicação de técnicas de inteligência artificial, dentre eles destaca-se o direcionamento à área biológica, a fim de desenvolver conceitos com o intuito de imitar as redes neurais.

Nos anos 60 essa ciência recebe a denominação de inteligência artificial, e os pesquisadores acreditavam ser possível que as máquinas pudessem raciocinar, assim como realizar tarefas humanas complexas.

Em 1980 os estudos sobre redes neurais são retomados, e nos anos 90 são consolidados com base em pesquisas realizadas. Atualmente existem várias aplicações de técnicas de inteligência artificial, como em: jogos, programas de computador, robótica, aplicativos de segurança para sistemas informacionais, dispositivos para reconhecimentos de escrita à mão e reconhecimento de voz, programas de diagnósticos médicos, entre outros.

Segundo Piatetsky-Shapiro (1999), Padmanabhan & Tuzhilin (1999), Barroso *et al.* (2003) *apud* Melanda (2004), a popularização da internet, maior e principal rede de computadores do planeta e os avanços tecnológicos em banco de dados, sistemas de comunicação, mecanismos de coleta e armazenamento de dados, têm proporcionado às organizações a capacidade de armazenar informações detalhadas sobre cada transação efetuada, gerando grandes bases de dados.

Bogorny (2008) afirma que os avanços nas tecnologias de coleta de dados espaciais e espaço-temporais (por exemplo, telefones celulares, dispositivos GPS) facilitaram a coleta de dados referenciados no espaço e no tempo. Essas enormes coleções de dados geralmente ocultam informações interessantes que os sistemas convencionais não conseguem descobrir. Dados espaciais e espaço-temporais requerem técnicas complexas de pré-processamento de dados, transformação, mineração de dados e pós-processamento para extrair padrões novos, úteis e compreensíveis.

Para Melanda (2004), a crescente demanda por informações precisas, detalhadas e atualizadas, necessárias ao gerenciamento das organizações, levou à busca por meios mais rápidos e eficientes de recuperação da informação e do conhecimento, incentivando a realização de pesquisas voltadas para o processo de transformação de dados em conhecimento.

A comunidade de inteligência artificial interessou-se pela mineração de dados (extração de conhecimento a partir de pequenas quantidades de dados) e a aprendizagem a partir de um pequeno conjunto de informações. A combinação dessas duas abordagens originou a área de Descoberta de Conhecimento em Banco de Dados (DCBD). As técnicas tradicionais de aprendizagem descobrem conhecimento a partir de dados extraídos do mundo real, enquanto as técnicas de mineração utilizam dados extraídos de bancos de dados. (BIGOLIN *et al.*, 2003)

De acordo com Bigolin *et al.* (2003) e Bogorny (2003), os dados de entrada para a mineração são representados em uma estrutura simples, e as informações relevantes são definidas pelo especialista. Já no segundo caso, além da estrutura e os tipos de dados serem complexos, o volume de informações é muito grande e, parte deste conjunto, pode ser irrelevante. Assim, é necessário efetuar dois tratamentos: um, para encontrar um subconjunto de informações apropriado e outro para adaptar essas informações em uma estrutura aceitável pelas ferramentas de mineração de dados. A seleção das informações relevantes é tão importante quanto o reconhecimento de padrões, pois sem os dados apropriados, dificilmente algum conhecimento útil será extraído do banco de dados.

Para selecionar a amostra adequada é fundamental que haja um entendimento do domínio da aplicação e um prévio conhecimento dos dados relevantes, para que se possa estabelecer as metas do processo de descoberta de conhecimento do ponto de vista do usuário.

2.2. Mineração de dados ou “Data mining”

O avanço da tecnologia permitiu às organizações manipular de forma eficiente a gigantesca quantidade de dados (capturar, organizar, armazenar) obtidas de suas operações diárias ou pesquisas científicas, porém, a maioria ainda não utiliza essas informações de forma adequada em suas atividades (comerciais ou científicas), para auxiliar precisamente na tomada de decisões.

Buscando automatizar o processo de análise, várias pesquisas têm sido desenvolvidas da área de mineração de dados. Segundo Bogorny (2008), a mineração de dados espaciais tornou-se mais importante devido à necessidade de avaliar a quantidade de informações já

disponíveis e os grandes conjuntos de dados geoespaciais, como mapas, repositórios de imagens de sensoriamento remoto, trajetórias de objetos móveis gerados por dispositivos móveis, etc.

Diferentes abordagens para a divisão do processo de mineração de dados em etapas são encontradas na literatura. Fayyad *et al.* (1996b) propõem a divisão do processo em nove etapas. De acordo com Weiss & Indurkha (1998) *apud* Rezende (2003), o processo de mineração é dividido em quatro etapas. Rezende *et al.* (2003) *apud* Melanda (2004) agrupam o processo em um ciclo com três grandes etapas: Pré-processamento, Extração de Padrões e Pós-processamento. Adicionalmente os autores propõem a inclusão de uma fase anterior ao processo de Mineração de Dados, que se refere à identificação do problema a ser abordado, e uma fase posterior ao processo, que se refere à utilização do conhecimento obtido.

Figura 1: Etapas do processo de mineração de dados.



Fonte: Rezende *et al.* (2003) *apud* Melanda (2004)

A mineração de dados (*Data Mining*) é uma ferramenta eficiente para descoberta de informações, pois revela estruturas de conhecimento, que orientam às decisões em condições de certeza limitada. Atualmente existe um crescente interesse no desenvolvimento de novas técnicas de análise de dados, em especial à mineração e seleção de atributos relevantes. No entanto, a mineração ainda está baseada em princípios conceituais de Análise de Dados Exploratórios (*Exploratory Data Analyses – EDA*) e de modelagem.

A mineração de dados representa a principal etapa para a Descoberta de Conhecimento em Bases de Dados - KDD (*Knowledge Discovery in Databases*), a qual integra os conceitos e

processos para a utilização de bases de dados em processos de tomada de decisão, transformados através de processamentos sucessivos de dados brutos, em informações e conhecimentos que poderão ser utilizados e aplicados com assertividade.

De acordo com Fayyad *et al.* (1996), a Descoberta de Conhecimento em Bases de Dados (KDD) é uma área interdisciplinar que envolve conhecimento sobre Bancos de Dados, Inteligência Artificial (Redes Neurais, Lógica Fuzzy, etc.) e estatística, entre outros.

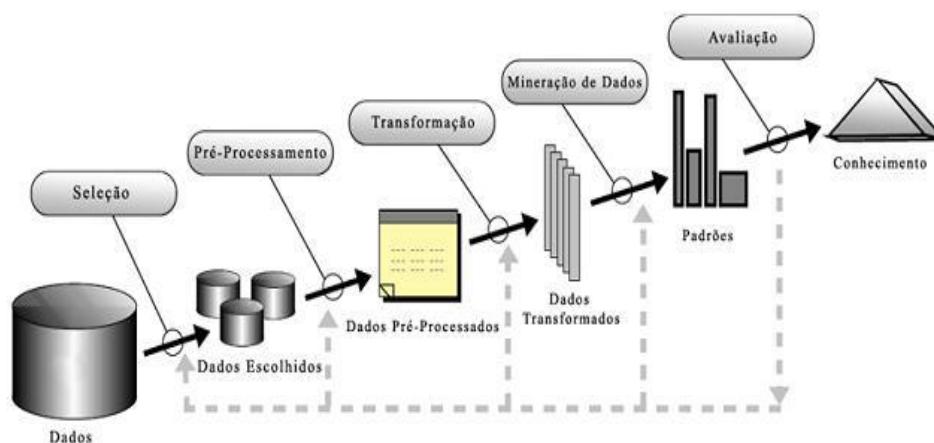
Segundo Melanda (2004), para melhor compreensão do Processo KDD, deve-se considerar individualmente cada componente:

- Dados - Conjunto de fatos ou casos em um repositório de dados. Por exemplo: dados correspondem aos valores dos campos de um registro de vendas em uma base de dados qualquer;
- Padrões - Denota alguma abstração de um subconjunto dos dados em alguma linguagem descritiva de conceitos;
- Processo - A extração de conhecimento de bases de dados envolve diferentes etapas - a preparação dos dados, a busca por padrões, a validação do conhecimento;
- Válidos - Os padrões descobertos devem satisfazer funções ou limiares que garantem os exemplos cobertos e os casos relacionados ao padrão encontrado sejam aceitáveis;
- Novos - Padrões identificados devem fornecer novas informações sobre os dados. O grau de novidade é um indicador da originalidade desses padrões;
- Úteis - Os padrões descobertos devem ser incorporados e utilizados;
- Compreensíveis - Um dos objetivos da realização do processo de Mineração de dados é identificar padrões descritos em alguma linguagem que possa ser compreendida pelos usuários permitindo uma análise mais aprofundada da base de dados;
- Conhecimento - O conhecimento é dependente do domínio da aplicação e está relacionado com a capacidade de inovação de um conceito, sua utilidade e inovação para o usuário.

Fayyad *et al.* (1996) representa o *KDD* através de uma sequência de etapas, que, quando executadas, poderá gerar conhecimento e informações úteis. As etapas de *KDD* estão apresentadas na Figura 2, sendo elas: seleção dos dados utilizados; preparação para a utilização dos dados através de um tratamento prévio (pré-processamento); transformação das informações para um formato apropriado; processamento do conjunto de dados por algoritmos

especialistas (mineração de dados) e a análise dos resultados obtidos para a sua aplicação no processo de decisão (interpretação/avaliação).

Figura 2: Processo de KDD



Fonte: Fayyad *et al.*, 1996

A) Etapa de seleção de dados

O estudo do domínio da aplicação e a definição de objetivos e metas a serem alcançados no processo de mineração de dados são identificados na etapa de seleção de dados e identificação do problema. Também é realizada identificação e seleção dos conjuntos de dados a serem utilizados durante o processo.

Para Melanda (2004), a participação de especialistas do domínio da aplicação fornecendo conhecimento prévio e dando suporte aos analistas na tarefa de identificação de padrões é aspecto determinante no sucesso do processo de extração do conhecimento.

De acordo com Fayyad *et al.* (1996), é imprescindível adquirir conhecimentos iniciais sobre o domínio da aplicação, antes de se proceder a execução das tarefas de Mineração de Dados. O conhecimento subsidia tomadas de decisão nas etapas subsequentes do processo de mineração de dados: pré-processamento, extração de padrões e pós-processamento.

Entender o domínio dos dados é naturalmente um pré-requisito para se extrair conhecimento útil - o usuário final do sistema deve ter algum grau de entendimento sobre a área de aplicação para que quaisquer informações significantes sejam obtidas (MELANDA, 2004).

Desta forma, diversos tipos de banco de dados e de formas de armazenamento podem ser atribuídos no processo de mineração, e de acordo com cada funcionalidade dos dados e de sua disponibilidade, pode-se definir quais padrões ou relações será possível minerar.

A etapa de seleção se refere justamente na análise da disponibilidade e relevância dos dados existentes em diferentes fontes. Quando escolhida de forma errônea o resultado se torna impreciso, prejudicando a tomada de decisão.

B) Etapa de pré-processamento

As fontes de dados disponíveis para Mineração de Dados podem ser encontradas nos mais diversos formatos - arquivos texto, planilhas, banco de dados, *Data Warehouse* (armazenamento de dados), dentre outros. Nesse sentido, é necessária a obtenção desses dados e sua unificação, constituindo-se assim uma única fonte de dados no formato atributo-valor.

Outro aspecto relevante se refere aos problemas gerados no processo de coleta de dados, tais como erros de digitação ou de leitura de dados por sensores, podem originar dados incorretos ou inconsistentes que necessitam de tratamento. Limitações na capacidade de memória ou no tempo de processamento podem impossibilitar a aplicação direta dos algoritmos de extração de padrões a todo o conjunto de dados. Nesse sentido, torna-se importante a aplicação de métodos de tratamento, limpeza, redução do volume de dados, antes de iniciar a etapa de extração de padrões (MELANDA, 2004).

Ao selecionar os dados a serem utilizados para mineração, é necessário prepará-los para o modelo. Essa etapa de tratamento é denominada de pré-processamento, o qual pode ser realizado de diferentes formas e aspectos, deixando-os com forma adequada para a mineração.

