

Universidade de Brasília - UnB
Faculdade do Gama - FGA
Engenharia Eletrônica

Previsão de preço justo para imóveis usando aprendizado de máquina

Autor: Eduardo Felipe Leão Ribeiro de Menezes
Orientador: Prof. Dr. Ricardo Chaim

Brasília, DF
2023



Eduardo Felipe Leão Ribeiro de Menezes

Previsão de preço justo para imóveis usando aprendizado de máquina

Monografia submetida ao curso de graduação em Engenharia Eletrônica da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia Eletrônica.

Universidade de Brasília - UnB

Faculdade do Gama - FGA

Orientador: Prof. Dr. Ricardo Chaim

Brasília, DF

2023

Eduardo Felipe Leão Ribeiro de Menezes

Previsão de preço justo para imóveis usando aprendizado de máquina/ Eduardo
Felipe Leão Ribeiro de Menezes. – Brasília, DF, 2023-

Orientador: Prof. Dr. Ricardo Chaim

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade do Gama - FGA , 2023.

1. Previsão de preço de imóveis. 2. Aprendizado de máquina. I. Prof. Dr.
Ricardo Chaim. II. Universidade de Brasília. III. Faculdade do Gama. IV. Previsão
de preço justo para imóveis usando aprendizado de máquina

Eduardo Felipe Leão Ribeiro de Menezes

Previsão de preço justo para imóveis usando aprendizado de máquina

Monografia submetida ao curso de graduação em Engenharia Eletrônica da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia Eletrônica.

Prof. Dr. Ricardo Chaim
Orientador

Prof. Dr. Fernando William
(FGA/UnB)
Convidado 1

Prof. Dr. Giovanni
Almeida(FGA/UnB)
Convidado 2

Brasília, DF
2023

Aos meus pais, Marcilon e Patrícia, pelos gestos de compressão, palavras de incentivo, paciência, generosidade, companheirismo e amizade em todos os momentos. Ao meu irmão Rafael, pelo carinho e amizade. À minha namorada, Carolina, que sempre está ao meu lado, dando apoio e incentivo

*“A disciplina é o freio que nos guia na direção certa.”
(Sêneca)*

Resumo

O mercado de imóveis é um pilar essencial da economia, tendo impacto direto na vida dos cidadãos brasileiros, uma vez que grande parte da riqueza existente está concentrada nesse tipo de ativo. Diante desse panorama, o presente estudo analisa os atuais modelos de precificação disponíveis no setor e busca, através da aplicação do aprendizado de máquina, o desenvolvimento de métodos alternativos para serem adotados tanto de forma comparativa com os modelos já existentes quanto como técnica praxe de mercado. A fim de cumprir esse objetivo, realiza-se a construção de uma base de dados com as principais variáveis de preço utilizando-se de API, em seguida estrutura-se essas informações para serem inseridas no novo modelo de aprendizado. Por fim, compara-se os resultados encontrados com os valores praticados atualmente no setor, estudando regiões subvalorizadas e supervalorizadas. Destaca-se que obteve os melhores resultados foi a Regressão Linear.

Palavras-chaves: mercado imobiliário, aprendizado de máquina, previsão de preço de imóvel.

Abstract

The Real State Market is an essential pillar of the economy, having a direct impact on the lives of Brazilian citizens, since a large part of the existing wealth is concentrated in this type of asset. Given this scenario, the present study analyzes the current pricing models available in the sector and seeks, through the application of Machine Learning, the development of alternative methods to be adopted both in a comparative way with the existing models and as a usual market technique. In order to fulfill this objective, a database is built with the main price variables using API, then there is the structuring of this information to be inserted in the new evaluation model. Finally, compare the results found with the cultural values currently in the sector, studying undervalued and overvalued regions. It is noteworthy that Linear Regression obtained the best results.

Key-words: real state , machine learning, real state price predictions.

Lista de ilustrações

Figura 1 – Captura de tela do site com ferramenta	13
Figura 2 – Preço médio do m ² de imóveis urbanos por Região	16
Figura 3 – Representatividade de imóveis urbanos por Região	17
Figura 4 – Tamanho do imóvel urbano por região	17
Figura 5 – Preço do imóvel urbano por Região	18
Figura 6 – Representatividade do mercado imobiliário por Região	19
Figura 7 – Aprendendo um modelo a partir de uma base histórica	21
Figura 8 – Diagrama de Floresta Aleatória	23
Figura 9 – O modelo do neurônio	24
Figura 10 – Uma Rede Neural Artificial totalmente conectada	25
Figura 11 – Diagrama de etapas da metodologia para a precificação de imóveis	27
Figura 12 – Diagrama de execução da base de dados	28
Figura 13 – Tabela representando a base de dados	28
Figura 14 – Diagrama representando a criação do preço justo	29
Figura 15 – Diagrama representando ações	29
Figura 16 – Correlação de Pearson em Brasília	35
Figura 17 – Correlação de Spearman em Brasília	36
Figura 18 – Reta da Regressão Linear: Aluguel vs Valor justo aluguel (RL) em Brasília	37
Figura 19 – Distribuição da variação dos valores em Brasília	38
Figura 20 – Bairros com maiores variações em Brasília	39
Figura 21 – Mapa do bairro de Pinheiros	40
Figura 22 – Correlação de Pearson em São Paulo	41
Figura 23 – Correlação de Spearman em São Paulo	42
Figura 24 – Reta da Regressão Linear: Aluguel vs Valor Justo Aluguel	43
Figura 25 – Distribuição da variação dos valores em São Paulo	44
Figura 26 – Variação de preço do aluguel justo vs preço de mercado por bairro em São Paulo	45
Figura 27 – Distribuição Yield projetado vs valor Venda em São Paulo	46
Figura 28 – Imóvel listado com maior Yield	47
Figura 29 – Reta da Regressão Linear: Venda vs Valor Justo Venda em São Paulo	48
Figura 30 – Representatividade da variação dos valores de venda em São Paulo	49
Figura 31 – Bairros com maiores variações em São Paulo	50
Figura 32 – Distribuição Yield vs valor Venda em São Paulo	51
Figura 33 – Imóvel listado com maior Yield Projetado	53

Lista de tabelas

Tabela 1 – Tamanho do mercado de imóveis urbanos por Região	18
Tabela 2 – Número de imóveis em Brasília - DF e São Paulo - SP	33
Tabela 3 – Comparação dos modelos de aprendizado de máquina	33
Tabela 4 – Maiores Variações de Valor Justo em Brasília	37
Tabela 5 – Maiores Variações de Valor Justo de aluguel em São Paulo	43
Tabela 6 – Maiores Yields em São Paulo	47
Tabela 7 – Maiores Variações de Valor Justo de venda em São Paulo	49
Tabela 8 – Maiores Yields Projetados São Paulo	52

Lista de abreviaturas e siglas

ANN	<i>Artificial Neural Network</i>
API	Interface de Programação de Aplicação
FIPE	Fundação Instituto de Pesquisas Econômicas
IBGE	Instituto Brasileiro de Geografia e Estatística
ML	<i>Machine Learning</i>
MSE	<i>Mean Squared Error</i>
PIB	Produto Interno Bruto
RF	Random Forest
RL	Regressão Linear
RMSLE	<i>Root Mean Squared Logarithmic Error</i>

Sumário

1	INTRODUÇÃO	13
1.1	Objetivo Geral	14
1.2	Objetivos Específicos	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Mercado Imobiliário	16
2.2	Aprendizado de Máquina	20
2.2.1	Regressão Linear	21
2.2.2	Floresta Aleatória	22
2.2.3	Redes Neurais	23
2.3	<i>Machine Learning</i> para Previsão de Preços de Imóveis	25
3	METODOLOGIA	27
3.1	Procedimento de Coleta de Dados	27
3.2	Tratamento de Dados	27
3.3	Linguagem Utilizada	30
3.4	Máquina Virtual	30
3.5	Containers	30
3.6	Docker	31
3.7	Métricas de Avaliação	32
3.8	Implementação e Testes	32
4	RESULTADOS	33
4.1	Brasília	34
4.2	São Paulo	40
4.2.1	Imóveis listados para aluguel em São Paulo	42
4.2.2	Imóveis listados para venda	48
5	CONCLUSÃO	54
	REFERÊNCIAS	55
	APÊNDICES	58
	APÊNDICE A – CÓDIGO UTILIZADO PARA OBTER A BASE DE DADOS SÃO PAULO	59

B	-	CÓDIGO UTILIZADO PARA OBTENÇÃO DA BASE DE DADOS DE BRASÍLIA	63
C	-	CÓDIGO UTILIZADO PARA APRENDIZADO DE MÁQUINA . . .	67

1 Introdução

O mercado de imóveis é uma área complexa e em constante mudança, impactado por vários fatores, como a localização da propriedade, o tipo de imóvel, taxas de juros, entre outros. Determinar o preço justo, ou seja, o preço que, de acordo com comparativos de mercado, melhor se adéqua para um imóvel, é desafiador tanto para o vendedor quanto para o comprador, uma vez que o mercado carece de ferramentais para essa decisão.

Neste estudo, propõe-se utilizar o aprendizado de máquina para avaliar o valor de um imóvel, utilizando-se variáveis e informações diversas sobre a propriedade para definir o preço justo, ou seja, de acordo com o que existe listado em sites de venda e aluguel de imóveis, o valor que uma propriedade deveria ser listada, a partir do aprendizado de máquina e com base nas características que se assemelham ao que já existe nos mercados.

Além disso, será feita uma análise dos resultados obtidos, apresentando informações sobre a localização estudada e comparando os preços dos imóveis, a fim de concluir se o preço está acima ou abaixo do que seria considerado justo.