Inicialmente são estabelecidas normas de representação dos dados utilizados, assim como a padronização das informações provenientes de diferentes fontes. De acordo com a técnica de mineração escolhida para ser aplicada, é realizada a definição as formas de representação da base de dados. Após a sua definição, os dados identificados são extraídos das suas diversas fontes, e integrados em um único repositório. A fase de pré-processamento é constituída por diversas atividades de tratamento dos dados selecionados, tais como verificação do significado, enriquecimento das informações, remoção de dados duplicados, unificação e discretização.

C) Etapa de transformação de dados

A etapa de transformação dos dados é a fase que antecede a fase de mineração de dados (*Data Mining*). Após os dados de interesse serem selecionados e pré-processados, é necessário que sejam armazenados e formatados de forma adequada para que os algoritmos possam ser aplicados.

Devido aos diferentes tipos de sistemas operacionais e Sistemas Gerenciadores de Banco de Dados (SGDB) que são utilizados em grandes corporações, os dados estão dispersos e é necessário ser agrupados em um repositório único.

D) Etapa de mineração de dados (*Data Mining*)

A fase de *Data Mining* é a etapa em que é realizada a exploração e análise dos dados obtidos. Devido à grande quantidade de dados disponíveis, a mineração é realizada de forma semiautomática, a fim de selecionar os padrões e regras que os dados obedecem.

Todas as etapas do processo são importantes para o sucesso da análise dos dados. Entretanto, é a etapa de Mineração de Dados (data mining) que recebe o maior destaque na literatura. Fayyad é considerado o “pai” do processo como um todo, porém a definição mais completa de *Data Mining* foi dada por Michael Berry Gordon Linoff, autor do livro *Data Mining Techniques: For Marketing, Sales, and Customer Support*. De acordo com os autores:

“Data Mining é a exploração e análise, de forma automática ou semi-automática, de grandes bases de dados com objetivo de descobrir padrões e regras. O objetivo do processo de mineração é fornecer às corporações informações que as possibilitem montar melhores estratégias de marketing, vendas, suporte, melhorando assim os seus negócios” (BERRY, M. J. A.; LINOFF, G., 1997, p. 07).

E) Etapa de Interpretação e Avaliação

A interpretação e avaliação dos dados obtidos através da mineração é uma etapa muito importante para que o objetivo final da pesquisa seja alcançado. Quando não obtido o resultado de forma satisfatória, é possível retornar a qualquer fase anterior ou até mesmo recomeçar, a fim de analisar e ajustar o banco de dados para um melhor resultado. Duas ações podem ser

feitas caso o produto final não se apresente satisfatório, a primeira opção é modificar o conjunto de dados inicial, a outra opção seria trocar o algoritmo de data mining (ou ao menos alterar suas configurações de entrada).

F) Diferentes técnicas de mineração de dados

Não existe ao certo uma técnica que consiga resolver todos os problemas de mineração de dados. Cada técnica deve ser usada para um propósito específico, pois possuem características particulares, vantagens e desvantagens. As técnicas de mineração estão fortemente relacionadas com o tipo de conhecimento que se deseja extrair, ou com o tipo de dado em que será aplicada. Desta forma, abaixo estão descritas as principais técnicas de data mining que serão utilizadas nesta pesquisa:

2.2.1. Árvores de decisão

A técnica de árvore de decisão é um modelo preditivo que pode ser visualizado na forma de uma árvore. Cada ramo da árvore representa uma classificação, onde cada folha é uma partição do conjunto de dados separados de acordo com a sua classificação.

Segundo Bigolin *et al.* (2003), a Árvore de decisão é uma técnica de aprendizagem que classifica exemplos em um número finito de classes, onde o objetivo é elaborar e testar hipóteses sem a intervenção do usuário.

A representação gráfica de uma árvore é dada por:

- um nó intermediário: corresponde a uma pergunta sobre a propriedade do objeto, representando um atributo.
- um arco: corresponde a um valor de atributo.
- um nó final corresponde a um conjunto da mesma classe.

Percorrendo a árvore (respondendo as questões dos nós intermediários e seguindo as ligações correspondentes), chega-se a um nó final que é a classe do objeto.

Para Zadeh (1965) *apud* Bigolin *et al.* (2003), a arquitetura proposta baseia-se em árvores de decisão nebulosas, que são um tipo especial de árvore. Elas são baseadas na teoria de conjuntos nebulosos, diferenciando-se das árvores de decisão tradicionais em dois aspectos: utilizam critérios de partição imprecisos e, referem-se ao modelo de raciocínio aproximativo. Nas árvores de decisão nebulosas, uma pergunta é realizada sobre um valor nebuloso, gerando

um resultado nebuloso. Como exemplo de valor nebuloso podemos citar: grande, próximo, longe, etc. Os dados de entrada desta técnica são valores nebulosos e a saída é uma árvore de decisão nebulosa. As árvores de decisão tradicionais utilizam uma condição precisa, onde um valor satisfaz uma única condição.

Ao selecionar um conjunto de dados, o usuário escolhe uma das variáveis como objeto de saída. Desta forma, o algoritmo descobre o fator mais relevante correlacionado com a variável de saída, em seguida já o aloca como o primeiro ramo (chamado de raiz).

Os demais fatores são posteriormente classificados como nós, até chegar ao último nível, a folha. Contudo, a árvore de decisão utiliza a estratégia de dividir um problema complexo em subproblemas mais simples, e, repetitivamente, essa mesma estratégia é aplicada a cada subproblema. Nesta pesquisa, foi utilizado o algoritmo *Random Forest*, o qual produz opcionalmente duas informações adicionais: uma medida da importância das variáveis preditoras, e uma medida da estrutura interna dos dados (a proximidade de dados diferentes apontam uns para os outros).

Por ter suas origens na forma mais básica e inicial de um algoritmo de suporte à decisão, o *Random Forest* resolve problemas tanto de regressão quanto de classificação.

O *Random Forest*, ou *floresta aleatória*, possibilita selecionar subconjuntos de *features* e apelar diversas miniárvores de decisão. Devido ao fato de possuir suas origens na forma mais básica de um algoritmo de suporte a decisão (as árvores de decisão), o *Random Forest* soluciona problemas que necessitem tanto da técnica de regressão quanto da classificação.

Para aplicação do algoritmo *Random Forest* deve-se percorrer quatro etapas:

1. Selecionar quais *features* irá utilizar (aleatórias);
2. Selecionar a *feature* para posição de nó raiz (principal);
3. Aplicar o algoritmo para geração dos nós secundários;
4. Repetir novamente os passos anteriores para calibração do modelo até que se obtenha o resultado almejado.

Após alcançar o modelo mais adequado, as previsões são obtidas através de “votações”. Cada miniárvore toma uma decisão a partir dos dados calibrados. A resposta do algoritmo apresenta a decisão mais votada.

Contudo, o algoritmo *Random Forest* possibilita a combinação de previsões de múltiplos algoritmos de *machine learning* em conjunto, de forma a alcançar resultados mais acurados do que qualquer modelo individual.

2.3. Método Comparativo de Dados de Mercado

O método comparativo de dados de mercado define o valor do imóvel ou da propriedade através da comparação com valores de mercado, com as características intrínsecas e extrínsecas semelhantes dos imóveis.

Para que esse método seja aplicado é necessário que se tenha um conjunto de dados disponíveis para análise estatística, como, por exemplo, levantamento de valores imobiliários e os atributos das regiões as quais estão inseridos. A aplicação do método de regressão linear permite estimar o valor de um imóvel a partir de uma amostra de dados de mercado, representativa da população (HOFFMAN & VIEIRA, 1977).

2.4. Valores hedônicos de imóveis

De acordo com Melanda (2004), o estabelecimento do valor hedônico de imóvel ou propriedade consiste na identificação de atributos ou características de um bem ambiental (bonita paisagem, alto risco, etc.) e sua relação com o preço da terra ou do trabalho. Existem duas técnicas para esse tipo de avaliação: diferenças nos preços dos imóveis e diferenças nos níveis salariais. O valor de uma casa, por exemplo, tem relação com diversas variáveis como área, padrão de acabamento, número de quartos, localização, etc. e pela qualidade ambiental do seu entorno, tais como vista panorâmica, poluição do ar ou sonora. As variações nos preços dos imóveis derivadas destas variáveis ambientais podem ser utilizadas para medi-las. Outra possibilidade de aplicação deste método é a utilização de diferenças salariais para trabalhos com diferentes graus de risco. Neste caso, salários maiores para trabalhar em áreas poluídas podem ser utilizados para valorar impactos ambientais.

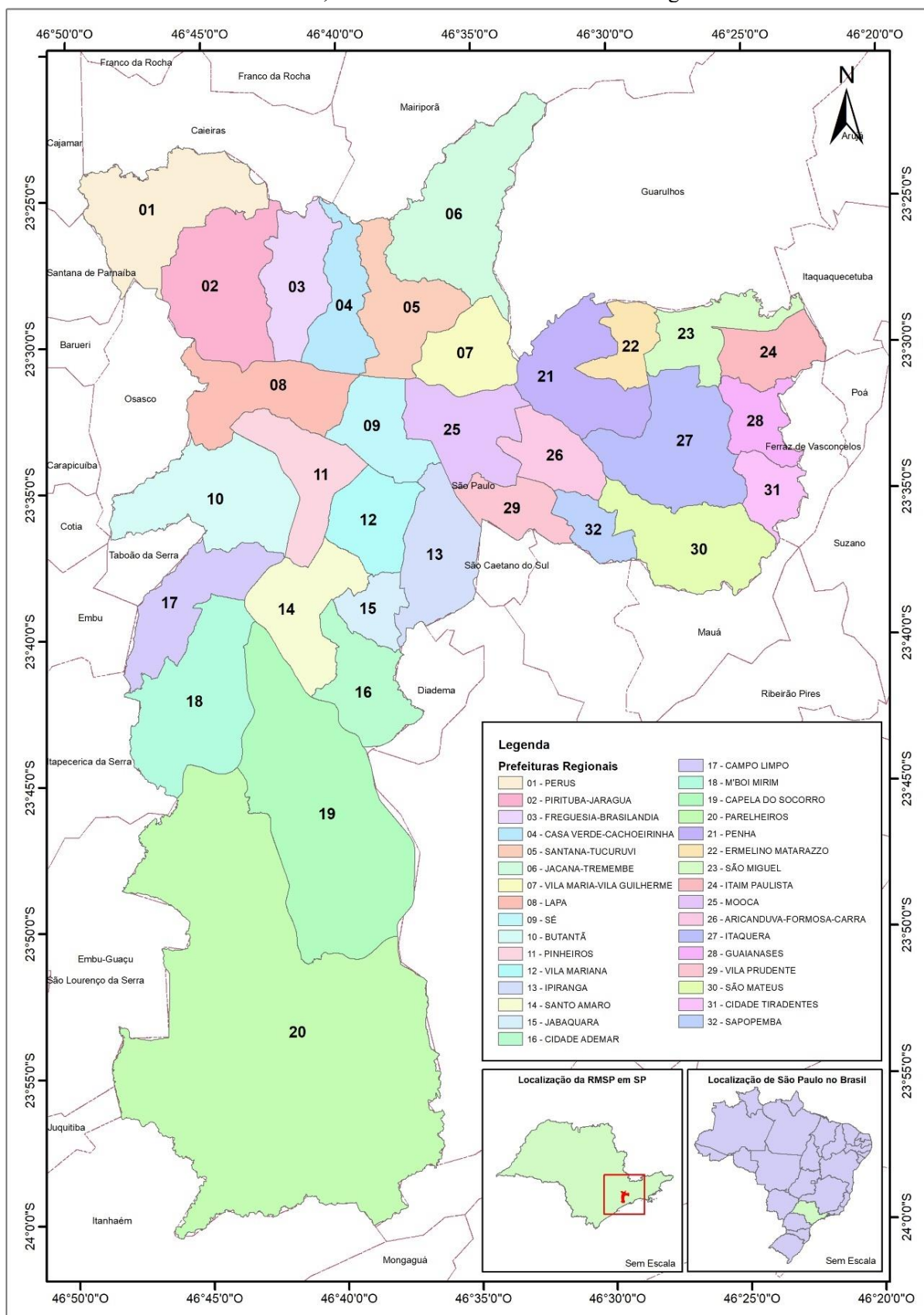
O importante no método dos valores hedônicos é segregar claramente o efeito ambiental dos gerados por outros fatores (localização com relação a serviços públicos, benfeitorias do próprio imóvel, etc.), o que requer a coleta de informações sobre numerosos imóveis semelhantes com e sem a influência ambiental que se quer mensurar.