A Figura 1 apresenta um exemplo do que é encontrado no mercado atualmente, onde o valor do aluguel é sugerido de acordo com as informações que o usuário fornece para o modelo, porém carece de informações com juízo de valor e comparativo para que se possa averiguar localizações e listagens que encontram-se com preço abaixo ou acima do valor justo, de acordo com modelo matemático.

Figura 1 – Captura de tela do site com ferramenta



Fonte: Quinto Andar (5ANDAR, 2023)

Conforme indicado por (PINDYCK; RUBINFELD, 2014), a compreensão do funcionamento dos mercados é crucial para muitos estudos econômicos, abordando temas como a intervenção do governo, a concorrência entre as empresas do mercado e a variação dos preços. Em resumo, o mercado pode ser entendido como um espaço (não necessariamente geográfico) onde ofertantes e demandantes de bens e serviços se encontram, segundo (PASSOS; NOGAMI, 2005).

A utilização de ferramentas de avaliação imobiliária tem evoluído ao longo dos anos (DANTAS, 2012). De acordo com a NBR 14653:2004 da ABNT (IBGE, 2017), os seguintes métodos são utilizados para avaliar imóveis:

- Método de Capitalização de Renda: utiliza a capitalização presente da renda líquida prevista.
- Método Comparativo Direto de Dados do Mercado: realiza uma comparação com base nos valores e informações disponíveis no mercado, para avaliar imóveis semelhantes em diferentes categorias.
- Método Evolutivo: o valor do imóvel é obtido pela soma das parcelas, incluindo o Fator de Comercialização.

Este estudo se concentrará no segundo método, uma vez que é o único método capaz de ter informações de uma ampla gama de imóveis para que se consiga estruturar um modelo replicável em diferentes localidades, visto que os dados utilizados podem ser obtidos através de sites de anúncios de imóveis.

1.1 Objetivo Geral

O presente trabalho tem como objetivo principal desenvolver um novo modelo de precificação de ativos no mercado imobiliário utilizando técnicas de aprendizado de máquina.

1.2 Objetivos Específicos

Para alcançar o objetivo geral desse trabalho será necessário:

- Coletar informações relevantes de preços históricos, informações sobre as características dos imóveis, entre outros, por meio de API.
- Estruturar e processar essas informações para que possam ser utilizadas na construção do novo modelo de precificação.

- Desenvolver um modelo de precificação utilizando técnicas de aprendizado de máquina, que possa ser aplicado a diferentes ativos do mercado de valores imobiliários.
- Validar o modelo em dados de teste e comparar seus resultados com os modelos já existentes.
- Identificar regiões subvalorizadas e valorizadas no mercado de valores mobiliários com base nos resultados do novo modelo de precificação.
- Avaliar a aplicabilidade do novo modelo como técnica praxe de mercado.

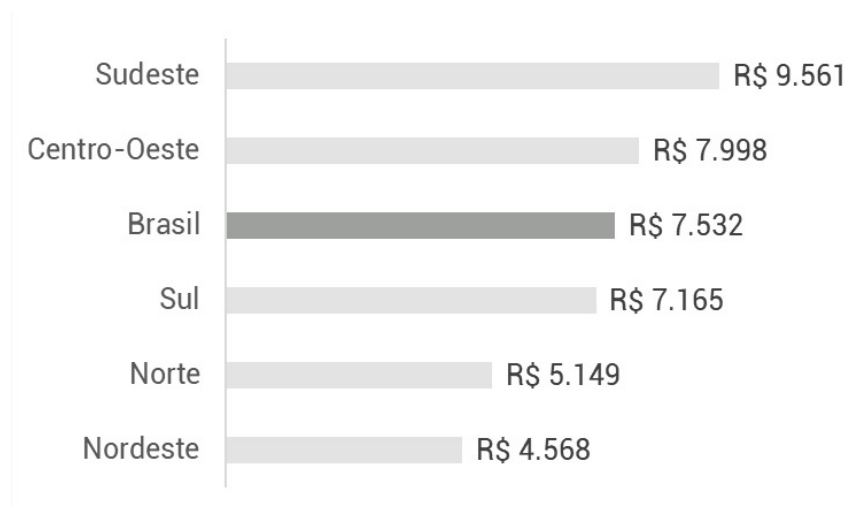
2 Revisão Bibliográfica

2.1 Mercado Imobiliário

No Brasil, existem de 72,4 milhões de domicílios particulares, dos quais 85,6% são casas (IBGE, 2019). Para a aferição dos dados mercadológicos, a Fundação Instituto de Pesquisas Econômicas (FIPE), em parceria com o portal de anúncios ZAP elabora relatórios mensais para avaliar o preço dos imóveis. Em janeiro de 2023, o valor médio do metro quadrado no Brasil foi de R\$ 8.399 (FIPE, 2023). De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE) referentes a 2019, a área média das residências urbanas no Brasil é de 80 metros quadrados (PNADC, 2020).

A Região Sudeste possui o maior valor de m² para imóveis urbanos, R\$ 9.561, abaixo dele está a Região Centro-Oeste com 16,3% a menos, resultando em um valor de R\$7.998. Na Figura 2, é mostrado o preço médio do m² urbano por Região:

Figura 2 – Preço médio do m² de imóveis urbanos por Região

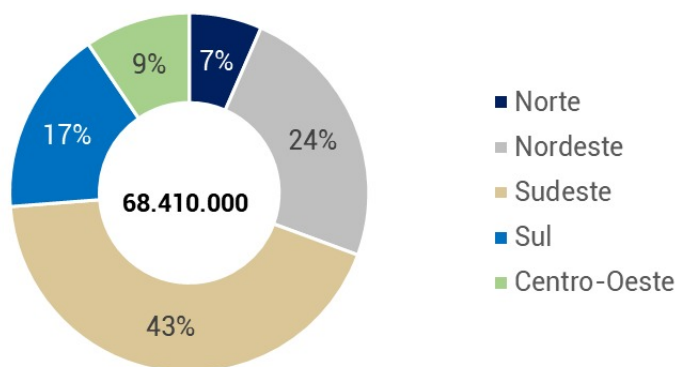


Fonte: (FIPE, 2023)

Quando se analisa a representatividade das Regiões na quantidade de imóveis, evidencia-se na Figura 3 a predominância da Região Sudeste, com 43% do total de imóveis urbanos no país. Em contrapartida, a Região Norte possui a menor representatividade, com 7%.

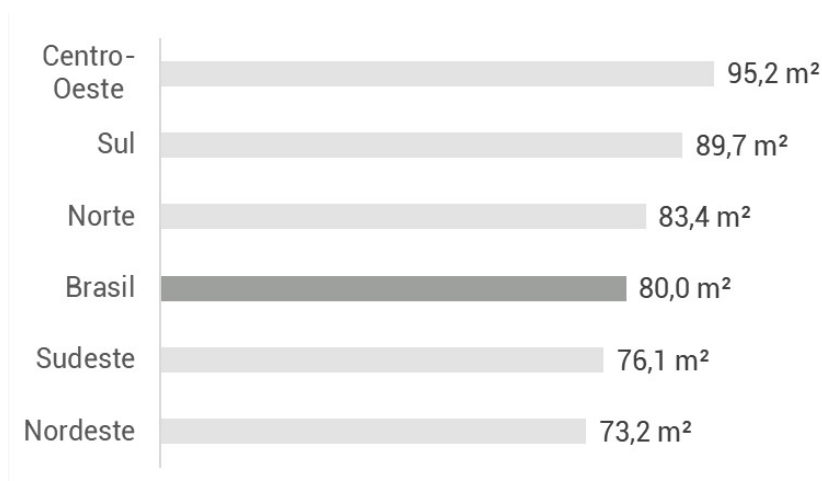
Da mesma forma, apresenta-se na Figura 4 o tamanho médio dos imóveis urbanos por Região, onde a Região Centro-Oeste lidera com 95,2 m², a média nacional resultante é de 80 m², conforme mostrado abaixo:

Figura 3 – Representatividade de imóveis urbanos por Região



Fonte: (PNADC, 2020)

Figura 4 – Tamanho do imóvel urbano por região



Fonte: (PNADC, 2020)

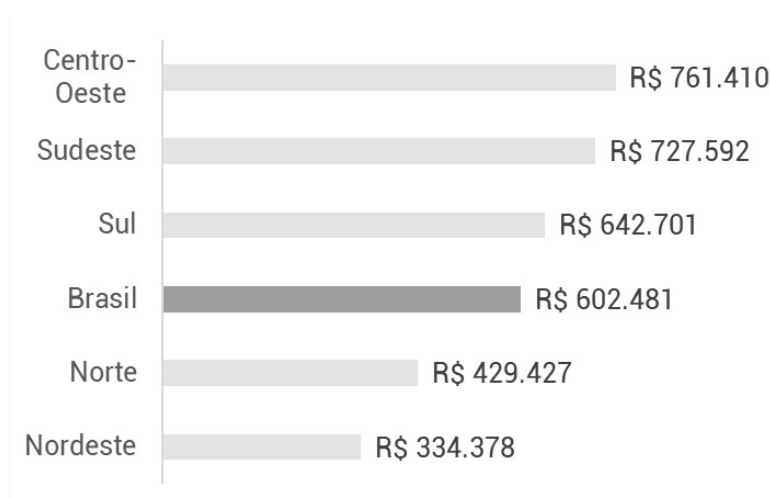
A partir dos dados apresentados, calcula-se o preço médio dos imóveis urbanos por Região, a partir da seguinte fórmula:

$$\text{Preço do imóvel} = \text{Preço do m}^2 \times \text{Tamanho do imóvel} \quad (2.1)$$

A partir da equação 2.1, calcula-se o valor do preço médio de imóveis urbanos por Região e Brasil.