2.5. Descrição da Área de Estudo

São Paulo é um município brasileiro, capital do estado de São Paulo e principal centro financeiro, corporativo e mercantil da América do Sul.

O município de São Paulo está, administrativamente, dividido em trinta e duas regiões administradas por prefeituras regionais (Figura 3), cada uma delas, por sua vez, divididas em distritos, sendo estes últimos, eventualmente, subdivididos em subdistritos (a designação "bairro", porém, não existe oficialmente, embora seja usualmente aplicada pela população). Os atuais distritos foram criados pela lei municipal nº 11.220 de 20 de maio de 1992, e as atuais subprefeituras pela Lei Municipal nº 13.399, de 1 de agosto de 2002. As subprefeituras estão oficialmente agrupadas em nove regiões (ou "zonas"), levando em conta a posição geográfica e história de ocupação. Entretanto, há certos órgãos e instituições (companhias telefônicas, zonas eleitorais, etc.) que adotam uma divisão diferente.

Figura 3: Localização das trinta e duas regiões administradas (Prefeituras Regionais), na Escala 1:280.000, Sistema de Coordenadas WGS84/Sirgas.



Elaboração: Autor, 2019.

A explosiva expansão urbana da Região Metropolitana de São Paulo (RMSP) vem acompanhada de conflitos relacionados à mineração, ao abastecimento público de águas (quantidade e qualidade), às áreas de riscos geológicos e/ou geotécnicos, à precária ocupação habitacional de favelas, aos colapsos de terrenos cárnicos, à disposição de resíduos sólidos, à vulnerabilidade por contaminação de aquíferos subterrâneos, entre outros.

2.5.1. Características ambientais

O clima do município de São Paulo é considerado subtropical úmido do tipo Cfa segundo Köppen, redução de chuvas no inverno e temperatura média anual de 19,2°C, tendo invernos frescos (com média de 15,8 °C no mês mais frio) e verões com temperaturas moderadamente altas (média de 22,4 °C no mês mais quente), resultante do efeito da poluição atmosférica e da grande concentração de edifícios (efeito estufa) (CLIMATEMPO, 2019).

Em relação à temperatura, a região metropolitana de São Paulo está entre as três cidades mais frias do Brasil. Ao levar em consideração a temperatura média anual, a cidade recebe o título de segunda capital mais fria. Ao adotar como referência a temperatura média do mês mais frio, a cidade recebe o título de terceira mais fria. Devido à proximidade do mar, a maritimidade é uma constante do clima local, sendo responsável por evitar dias de calor intenso no verão ou de frio intenso no inverno e tornar a cidade mais úmida, com temperaturas mínimas que raramente são superiores a 23 °C em um período de 24 horas no verão (SMVMA, 1993).

Um estudo realizado pela Secretaria Municipal do Verde e Meio Ambiente, com o apoio do Instituto Oceanográfico da Universidade de São Paulo (USP), indica que a diferença de temperatura, entre o centro e as áreas mais afastadas do centro, pode chegar a até 10 °C, resultado do efeito estufa (“ilhas de calor”), que ocorre devido à quantidade de prédios, asfalto, concreto, e pelo número reduzido de áreas verdes na região central da cidade, recorrendo em dias quentes e secos durante o inverno.

Em relação à vegetação, São Paulo era um ecótono – uma região de transição entre diferentes vegetações – aqui existia um mosaico de formações, e não um tapete contínuo de Mata Atlântica como aquela que vemos na Serra do Mar.

A Mata de Araucária, originária do Sul, o Cerrado, do Centro-Oeste, a Mata Atlântica de Encosta, da Serra do Mar, e a Mata Semidecidual, do interior do Estado, se encontravam nessa área, em um fluxo de diferentes floras. O resultado foi uma área com grande diversidade

de paisagens. Hoje, raros são os locais dentro da Cidade onde encontramos resquícios do antigo mosaico de vegetação.

A abertura de campos de cultivo, a formação de paisagens artificiais e o uso da madeira como combustível ou como matéria-prima, levaram a uma quase completa destruição das florestas paulistas. Essa cobertura florestal revestia outrora cerca de 82% da área estadual; atualmente subsistem florestas em áreas impróprias para pastagens e culturas (como nas encostas íngremes das serras do Mar e da Mantiqueira), ou ainda não incorporadas à economia agropastoril (como algumas porções do extremo oeste).

Nessas áreas, onde se fez sentir com menos intensidade a interferência humana, ainda é possível observar as características originais da floresta e reconhecer a marca de sua diferenciação regional. Ao longo da costa e no rebordo do planalto, graças à elevada pluviosidade, a floresta se apresentava como uma mata densa, perene e muito rica em epífitas e lianas.

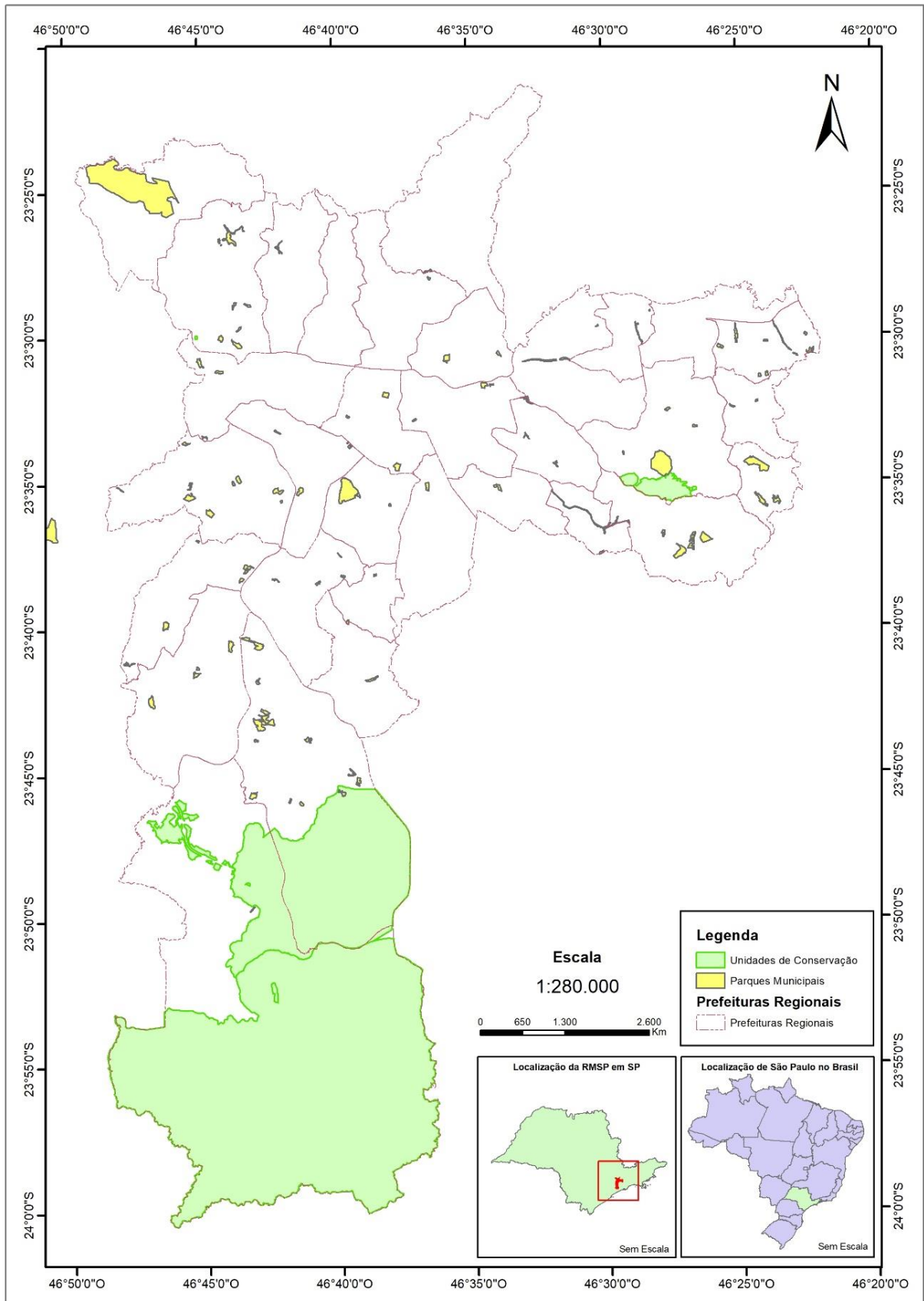
No interior do planalto, a umidade mais reduzida resultava no desenvolvimento de uma Floresta Estacional Semidecidual, onde contrastavam as formações desenvolvidas em solos de arenito com as formações mais exuberantes, dos solos de terra roxa.

Nos trechos mais elevados da Mantiqueira e da Bocaina, a floresta assumia caráter mais subtropical, assinalado pela presença do Pinheiro-do-Paraná (*Araucária*). A mata de pinheiros aparecia, também, em manchas dispersas no leste do estado, das quais a maior situava-se nas proximidades de Botucatu.

Outros tipos de vegetação existiam em São Paulo. Campos cerrados (10% da superfície total) formavam manchas dispersas no interior do planalto, sobretudo na margem oriental do planalto ocidental e na depressão interior. A sul dessa última, aparecia, como ainda hoje aparece, a extremidade setentrional dos Campos Gerais do Paraná, ocupando 1,3% da área estadual.

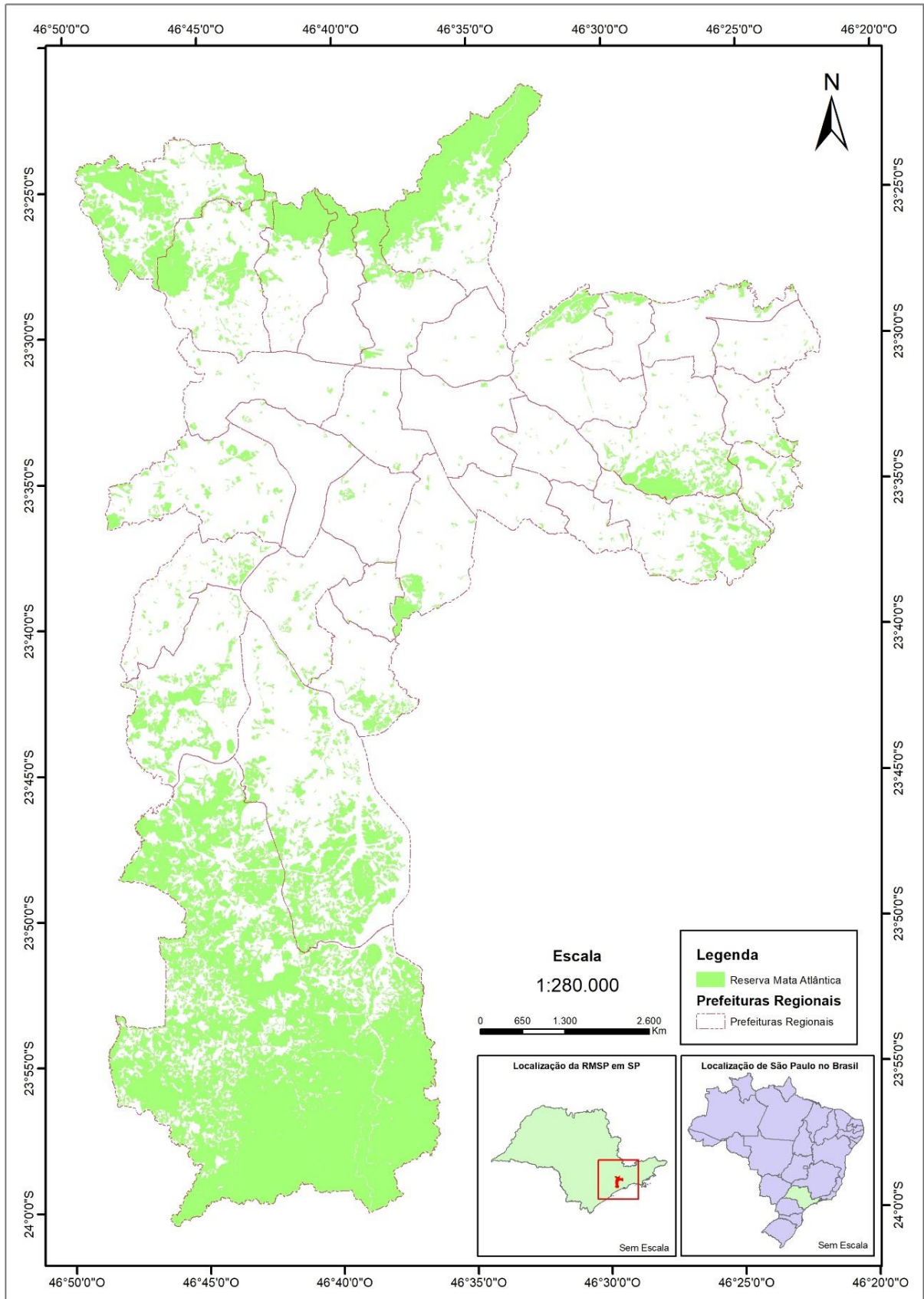
As Figuras 4 a 7 apresentam a distribuição dos recursos naturais, da arborização urbana, das Unidades de Conservação, parques e reservas da Mata Atlântica no município de São Paulo.