Analisa-se as regiões por valor médio do preço de imóveis urbanos. Destaca-se na Figura 5 a liderança da Região Centro-Oeste frente as demais com valor médio de R\$ 761.410 enquanto a média nacional é de R\$ 602.481, representando um valor 20,9% menor.

Figura 5 – Preço do imóvel urbano por Região



Fonte: De autoria própria

De acordo com o IBGE, o Produto Interno Bruto (PIB) do Brasil em 2021 foi de R\$ 8,7 trilhões (IBGE, 2022). O valor total dos imóveis urbanos existentes no Brasil é de R\$ 41,2 trilhões no ano de 2023. Isso representa um valor 4,62x maior do que o PIB nacional frente o valor divulgado em 2022.

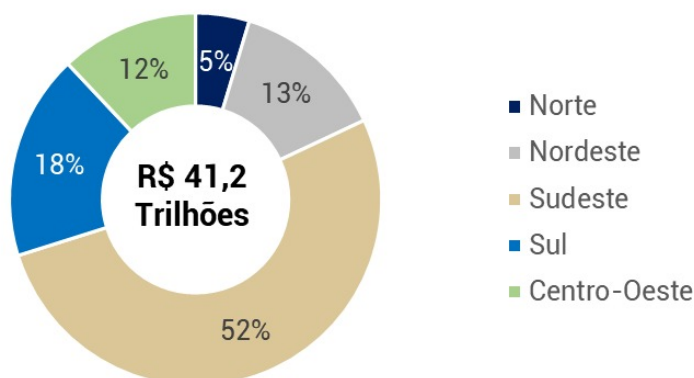
Tabela 1 – Tamanho do mercado de imóveis urbanos por Região

Região	Tamanho do mercado de imóveis urbanos
Norte	R\$ 1.925.119.447.800
Nordeste	R\$ 5.504.189.673.600
Sudeste	R\$ 21.483.611.936.700
Sul	R\$ 7.342.210.512.000
Centro-Oeste	R\$ 4.960.583.544
Brasil	R\$ 41.215.715.114.100

Fonte: De autoria própria

Assim como a quantidade de imóveis, na Figura 6 a Região Sudeste lidera o mercado imobiliário urbano com 52% do total de ativos estimados. Em contrapartida, a Região Norte possui a menor representatividade, com 5% do total.

Figura 6 – Representatividade do mercado imobiliário por Região



Fonte: De autoria própria

Como abordado, o mercado de imóveis abrange uma vasta quantidade de propriedades, interagindo com amplas áreas da sociedade. O processo de avaliação de um imóvel busca a determinação do preço mais provável de acordo com o que o mercado estaria disposto a pagar.

Fatores qualitativos que influenciam o preço do imóvel são (GOMES; MACIEL; KUWAHARA, 2012):

- Características:
- Acessibilidade:
- Especificações:

No mesmo estudo, ainda é destacado que a distância ao metrô e área útil não afetam o valor do apartamento significativamente, uma vez que áreas mais próximas implicam em maior poluição sonora e atmosférica, maior fluxo de veículos e maior quantidade de pedestres.

Por outro lado, mostraram que os seguintes atributos possuem uma correlação positiva com o aumento ou diminuição do preço do apartamento:

- tamanho da área total construída,
- menores custos de produção,
- quantidade de vagas de estacionamento,
- número de banheiros,

- número de quartos,
- regiões com maior qualidade de vida (localização).

Dessa forma, como apresentado no estudo, devido à sua correlação positiva, o aumento do número de quartos implica no acréscimo do preço da propriedade, assim como sua diminuição implica na desvalorização.

No viés determinístico do que é importante para o preço do imóvel, (BOURASSA; CANTONI; MARTIN, 2005) conclui que o preço do imóvel é provável estar relacionado com seus adjacentes, portanto se um modelo de previsão de preços não abrange a localização seus valores carecerão de informações determinísticas.

Diferentemente dos outros mercados, o setor imobiliário possui uma alta vida útil, singularidades e localização (NETO, 2003). Ainda é citado que mesmo existindo similaridades entre as propriedades, cada imóvel é único, por exemplo um imóvel pode possuir maior preço no mesmo local, devido ao andar que se localiza. Em outros casos, propriedades com as mesmas características estruturais, podem possuir variação de preços devido à sua localização (GOMES; MACIEL; KUWAHARA, 2012).

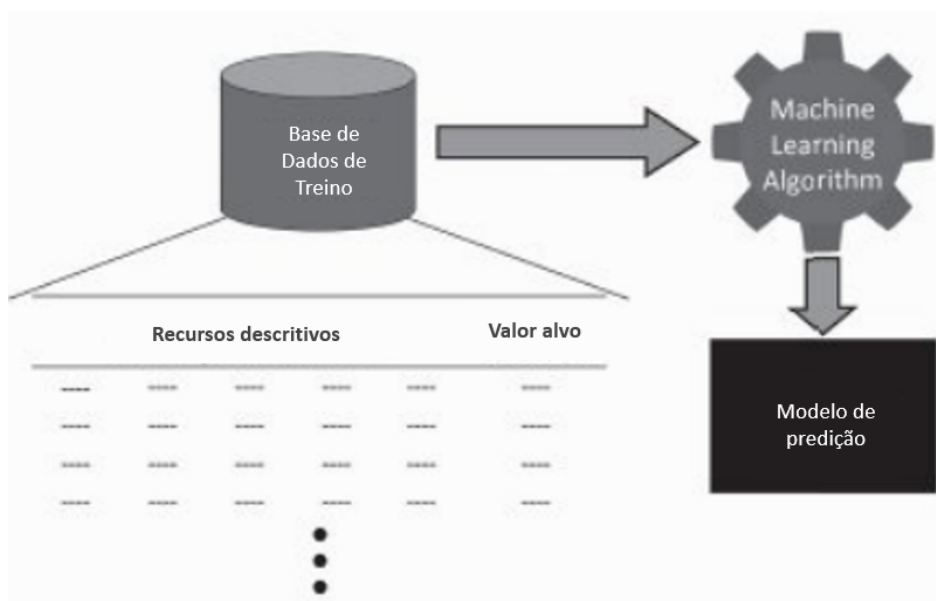
2.2 Aprendizado de Máquina

O aprendizado de máquina é um campo em rápido crescimento na área de inteligência artificial, que visa desenvolver algoritmos que permitem que sistemas de computador aprendam com dados e façam previsões ou executem ações com base na entrada de informações. Existem várias abordagens para aprendizado de máquina, incluindo aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado por reforço (MURPHY, 2012). A Figura 7 apresenta o fluxo de informações para que ocorra o aprendizado de máquinas, onde um conjunto de dados é separado em teste e o restante utiliza-se para a própria utilização do modelo.

O aprendizado supervisionado é a abordagem mais amplamente usada em aprendizado de máquina, no qual um conjunto de dados é usado para treinar um modelo para fazer previsões com base em exemplos rotulados (KELLEHER, 2015). Um dos exemplos mais conhecidos de aprendizagem supervisionada é a regressão linear, que é usada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes (HASTIE; FRIEDMAN, 2009).

Por outro lado, o aprendizado não supervisionado lida com dados não rotulados e visa encontrar padrões ou relacionamentos dentro dos dados. Uma das técnicas mais comuns nessa abordagem é o agrupamento, que baseia-se no agrupamento de pontos de dados semelhantes referentes em seus atributos (YU; WUNSCH, 2005).

Figura 7 – Aprendendo um modelo a partir de uma base histórica



Fonte: (KELLEHER, 2015)

O aprendizado semissupervisionado combina as ideias de aprendizado supervisionado e não supervisionado, usando dados rotulados e não rotulados para treinar um modelo (CHAPELLE, 2006). Essa abordagem pode ser útil em situações em que é difícil ou caro obter dados rotulados.

O aprendizado por reforço é um tipo de aprendizado de máquina no qual um agente interage com um ambiente e aprende como realizar determinadas ações para maximizar um sinal de recompensa (BARTO, 2018). Essa abordagem é comumente usada em robótica e sistemas de controle, bem como em jogos e outras simulações.

Mesmo com grande diversidade de modelos, faz-se por necessário mensurar e saber qual é o melhor modelo a ser utilizado para a situação em que se encontra (KUHN; JOHNSON, 2013).

A seguir, apresenta-se uma descrição e caracterização dos modelos mais usados no contexto do mercado imobiliário.

2.2.1 Regressão Linear

A regressão linear é a forma mais básica de regressão, na qual a relação entre as variáveis dependentes e independentes é modelado como uma equação linear (HASTIE; FRIEDMAN, 2009). Neste modelo, os parâmetros da equação são estimados usando os dados de treinamento, e o modelo resultante pode ser usado para fazer previsões para novos pontos de dados.

A regressão linear é um método usado para modelar a relação entre uma variável

dependente e uma ou mais variáveis independentes. É um algoritmo de aprendizado supervisionado que usa uma função linear para prever o valor da variável dependente com base nos valores das variáveis independentes.

A forma básica do modelo de regressão linear é mostrada na Figura 2.2:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.2)$$

onde y é a variável dependente, x_1, x_2, \dots, x_n são as variáveis independentes, e $b_0, b_1, b_2, \dots, b_n$ são os coeficientes que representam a inclinação da linha para cada variável independente.

O objetivo da regressão linear é encontrar os valores dos coeficientes que minimizam a diferença entre os valores previstos e os valores reais da variável dependente. Isso geralmente é feito usando um método chamado mínimos quadrados, que minimiza a soma das diferenças quadradas entre os valores previstos e reais (HASTIE; FRIEDMAN, 2009).