Figura 4: Localização dos Parques Municipais e Unidades de Conservação em São Paulo-SP.



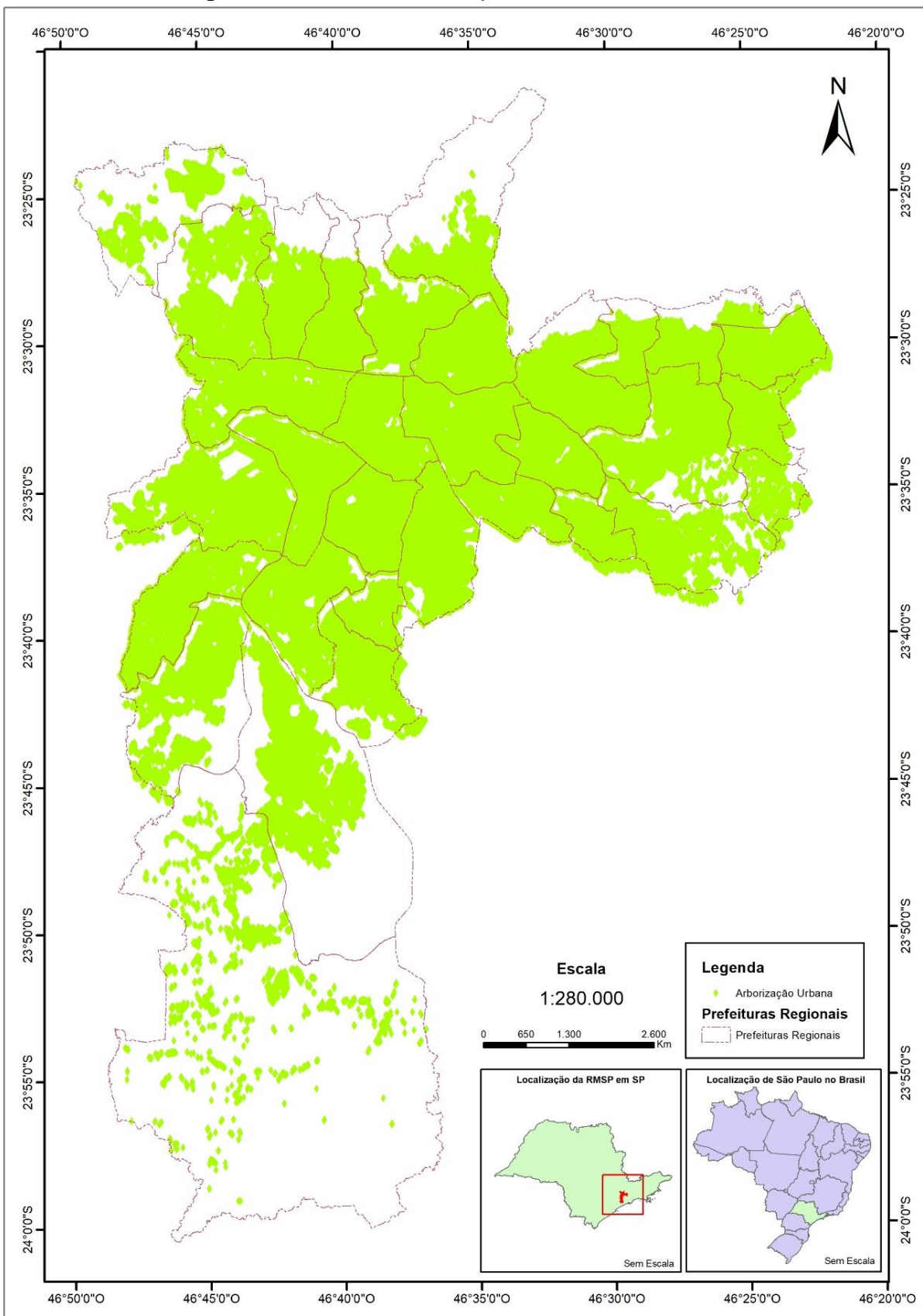
Elaboração: Autor, 2019.

Figura 5: Localização das Reservas de Mata Atlântica em São Paulo-SP.



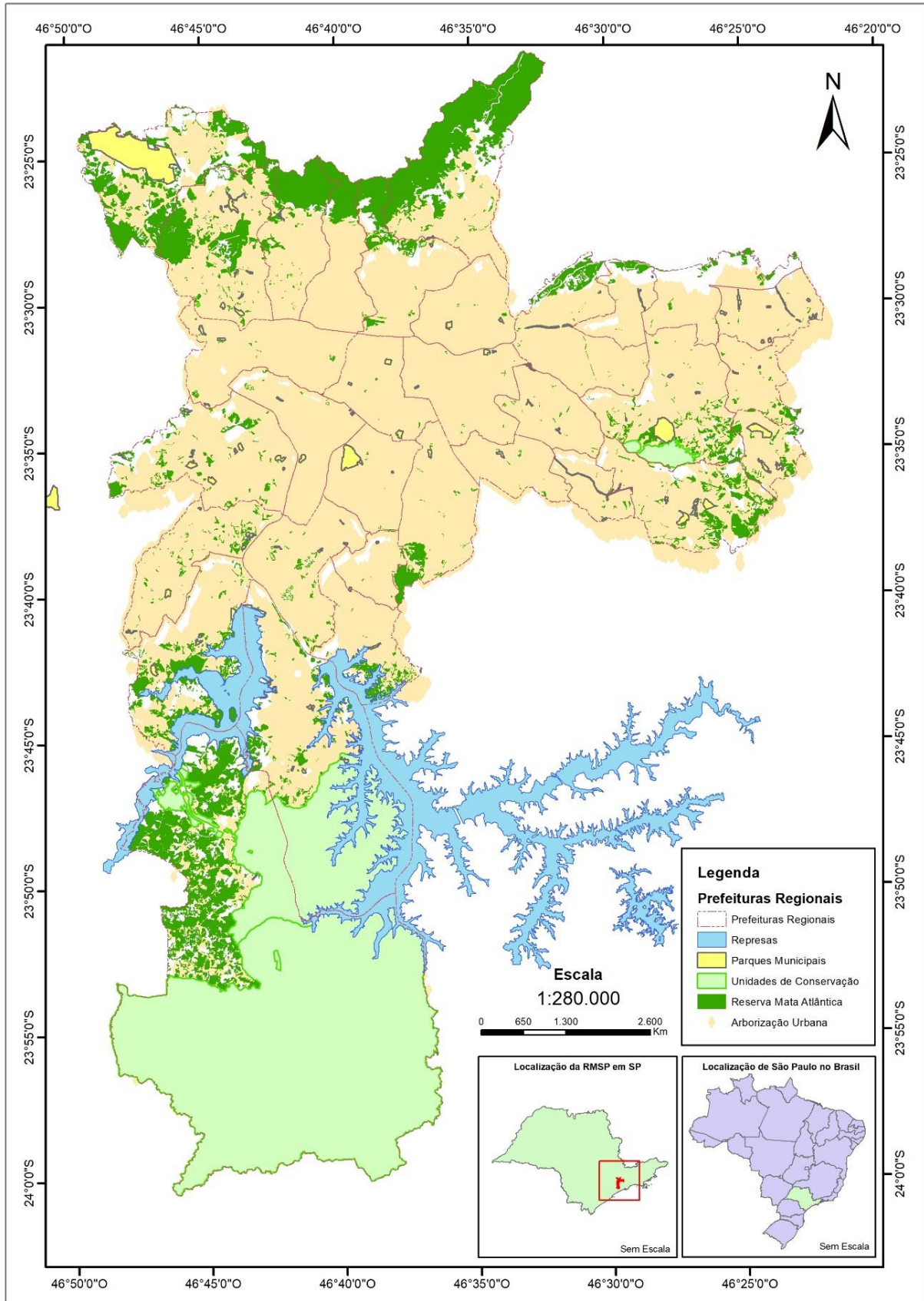
Elaboração: Autor, 2019.

Figura 6: Densidade da arborização urbana em São Paulo-SP.



Elaboração: Autor, 2019.

Figura 7: Verde e Recursos Naturais em São Paulo-SP.



Elaboração: Autor, 2019.

Conforme pode-se verificar nos mapas, as unidades de conservação e reservas de Mata Atlântica estão localizadas nas regiões periféricas do município. Os parques municipais possuem pequena densidade e escassez em muitas regiões administrativas. Em relação à arborização urbana, a região central do município apresenta uma densidade considerável, porém não se pode afirmar que é suficiente para minimizar as ilhas de calor e efeito estufa ocasionados pela dinâmica da cidade.

2.5.2. Aspectos Sociais

São Paulo foi a capital brasileira que mais cresceu em todo o século XX, atingindo a marca de um milhão de habitantes na década de 1930 e se configurando como o município mais populoso do Brasil desde 1960, quando ultrapassou o Rio de Janeiro em população.

De acordo com informações do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), a população de São Paulo em 2010 era de 11.253.503 pessoas, apresentando uma densidade populacional de 7.765,06 hab./km .

Ainda segundo o mesmo censo, 11.152.344 habitantes viviam na zona urbana (99,1%) e 101.159 na zona rural (0,9%). Naquele ano, o distrito mais populoso de São Paulo era Grajaú, com 360.787 habitantes, e Marsilac, no extremo sul do município, o menos populoso, com uma população de 8.258 pessoas. Para 2018, a estimativa populacional é de 12.176.866 habitantes.

O Quadro 1 apresenta a população por Prefeitura Regional de São Paulo, com suas respectivas áreas e densidade populacional. Os dados de densidade demográfica e localização das favelas estão representados na Figura 8.

Quadro 1: Densidade Populacional por Prefeitura Regional, com suas respectivas áreas.