Uma das limitações da regressão linear é a suposição de uma relação linear entre a variável dependente e independente, o que pode não ser adequado em todos os casos. Além disso, a presença de valores atípicos nos dados pode comprometer os resultados, afetando a precisão do modelo. Nesses casos, outras técnicas que podem ser aplicadas para reduzir a influência dos *outliers* e melhorar a qualidade do modelo (HASTIE; FRIEDMAN, 2009).

2.2.2 Floresta Aleatória

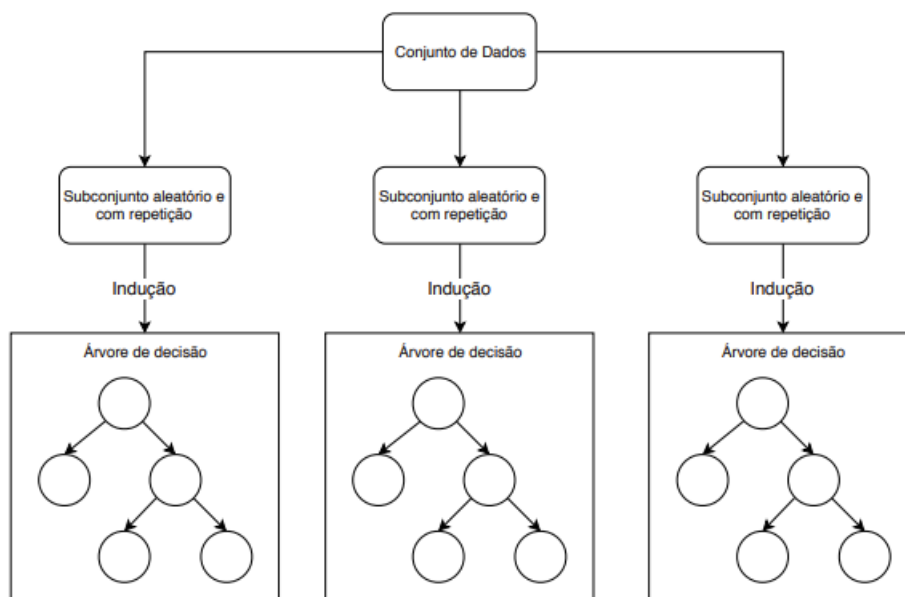
A *Random Forest*, em português Floresta Aleatória, é um método de conjunto que combina várias árvores de decisão para fazer previsões. Ele usa uma técnica chamada *bagging*, em que várias árvores de decisão são criadas usando subconjuntos aleatórios de dados e as previsões são combinadas para obter a previsão final (BREIMAN, 2001).

Uma árvore de decisão é uma estrutura de árvore semelhante a um fluxograma, onde um nó interno representa o recurso e cada nó folha representa uma decisão. O nó mais alto em uma árvore de decisão é conhecido como nó raiz. Ele aprende a particionar os dados em subconjuntos.

Segundo (PANDIMURUGAN, 2021), com o aumento do número de árvores, maior será a precisão do resultado previsto, devido à média de todas estas árvores de decisão ser utilizada no aumento da eficiência e da precisão do poder preditivo do algoritmo.

Florestas aleatórias demonstraram ter bom desempenho em uma variedade de aplicações, incluindo classificação de imagens, reconhecimento de fala e bioinformática (ZHANG H; ZHU, 2019). Além disso, florestas aleatórias são relativamente fáceis de usar e não requerem muito ajuste, pois possuem poucos hiperparâmetros e não são sensíveis à escolha de hiperparâmetros, uma vez que eles não são aprendidos durante o treina-

Figura 8 – Diagrama de Floresta Aleatória



Fonte: (PANDIMURUGAN, 2021)

mento do modelo, mas sim definidos antes do treinamento e afetam o comportamento e a performance do modelo.

Por outro lado, existem algumas limitações para o uso de florestas aleatórias. Um dos problemas apontados é que o tamanho do modelo pode se tornar muito grande, levando a longos tempos de treinamento e previsões lentas, especialmente para grandes conjuntos de dados. Além disso, a interpretação do modelo pode ser difícil, pois a contribuição de cada característica para a previsão final não é tão clara quanto em uma única árvore de decisão (NICULESCU-MIZIL, 2006).

A ideia básica por trás da *Random Forest Regression* é criar um grande número de árvores de decisão usando subconjuntos aleatórios de dados e recursos e, em seguida, calcular a média de suas previsões para obter a previsão final. Esse processo de média ajuda a reduzir a variação nas previsões e leva a um modelo mais robusto (BREIMAN, 2001).