Subprefeituras	Área (km)	População (2010)	Densid. Demográfica (Hab/km)
Aricanduva	21,50	267.702	12.451
Butantã	56,10	428.217	7.633
Campo Limpo	36,70	607.105	16.542
Capela do Socorro	134,20	594.930	4.433

(Continuação)

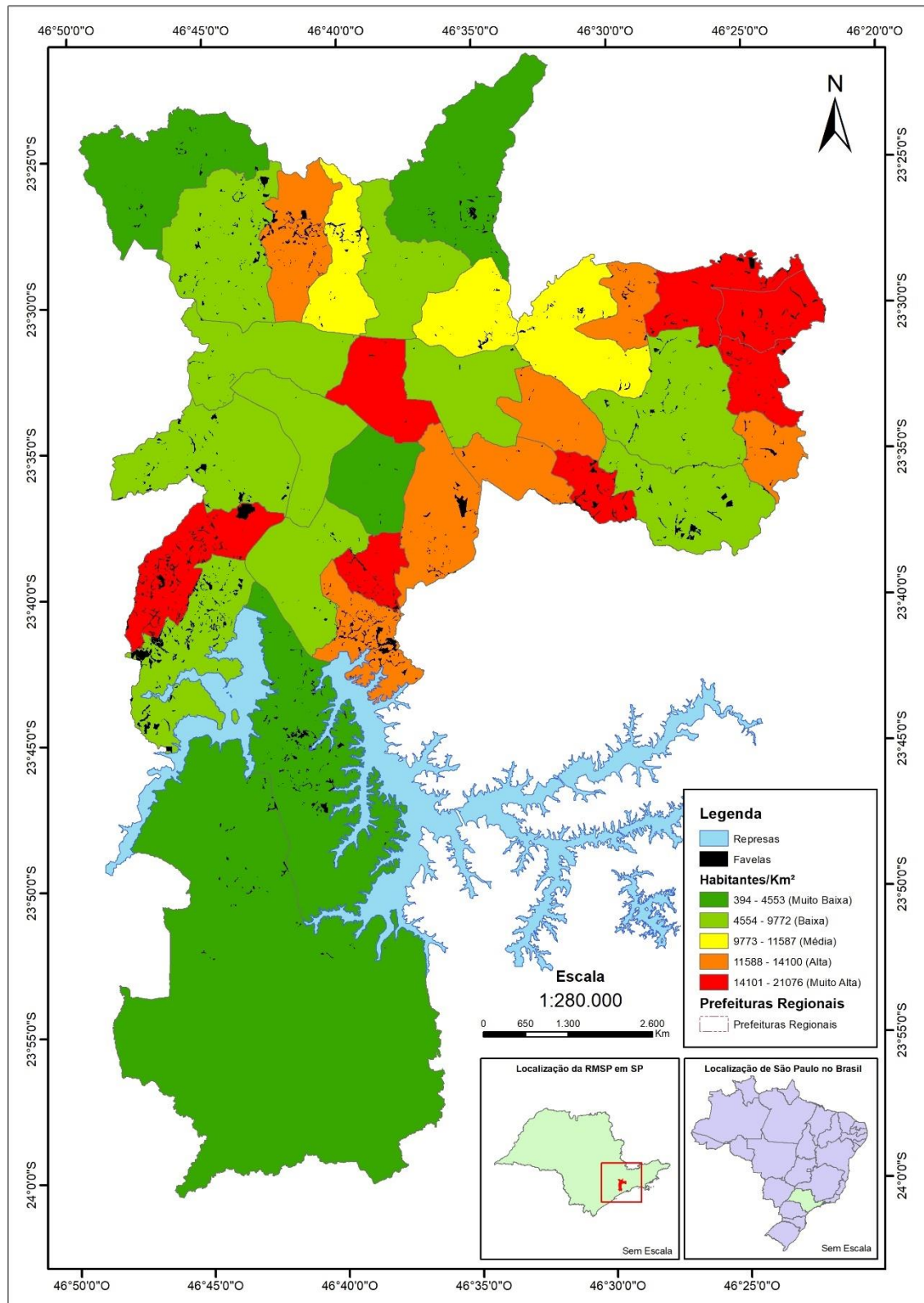
Subprefeituras	Área (km)	População (2010)	Densid. Demográfica (Hab/km)
Casa Verde	26,70	309.376	11.587
Cidade Ademar	30,70	410.998	13.388
Cidade Tiradentes	15,00	211.501	14.100
Ermelino Matarazzo	15,10	207.509	13.742
Freguesia do Ó	31,50	407.245	12.928
Guaianases	17,80	268.508	15.085
Ipiranga	37,50	463.804	12.368
Itaim Paulista	21,70	373.127	17.195
Itaquera	54,30	523.848	9.647
Jabaquara	14,10	223.780	15.871
Jaçanã	64,10	291.867	4.553
Lapa	40,10	305.526	7.619
M'Boi Mirim	62,10	563.305	9.071
Mooca	35,20	343.980	9.772
Parelheiros	353,50	139.441	394
Penha	42,80	474.659	11.090
Perus	57,20	164.046	2.553
Pinheiros	31,70	289.743	9.140
Pirituba	54,70	437.592	8.000
Santana	34,70	324.815	9.361

(Continuação)

Subprefeituras	Área (km)	População (2010)	Densid. Demográfica (Hab/km)
Santo Amaro	37,50	238.025	6.347
São Mateus	45,80	426.794	9.319
São Miguel	24,30	369.496	15.206
Sapopemba	13,50	284.524	21.076
Sé	26,20	431.106	16,454
Vila Maria-Vila Guilherme	26,40	297.713	11,277
Vila Mariana	26,50	344.632	13,005
Vila Prudente	19,80	246.589	12.454

Fonte: Adaptado de Prefeitura Municipal de São Paulo (Infocidade), 2019.

Figura 8: Dados de densidade demográfica e localização das favelas nas Prefeituras Regionais de São Paulo-SP.



Elaboração: Autor, 2019.

O mapa acima representa a densidade demográfica por subprefeitura na cidade. Observa-se as maiores concentrações populacionais definido pelo numero de habitantes por km na região do Itaim Paulista (17.195), seguido por Campo Limpo (16.652), Sé (16.454), altamente edificada, Jabaquara (15.871) e Guaianases (15.085). Das prefeituras regionais com menores taxas de densidade estão: Palheiros (394), Perus (2.553), Capela do Socorro (4.433), Jaçanã/Tremembé (4.553) e Santo Amaro (6.347). É importante ressaltar que grande parte das regiões com menores taxas de densidade demográfica estão protegidas por unidades de conservação, como o Parque Estadual Serra da Cantareira e o Anhanguera ao norte, o Parque Estadual Serra do Mar, APAs (Áreas de Proteção Ambiental) e parques municipais oriundos de compensação ambiental no entorno do Rodoanel.

2.5.3. Características Econômicas

A economia da cidade de São Paulo, com um orçamento anual de R\$ 42 bilhões e arrecadação de mais de R\$ 17 bilhões, o Produto Interno Bruto (PIB) da capital paulista é bastante expressivo (cerca de 389 bilhões de reais) (IBGE, 2017).

Nos últimos anos, São Paulo tem passado por uma nítida transformação em sua economia. Durante muito tempo a indústria constituiu uma atividade econômica bastante presente no município.

Segundo dados da Agência Paulista de Promoção de Investimentos e Competitividade (2011), a composição econômica da cidade de São Paulo é composta por:

- Serviços: 69,0%
- Agropecuária: 2,0 %
- Indústria: 29,0%

Porém, São Paulo, nas últimas três décadas, teve uma clara mudança em seu perfil econômico: de uma cidade com forte caráter industrial, o município tem cada vez mais assumido um papel de cidade terciária, polo de serviços e negócios para o país.

O turismo cultural também se destaca na metrópole, devido ao número de museus, teatros, centros culturais e salas de concerto.

Segundo dados do IBGE, em 2015, o salário médio mensal era de 4,4 salários mínimos. A proporção de pessoas ocupadas em relação à população total era de 48,9%. Na comparação com os outros municípios do estado, ocupava as posições 8 de 645 e 23 de 645,

respectivamente. Já na comparação com cidades do país todo, ficava na posição 28 de 5.570 e 79 de 5.570, respectivamente. Considerando domicílios com rendimentos mensais de até meio salário mínimo por pessoa, tinha 31,6% da população nessas condições, o que o colocava na posição 305 de 645 dentre as cidades do estado e na posição 4.372 de 5.570 dentre as cidades do Brasil (IBGE, 2017).

2.5.4. Infraestrutura Urbana

Internamente, o Município de São Paulo é subdividido em 32 subprefeituras administrativas que exercem um papel de atendimento local compartilhado com Poder Executivo Municipal. Cabe às Subprefeituras: fortalecer as formas participativas; participar da coordenação de planos regionais; estabelecer formas articuladas de ação local; ampliar a oferta dos serviços municipais e praticar a gestão, direção e controle dos assuntos municipais em nível local, respeitadas as legislações vigentes e observadas as prioridades estabelecidas pela Prefeitura.

Segundo o diagnóstico da mobilidade urbana realizado pela Prefeitura Municipal de São Paulo, o processo de urbanização foi marcado por uma progressiva motorização que espelhava a implantação da indústria automobilística. Desde a década de 1950 as políticas federais, estaduais e municipais relativas ao transporte urbano incorporaram uma postura francamente favorável à fabricação e uso do automóvel. Pode ser incluído nessa política a construção de rodovias, a abertura de grandes vias expressas, o sucateamento de redes de transporte coletivo como a rede de bondes urbanos, a separação de fluxos veiculares por meio de túneis, pontes e viadutos, a adoção de uma engenharia de tráfego voltada à fluidez dos automóveis, os programas de financiamento para aquisição de veículos, o programa de incentivo ao uso do álcool combustível (etanol), subsídios aos combustíveis e às facilidades dadas ao estacionamento e provisão de vagas, entre outras posturas (PlanMob/SP, 2015).

No mesmo período e no caso específico de São Paulo, além das diversas ações de apoio ao transporte individual, à baixa eficiência e qualidade do transporte coletivo sobre pneus aliada à lenta expansão da rede de alta capacidade sobre trilhos estimularam o uso crescente do veículo particular. O resultado dessa política levou a uma divisão modal desequilibrada do transporte motorizado. No final da década de 1960 a divisão modal que apresentava 30% de viagens individuais (auto) contra 70% por transporte coletivo, condição hoje considerada ideal, passou para uma situação inversa na virada do século, quando menos da metade das viagens era feita

por coletivos e o transporte individual motorizado se tornou majoritário. A prática tradicional na mobilidade urbana resultou em efeitos negativos para toda a sociedade.

Dentre todos os fatores que contribuem para a baixa atratividade do transporte público, destaca-se o percentual muito alto de ônibus que utilizam vias compartilhadas. De um total aproximado de 4.500 km de vias por onde circulam ônibus aproximadamente 87% são em vias onde o espaço é compartilhado com outros modos, ou seja, não são corredores ou faixas exclusivas, de forma que os coletivos sofrem interferência do tráfego, perdendo velocidade em relação ao individual e tornando o transporte coletivo menos atrativo. Desse modo, a ampliação do uso de faixas exclusivas emerge uma resposta necessária em muitos casos.

Comparando as taxas de ocupação do viário pelo transporte público e pelo individual, a solução para problemática da mobilidade do município se dá pela melhoria no transporte público urbano, com destinação de mais faixas de fluxo viário, e desta forma torná-lo mais atrativo quando comparado ao transporte individual. Assim, no contexto da formulação de prioridades do PlanMob/SP (2015) foi realizada a ampliação das faixas de uso exclusivo do transporte público coletivo. Com a redução do espaço viário para os automóveis, devem ser adotadas medidas de gestão de trânsito que possam tornar mais eficiente o uso do viário que restou ao tráfego geral, reduzindo assim o impacto entre os usuários cativos do transporte individual.

3. METODOLOGIA

Com relação ao caráter desta pesquisa, o trabalho desenvolvido é exploratório e aplicado. De acordo com Gerhardt e Silveira (2009), a pesquisa exploratória tem como objetivo conhecer melhor a problemática, assim como possíveis técnicas para encontrar soluções, com vistas a torná-las mais explícitas ou a construir hipóteses. No que tange a pesquisa aplicada, esta tem como motivação a necessidade de produzir conhecimento para aplicação de seus resultados, com fins práticos, visando à solução mais ou menos imediata do problema encontrado na realidade (BARROS; LEHFELD, 2000).

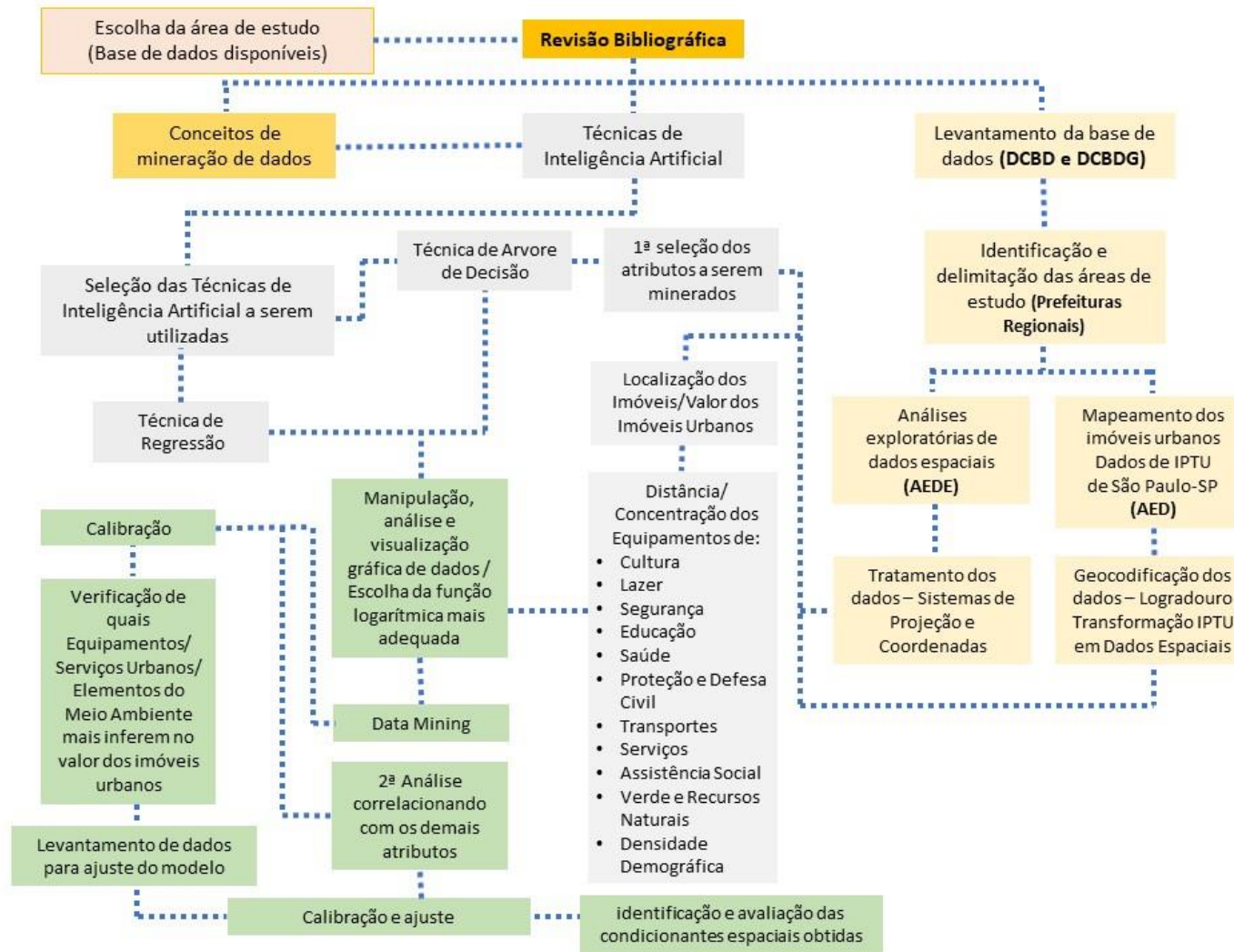
O desenvolvimento da pesquisa encontra-se alicerçado com base nas premissas expressas na literatura obtida, assim como na análise dos resultados obtidos através da mineração de dados e da distribuição infraestrutural na cidade de São Paulo, com o valor do solo urbano aonde se encontram.

Será considerada infraestrutura, fatores que vão desde os serviços até a disponibilidade de equipamentos culturais, transporte, saúde, meio ambiente, dentre outros.

Foram levantados dados de 21 categorias espaciais, totalizando 57 subcategorias de variáveis (equipamentos e serviços urbanos e aspectos ambientais), nas 32 subprefeituras de São Paulo, totalizando mais de 2,1 milhões de registros.

A partir de uma planilha nominada “IPTU.2019.csv” e baixada no portal “GeoSampa”, foi obtido o valor da área construída e do metro quadrado do terreno de cada contribuinte na cidade. A Figura 9 apresenta um fluxograma das etapas de pesquisa, que serão detalhadas em seguida.

Figura 9: Processo metodológico



Elaboração: Autor, 2021.

3.1. Etapas para tratamento dos dados

Para identificação, análise, sistematização e tratamento dos dados obtidos referente ao município de São Paulo, adotou-se primeiramente como unidades de estudo as 32 regiões administrativas denominadas como Prefeituras Regionais. Desta forma, através da pesquisa bibliográfica e da Descoberta do Conhecimento de Bancos de Dados Geográficos (DCBDG), foi possível obter mais de 2,1 milhões de registros referentes à equipamentos, serviços urbanos e elementos do meio ambiente das 32 Prefeituras Regionais. Os dados foram obtidos por meio do site da Internet Geosampa (Interface de consulta de dados espaciais da Prefeitura de São Paulo, implementado em 2015), os quais estão apresentados no Quadro 2, a seguir:

Quadro 2: Número de registros referentes à 20 atributos principais e 74 atributos secundários das 32 Prefeituras Regionais de São Paulo-SP.

Atributo Principal	Atributo Secundário	Nº de Registros
Cadastro	IPTU	1.048.575
	Uso do solo	54.821
Equipamentos Cultura	Espaços Culturais	320
	Museus	135
	Outros	7
	Bibliotecas	171
	Teatro, Cinema, Shows	1.058
Declividade	Declividade	24.637
Equipamentos Educação	CEU	184
	Ensino Técnico Público	47
	Educação rede Privada	2.397
	Outros	110
	Rede Pública Educação Infantil	2.526
	Rede Pública Ensino Fundamental e Médio	1.590
	Sesi, Senai, Sesc	43
Equipamentos Abastecimento	Feiras Livres	883
	Mercados Municipais	16
	Sacolão	18
Equipamentos Assistência Social	Equipamentos Assistência Social	59
Equipamentos Esporte	Centro Esportivo	41
	Clube	80
	Clube da Comunidade	293
	Estádios	9
	Outros	3
Habitação	Favelas	2.195
IDH (Excel)	2000-2010 Unidades de Desen. Humano	3.187

(continuação)

Atributo Principal	Atributo Secundário	Nº de Registos
Infraestrutura Urbana	Linhas de Alta Tensão	2.242
	Reservatórios	65
Legislação Urbana	Perímetros de Incentivo	20
	Utilidade Pública - Desapropriação	369
População	Setores Censitários	18.953
População	Prefeituras Regionais	32
Proteção e Defesa Civil	Risco de ocorrências	3.282
	Risco Geológico	1.183
Equipamentos de Saúde	Ambulatórios Especializados	157
	Hospital	228
	Saúde Mental	113
	UBS (Posto e Centro de Saúde)	455
	Unidades DST / AIDS	27
	Urgência / Emergência	49
Equipamentos de Saúde	Vigilância em Saúde	27
Equipamentos Segurança	Bombeiros	97
	Casa de Mediação	5
	Polícia Civil	190
	Polícia Militar	189
	Guarda Civil metropolitana	44
Equipamentos Serviços	CAT	24
	CCZ	4
	Consulado	49
	Correios	233
	Ecopontos / Centrais de Triagem	126
	Eletropaulo	11
	Poupatempo	5
	Receita Federal	10
	Sabesp	18
	Sede Prefeitura Regional	32
	Cemitérios	46
	Shopping Center	54
Sistema Viário	Logradouro	250.350
	Minianel Viário / Restrição Veículos	1
	Obras de Arte	7.303
Transportes	Estação de Metrô	88
	Linha de Metrô	6
	Pontos de Ônibus	20.386
	Redes Cicloviárias	253
	Terminais de Ônibus	32
	Estação de Trem	103
	Ferrovias	782
	Linhas Trem	7

(continuação)

Atributo Principal	Atributo Secundário	Nº de Registros
Verde e Recursos Naturais	Arborização Viárias	652.385
	Parques Municipais	165
	Reserva Mata Atlântica	4.496
	Unidades de Conservação	38
Total		2.108.139

Fonte: Adaptado de Geosampa, 2019.

Inicialmente os dados foram baixados em formato *shapefile* e *xls*, e organizados em pastas.

Em seguida, foi estruturado o banco de dados geográfico utilizando *Software* de geoprocessamento *ArqGis*, onde as camadas de dados geográficos foram importadas para o banco de dados. Após a separação das camadas através da interface gráfica criada utilizando *Software* de geoprocessamento, os dados foram ajustados na mesma projeção e no mesmo sistema de coordenadas geográficas. Adotou-se o Sistema de Coordenadas *Universal Transversa de Mercator (UTM)* e a Projeção *Sirgas 2000, Fuso 23S*.

Posteriormente, foi necessário tratar os dados do IPTU relativos aos imóveis, obtidos no formato de planilha, para conseguir estimar o valor do metro quadrado dos terrenos, e relacioná-los ao banco de dados geográfico denominado “logradouro”, o qual apresenta em formato *shapefile* a posição geográfica de cada endereço.

Para isso utilizou-se a ferramenta de geocodificação de endereços e pesquisa geográfica, função “*Geocode*” do *Software* de geoprocessamento. O localizador de endereços permitiu encontrar endereços e lugares para que se possa visualizar em um mapa e carregar como entrada para análise espacial.

A próxima etapa foi criar uma base de dados no *ArcCatalog*, denominada *Geodatabase*, e incluir todos os dados geográficos os quais serão utilizados na etapa de *Data Mining* (mineração de dados).

A etapa seguinte foi selecionar os equipamentos e serviços urbanos, assim como elementos do meio ambiente mais relevantes para a primeira análise, e posteriormente, utilizá-los na etapa de *data mining*, a qual será feita com a utilização do *Software Python* e do algoritmo *Randon Forest*.

3.2. Mineração dos dados

Para a mineração de dados foi utilizada a técnica de árvore de decisão aplicada através do algoritmo *Random Forest*, o qual está disponível na biblioteca do *Python*.

As etapas de processamento dos dados foram:

I) Seleção das *features* aleatórias e principal para calibração do modelo

As primeiras bases de dados selecionadas para mineração foram:

- Equipamentos de Cultura;
- Equipamentos de Lazer;
- Equipamentos de Segurança;
- Equipamentos de Educação;
- Equipamentos de Saúde;
- Proteção e Defesa Civil;
- Equipamentos de Transportes;
- Serviços;
- Assistência Social;
- Verde e Recursos Naturais;
- Densidade demográfica.

A etapa subsequente foi selecionar a *feature* a qual será irá compor a posição de nó raiz (principal), que, neste caso, são os valores do metro quadrado do terreno (*target*), obtidos na etapa de geocodificação de endereços.

As *features* hidrografia, ciclovias e linhas de metrô não foram utilizadas na mineração devido ao seus formatos serem segmentos (“*polyline*”), dificultando o cálculo da distância euclidiana delas em relação aos imóveis.

II) Leitura do valor do metro quadrado (*target*) dos terrenos

Os arquivos estavam organizados no formato “.gdb”. Esse formato possui uma estrutura complexa, com diversos arquivos que descrevem partes específicas da informação que neles contém, separando os tipos de informações em diferentes camadas. Para a leitura do arquivo foi usada a biblioteca em *Python fiona* que permite ler os

diferentes *layers* de informações separadamente. Dentre as informações contidas nestes arquivos, foram utilizadas as *features*:

- a) id : identificador único;
- b) Latitude;
- c) Longitude;
- d) Valor do metro quadrado do terreno.

Essas informações foram salvas em um arquivo do tipo “.csv” separadamente, para que pudessem ser facilmente acessadas e usadas na computação das demais variáveis.

III) Leitura e processamento dos dados pré-treino

Os arquivos estavam organizados no formato *shapefile*. Esse formato possui uma estrutura simples, e contém as informações geográficas das variáveis (equipamentos, serviços urbanos e aspectos ambientais). Para extraí-las foi usada a biblioteca em *Python geopandas*, e todas as variáveis, assim como sua respectiva classe e posições geográficas, foram extraídas e salvas em um arquivo “.csv” para posterior computação dos dados.

IV) Foi computada a distância euclidiana entre todos os imóveis e todas as variáveis, assim como a distância mínima, média, e a mediana das distâncias.

A Etapa subsequente foi selecionar todas as *features* que possuíam importância superior a 0,001. Para diminuir a quantidade de testes necessárias em uma busca em *grid*, cada parâmetro da floresta aleatório foi variado individualmente, selecionado o melhor, e em seguida repetido o processo para os demais parâmetros.

Por exemplo:

Modelo 1:

Número de estimadores = 100,

Número mínimo de arvores na ramificação = *default*

Acc = 91%

Modelo 2:

Número de estimadores = 250,

Número mínimo de árvores na ramificação = *default*

Acc = 92%

Modelo 3:

Número de estimadores = 500,

Número mínimo de árvores na ramificação = *default*

Acc = 91%

Neste caso o melhor parâmetro é o 250, então é mantido nos modelos subsequentes.

Modelo 4:

Número de estimadores = 250,

Número mínimo de árvores na ramificação = 2

Modelo 5:

Número de estimadores = 250,

Número mínimo de árvores na ramificação = 4

Modelo 6:

Número de estimadores = 250,

Número mínimo de árvores na ramificação = 8

...

Os parâmetros variados são: **Número de estimadores** e **Número mínimo de árvores na ramificação**.

A fim de verificar a acurácia dos resultados foram avaliados os resultados obtidos na etapa de treino, assim como no teste.

As métricas foram calculadas a partir da equação (1):

$$\text{Erro absoluto médio} = \text{Mean}(E) \quad (1)$$

Na qual:

$E = |P - VR|$, e:

P = Valor predito pelo modelo

VR = valor real

E a acurácia é estabelecida pela equação (2):

$$\text{Mean}(100 * E / VR) \quad (2)$$

4. RESULTADOS E DISCUSSÃO

A revisão bibliográfica contribuiu satisfatoriamente ao entendimento sobre as técnicas de inteligência artificial mais aplicadas atualmente nas pesquisas nacionais e internacionais, porém, não foi possível encontrar aplicação semelhante ao tema proposto nesta pesquisa para ser utilizada como base comparativa ou para aprimoramento dos processos de atribuição dos valores imobiliários.

Os dados utilizados referentes aos imóveis, equipamentos e serviços urbanos e comunitários nas 32 subprefeituras de São Paulo, foram tratados, minerados utilizando-se a técnica de árvores de decisão, e através da aplicação do algoritmo *Random Forest*, disponível na biblioteca do *Python*, tornou-se possível verificar quais variáveis possuem maior e menor influência na formação do valor do metro quadrado dos terrenos no município de São Paulo/SP.

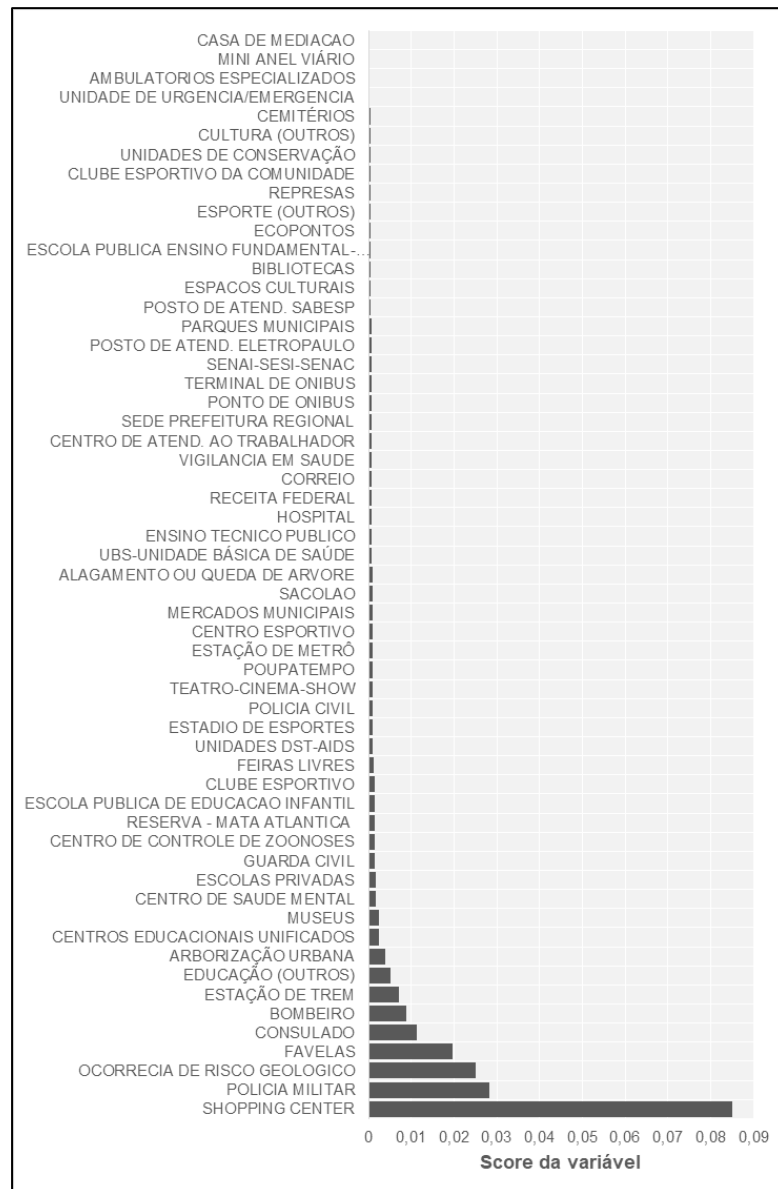
O *dataset* original utilizado para mineração é composto por 57 categorias de classes, sendo elas: equipamentos e serviços urbanos disponíveis à comunidade, assim como atributos ambientais.

De acordo com os testes realizados, o melhor modelo alcançado foi obtido com 100 (*default*) estimadores e 15 árvores mínimas. Após a seleção dos parâmetros, um modelo final foi criado com todas as *features*, e os parâmetros selecionados com o *subset* de *features*.

Os resultados obtidos no treino obtiveram a acurácia de 95,5% no treino e 92,9% na mineração (teste).

O Gráfico 1 apresenta os resultados obtidos, demonstrando o *Score* das principais variáveis. Quanto maior o *score*, maior é a relação com valor do metro quadrado do terreno. Desta forma, a análise possibilitou verificar quais as variáveis que mais influenciam no valor dos imóveis no município de São Paulo.

Gráfico 1: Score das variáveis em relação ao valor do metro quadrado do terreno



A *feature* que apresentou maior *score* foi Shopping Center. O impacto de shoppings na valorização de imóveis na maior capital do País pode estar ligado, além da preferência do consumidor, à falta de terreno e à alta procura no mercado.

Outras *features* que se mostraram importantes foram a Polícia Militar, as Favelas, o Bombeiro e a Guarda Civil, as quais estão relacionadas à segurança pública. Aparentemente, pode-se entender que a proximidade de áreas de baixa renda está relacionada às regiões com elevados índices criminais, o que pode influenciar negativamente o valor dos imóveis, assim como a proximidade à equipamentos de segurança pública pode influenciar positivamente.

Também foi possível identificar que variáveis ambientais possuem forte influência no valor dos imóveis em São Paulo. A ocorrência de riscos geológicos se apresenta-se em terceiro lugar no *ranking* e a arborização urbana em nono lugar. Muitos municípios tem instruído através de leis um tratamento tributário diferenciado (coeficiente menor de tributação) aos imóveis que são atingidos por Áreas com Potencial Risco Geológico, o que pode explicar o resultado.

Apesar do município de São Paulo apresentar um padrão de arborização urbana em grande parte das 32 regiões administrativas, seja planejado ou não, foi possível observar uma correlação direta entre a arborização urbana e o preço do metro quadrado dos imóveis. Por outro lado, apesar de algumas regiões intermediárias ou periféricas, por vezes apresentarem bons índices de concentração de equipamentos em comparação com o centro, estes ocorrem de forma isolada e em temas específicos, mais comumente ligados ao meio ambiente e educação. Essa condição favorável está possivelmente ligada à Lei de Diretrizes e Bases da Educação (LDB) e ao remanescente de Mata Atlântica do cinturão verde de São Paulo, protegido por Unidades de Conservação.

Outra variável que se apresentou relevante está relacionada à mobilidade urbana, como estação de trem em oitavo lugar no *ranking*. Embora o automóvel ainda seja o principal modo de transporte adotado, devido ao tráfego intenso no município de São Paulo, muitas pessoas já estão buscando a comodidade de morar em um bairro que ofereça opções mais sustentáveis de transporte. A busca por imóveis perto de estações de trem e metrô tem crescido bastante, refletindo no valor dos imóveis. Os resultados obtidos na mineração mostraram, ainda, que os serviços de educação também têm forte influência na valorização imobiliária.

5. CONCLUSÕES

No caso de São Paulo, como grande aglomerado urbano, com características peculiares de uma cidade que cresceu de forma desordenada, nota-se uma relação íntima entre a questão geográfica habitacional e o acesso à infraestrutura e ao direito à cidade.

O fato de um terreno ou benfeitoria com características idênticas possuírem valores imobiliários extremamente discrepantes dependendo da sua posição geográfica na cidade está relacionado ao que se chama de “renda imobiliária” ou “renda fundiária urbana”, e está diretamente correlacionado à disponibilidade de infraestrutura local.

Para São Paulo-SP, com as informações disponibilizadas no banco de dados do município, pôde-se obter uma série de vantagens no processo de análise através da mineração de dados, e, sobretudo, subsidiar o diagnóstico almejado.

Ao utilizar a mineração de dados e analisar a correlação entre os atributos mais expressivos, foi possível concluir que as técnicas de inteligência artificial são extremamente importantes para avaliação estatística de grandes bancos de dados, assim como, verificar a eficácia da mineração de dados e a sua aplicação ao planejamento urbano, implementação de políticas públicas, gestão democrática e para as tomadas de decisão.

O resultado da pesquisa aponta que a localização de terrenos próximos à Shoppings Centers apresenta maior valorização imobiliária. Conforme explicitado anteriormente, esse fator pode estar ligado à preferência do consumidor, à escassez de terrenos e à alta procura.

O acesso à segurança pública também se mostrou um fator determinante no valor dos terrenos em São Paulo. Foi possível verificar através da mineração de dados que a proximidade de equipamentos de segurança (delegacias de policias militares, corpo de bombeiro, consulados), assim como, ocorrência de riscos geológicos, tem forte influência no valor dos imóveis.

A mobilidade urbana foi outra amenidade que se mostrou importante para atribuir um valor aos imóveis. Residir em uma região que ofereça uma boa rede de transporte público proporciona maior qualidade de vida na metrópole. Ainda que a preferência dos brasileiros seja o automóvel, muitos já procuram por comodidade em residir próximo à equipamentos de educação, lazer e transporte.

A questão da mobilidade urbana tem forte influencia não somente na qualidade de vida da população, mas também no coletivo da cidade. O extensivo fomento à indústria automobilística, a facilitação do crédito para obtenção de automóveis e políticas públicas

direcionadas durante muitas décadas ao uso do carro em detrimento ao transporte coletivo, tornou insustentável a mobilidade na cidade como um todo.

A concentração de infraestrutura nas áreas centrais, somada ao problema da mobilidade e ao espraiamento populacional, motivado pela especulação imobiliária, faz com que a população residente nas regiões periféricas seja excluída do direito de usufruir da maioria dos equipamentos disponíveis na cidade.

Ainda que o centro antigo disponha dos maiores índices infraestruturais, e esteja entre as 5 regiões mais adensadas demograficamente do município, existe nela uma tendência de esvaziamento populacional que vem se confirmando nos últimos anos. Isso se dá tanto pela obsolescência dos antigos edifícios comerciais, quanto pela nova tendência de a população mais abastada financeiramente mudar-se para outras cidades próximas em busca de melhores condições e qualidade de vida.

Em uma direção contrária à lógica urbanística, muitos dos prédios que se tornaram comercialmente obsoletos no centro, ao invés de se readequarem e passarem por reformas que lhes possibilitem novos usos, permanecem fechados à espera de melhores índices de valorização.

Em relação aos aspectos ambientais, a pesquisa reflete a necessidade de se impedir e amenizar a expansão horizontal da mancha urbana na cidade, a fim de se preservar os últimos remanescentes verdes nela disponíveis. No entanto, se faz necessário ao poder público, direcionar seus esforços para mitigação dos riscos geológicos e de inundação os quais ocorrem em várias regiões do município, corroborando para riscos de mortalidade e da depreciação dos terrenos e imóveis urbanos.

Desta forma, o cumprimento da função social da propriedade, assim como, a aplicação do IPTU progressivo é apenas uma das soluções mitigadoras da problemática geográfica da habitação na cidade.

Se torna necessário, ainda, reestruturar o transporte coletivo para aliviar o trânsito, criar novas centralidades e condições para o estabelecimento de empresas em novas regiões, criar infraestrutura e equipamentos de qualidade nas regiões menos abastadas, incentivar e fortalecer a participação popular no processo de tomada de decisões, transformar o espaço público em áreas de convivência, e, preservar e ampliar a disponibilidade de arborização, áreas verdes e solo permeável em toda cidade.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABDALA, L.N.; SCHREINER, T.; COSTA, E.M.; SANTOS, N. **Como as cidades inteligentes contribuem para o desenvolvimento de cidades sustentáveis? Uma revisão sistemática de literatura.** Int. J. Knowl. Eng. Manag. v. 3, n.5, p. 98-120. 2014.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES GONÇALVES, J. L.; SPAROVEK, G. **“Köppen's climate classification map for Brazil”.** 2013.

ARCURI, M. E. P. **Planejamento Ambiental através de análise de informações em bases cartográficas: o caso de estudo da microbacia Santa Eliza, Rio Claro – SP.** Dissertação (Mestrado), UNESP – Universidade Estadual Paulista, IGCE, Rio Claro, 1997.

ARONOFF, S. **Geographic information systems: a management perspective.** Ottawa, WDL Publications, 1989. 295p.

BARROS, A. J. S.; LEHFELD, N. A. S. **Fundamentos de Metodologia: Um Guia para a Iniciação Científica.** 2 Ed. São Paulo: Makron Books, 2000.

BERRY, M. J. A.; LINOFF, G. **Data Mining Techniques: For Marketing, Sales, and Customer Support.** New York: Wiley Computer Publishing, 1997.

BIGOLIN, N. M.; BOGORNY, V.; ALVARES, L. O. C. Uma Linguagem de Consulta para Mineração de Dados em Banco de Dados Geográficos Orientado a Objetos. Em: XXIX Conferência Latinoamericana de Informática, 2003, La Paz. CLEI 2003, v. 1, 2003.

BOGORNY, V. **Algoritmos e ferramentas de descoberta de conhecimento em bancos de dados geográficos.** Dissertação - PPGC, UFRGS, Porto Alegre, 2003.

BOGORNY, V; SHEKHAR, S. Spatial and spatio-temporal data mining. **23rd Brazilian Symposium on Databases.** 315-316 pp. Campinas, São Paulo, Brazil — October 13 - 17, 2008.

BOGORNY, V. **Algoritmos e ferramentas de descoberta de conhecimento em bancos de dados geográficos.** Dissertação - PPGC, UFRGS, Porto Alegre, 2003.

BOTELHO, A. **A renda fundiária urbana: uma categoria de análise ainda válida.** Revista Geographia, v. 10, n. 19, p. 25-45, 2008.

BRASIL (2015). **Decreto nº 56.834, de 24 de fevereiro de 2016: institui o Plano Municipal de Mobilidade Urbana de São Paulo – PlanMob/SP 2015.** São Paulo: Prefeitura do Município de São Paulo. Disponível em: <https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/planmobsp_v072__1455546429.pdf>. Acesso em 04/07/2019.

BRASIL, **Decreto Federal no 4.297 de 10 de julho de 2002: Regulamenta o art. 9o, inciso II, da Lei no 6.938, de 31 de agosto de 1981, estabelecendo critérios para o Zoneamento Ecológico-Econômico do Brasil - ZEE, e dá outras providências.** Disponível em

<http://www.planalto.gov.br/ccivil_03/decreto/2002/d4297.htm>, acessado em 15 de abril de 2017;

BURROUGH, P. A. **Principles of geographical information systems**. Oxford University Press, Oxford, 1986.193p.

CAMILO, C.O.; SILVA, J.C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Instituto de Informática Universidade Federal de Goiás. 2009.

CLIMATEMPO. **Climatologia de São Paulo**. Disponível em: <<https://www.climatempo.com.br/climatologia/558/saopaulo-sp>>. Consultado em 30 de janeiro de 2019.

DINIZ, C. A.; LOUZADA NETO, F. **Data Mining: uma introdução**. São Paulo: ABE, 2000.

FAYYAD U.M.; PIATETSKY-SHAPIRO G.; SMYTH P. "**From data mining to knowledge discovery: An overview**", in: Fayyad U.M., Piatetsky-Shapiro G., Smyth P. (eds.), *Advances in Knowledge Discovery and Data Mining*, Menlo Park, USA, American Association for Artificial Intelligence. 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, v. 39, n.11, p. 27-34. 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996.

FREITAS, C.G.L. **Cartografia geotécnica de planejamento e gestão territorial: proposta teórica e metodológica**. Universidade de São Paulo. Tese de Doutorado, Faculdade de Filosofia, Letras e Ciências Humanas, Departamento de Geografia. São Paulo, Universidade de São Paulo, 230 p., 2000.

FREITAS, E. A. S. F.; PEREIRA, B. D.; MAIA, J. C. de S. Valoração contingente de ativos ambientais na suinocultura: um estudo de caso no município de Diamantino-MT. 2004. 167f. Dissertação (Mestrado) – Universidade Federal do Estado de Mato Grosso, Cuiabá, MT. 2004.

GERHARDT, T. E.; SILVEIRA, D. T. (organizadores). **Métodos de Pesquisa**. Coordenado pela Universidade Aberta do Brasil – UAB/UFRGS e SEAD/UFRGS. Porto Alegre: Editora da UFRGS, 2009.

GEOSAMPA – PREFEITURA DO MUNICÍPIO DE SÃO PAULO. Disponível em: <http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx>. Acesso em: 05/06/2018.

HAN, J., PEI, J.; KAMBER, M. **Data mining: concepts and techniques**.(3a ed.) . Oxford, UK: Elsevier. 2011.

HAN, J.; KAMBER, M.; TUNG, A. Spatial clustering methods in data mining: a Survey. School of Computing Science, Simon Fraser University, Burnaby, BC Canada, v5a156, 2001.

HIDALGO, B. D. **As divisões territoriais do município de São Paulo: uma proposta de classificação por meio da análise dos Distritos**. São Paulo, 2013. 115 pp.

HOFFMANN, R.; VIEIRA, S. **Análise de regressão: Uma introdução à econometria**. Hucitec, São Paulo, 2 ed., 1983. 379p.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Atlas do censo demográfico 2010**. Brasília, 2013.

IBGE, **Produto Interno Bruto dos Municípios 2005**. Disponível em: <<https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>>. Acesso em 20/12/2018.

INVEST SÃO PAULO. Agência Paulista de Promoção de Investimentos e Competitividade. 2011. Disponível em: <<https://www.investe.sp.gov.br/por-que-sao-paulo/economia-diversificada/>>. Acesso em: 10/01/2019.

LÉFÈBVRE, H. **A Revolução Urbana**. Belo Horizonte: EDUFMG, 2004.

MARSON, M. D. A industrialização brasileira antes de 1930: uma contribuição sobre a evolução da indústria de máquinas e equipamentos no estado de São Paulo, 1900-1920. SciELO, São Paulo, out/dez. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-41612015000400753>. Acesso em: 12/07/2020.

MELANDA E.; HUNTER A.; BARRY M. Identification of locational influence on real property values using data mining methods. **European Journal of Geography**. 2016. Link de acesso: <<https://journals.openedition.org/cybergeog/27493?lang=es>>.

MELANDA, E. A. **"Pós-processamento de regras de associação"**. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação, USP, São Carlos, SP, Brasil. 2004.

MELANDA, E. A.; REZENDE, S. O. Combining quality measures to identify interesting association rules. **Anais**. Heidelberg: Springer, 2004.

ORFORD, S. **Valuing locational externalities: a GIS and multilevel modelling approach**. Environment and Planning B: Planning and Design, v.29, n.1, p. 105-127. 2002.

PAVLOV, A. D. "Space-varying regression coefficients: A semi-parametric approach to real estate markets". **Real Estate Economics**. v.28, n.2, p. 249-283. 2000.

PELLI NETO, A. **Redes neurais artificiais aplicadas às avaliações em massa – estudo de caso para a cidade de Belo Horizonte**. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica, UFMG, Belo Horizonte, 2006.

ROSS, J. L. S. Análise empírica da fragilidade dos ambientes naturais e antropizados. **Revista do Departamento de Geografia – FFLCH-USO**, São Paulo, n. 8, p 63-74, 1994.

REZENDE, S. O. **Mineração de dados**. Disponível em:
<<http://www.lbd.dcc.ufmg.br/colecoes/enia/2005/0102.pdf>>. Acesso em: 11/08/2018.

REZENDE, S. O. **Sistemas Inteligentes: Fundamentação e Aplicações**. Editora Monole Ltda, Bueri, SP: Manole, 2003.

SÃO PAULO (Cidade). **Infocidade: Dados estatísticos do desenvolvimento urbano**. Secretaria Municipal de Desenvolvimento Urbano. Disponível em:
<https://www.prefeitura.sp.gov.br/cidade/secretarias/urbanismo/dados_estatisticos/info_cidade/index.php/>. Acesso em: 17/03/2019.

SÃO PAULO (Cidade). **Gestão urbana: Rede de Espaços Públicos**. Disponível:
<<http://gestaourbana.prefeitura.sp.gov.br/rede-de-espacos-publicos/>>. Acesso: 18/06/2019.

SÃO PAULO (Cidade). **Infocidade: Dados demográficos por Prefeituras Regionais de São Paulo** -. Secretaria Municipal de Prefeituras Subprefeituras. Disponível em:
<https://www.prefeitura.sp.gov.br/cidade/secretarias/subprefeituras/subprefeituras/dados_demograficos/index.php?p=12758>. Acesso em 14/07/2019.

SÃO PAULO (Cidade). Secretaria Municipal do Verde e do Meio Ambiente (SMVMA). **A Questão Ambiental Urbana**. São Paulo, 1993.

SÃO PAULO. **Mapa Pedológico do Estado de São Paulo**. Instituto Agrônomo de Campinas – IAC. Campinas, 2002.

SÃO PAULO. **Lei Municipal nº 13.885 de 25 de agosto de 2004**. Estabelece normas complementares ao Plano Diretor Estratégico, institui os Planos Regionais Estratégicos das Subprefeituras, dispõe sobre o Parcelamento, disciplina e ordena o Uso e Ocupação do Solo do Município de São Paulo.

SÃO PAULO. **PlanMob/SP 2015 - Plano de Mobilidade de São Paulo**. Prefeitura Municipal de São Paulo. Secretaria Municipal de Transportes. São Paulo. 2015

SÃO PAULO. **Unidades Básicas de Compartimentação do Meio Físico (UBC) do Estado de São Paulo**. INSTITUTO GEOLÓGICO/COORDENADORIA DE PLANEJAMENTO AMBIENTAL, SECRETARIA DO MEIO AMBIENTE DO ESTADO DE SÃO PAULO, 2014.

SECRETARIA DO MEIO AMBIENTE e INSTITUTO FLORESTAL. **Inventário florestal da vegetação natural do estado de São Paulo**. Secretaria do Meio Ambiente / Instituto Imprensa Oficial, São Paulo, 2010.

VERAS, M. P. B. “Tempo e Espaço na Metrópole: breves reflexões sobre assincronias urbanas”. In: São Paulo em Perspectiva, São Paulo, Fundação Seade, v.15, n.1, jan./mar., 2001.

WEISS, S. M.; INDURKHYA, N. **Predictive Data Mining: A Practical Guide**. Morgan Kaufmann Publishers, Inc., San Francisco, CA. 1998.

ZADEH, L. A. **Fuzzy sets.** *Information and Control.* 8:338-353, 1965.