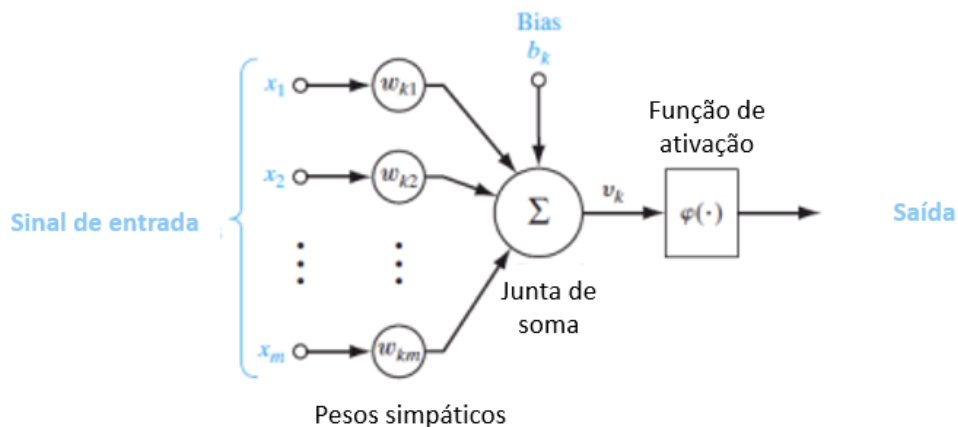
Logo, *Random Forest Regression* pode lidar com dados de alta dimensão e também não é sensível a outliers e também pode lidar com variáveis categóricas. Mas pode ser computacionalmente caro se o número de árvores for grande e também pode ser difícil de interpretar o modelo.

2.2.3 Redes Neurais

As Redes Neurais Artificiais, em inglês Artificial Neural Networks (ANN), são um tipo de modelo de aprendizado de máquina inspirado tanto na estrutura quanto na

função dos neurônios biológicos, de forma simples, (HAYKIN, 2009) define-a como uma máquina desenhada como um cérebro humano performa. Eles são compostos de camadas de "neurônios" interconectados, que processam e transmitem informações. A Figura 9 exemplifica a estrutura onde cada neurônio recebe entrada de outros neurônios, executa uma computação nessa entrada e então passa o resultado para outros neurônios na próxima camada. Dentre os benefícios trazidos pela rede neural destacam-se: não linearidade, adaptabilidade.

Figura 9 – O modelo do neurônio



Fonte: (HAYKIN, 2009)

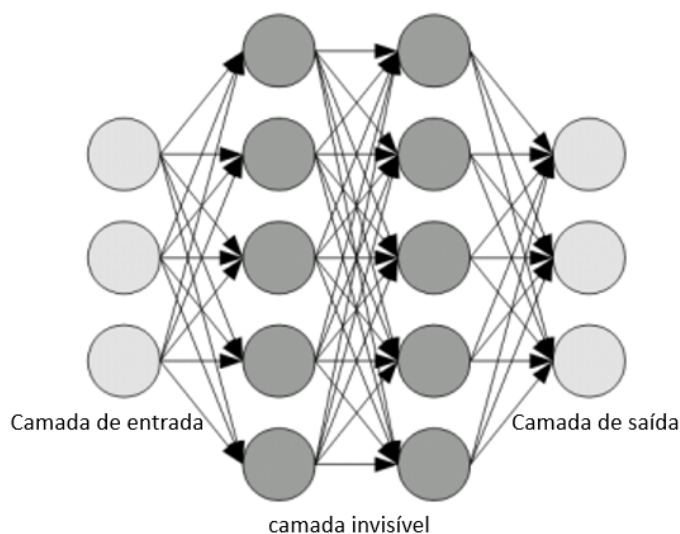
Uma de suas vantagens importantes das redes neurais é a de não existir a necessidade de confirmar o modelo antecipadamente. Ao treinar os dados de entrada da amostra, ela se adapta para reproduzir a saída. A ANN se destaca em funcionar bem para modelar a relação não linear por causa de suas características de regressão semi paramétrica (WANG; LI, 2019).

A Figura 10 apresenta a estrutura de uma rede neural artificial que se conecta a partir da camada de entrada, utiliza-se das camadas invisíveis e como retorno possui a camada de saída.

Uma qualidade importante é o fato da ANN poder ser usada para tarefas de aprendizado supervisionadas e não supervisionadas e tem a capacidade de aprender relacionamentos não lineares. (HAYKIN, 2009).

Segundo (HASTIE; FRIEDMAN, 2009), o modelo geralmente é super parametrizado, dessa forma sua otimização é dificultosa e instável. Alguns dos problemas mais recorrentes são: valores iniciais, sobre-ajustes, inputs e números de unidades e camadas escondidas.

Figura 10 – Uma Rede Neural Artificial totalmente conectada



Fonte: (AGGARWAL, 2018)

2.3 *Machine Learning* para Previsão de Preços de Imóveis

Em seu estudo de 104 artigos sobre a utilização de modelos para o cálculo dos preços de imóveis no século XXI, (WANG; LI, 2019) conclui que a tendência de utilização de modelos mistos de metodologias ganha cada vez mais força, uma vez que mais métodos podem ser testados e a inteligência artificial evolui juntamente com outras ferramentas como o Geo processamento.

No Brasil, (AFONSO et al., 2019), testa o uso de dois métodos diferentes de forma separada e realiza um novo teste combinando ambas para a previsão de preços, destaca-se o resultado apontado de que dois métodos resultou no menor valor de Raiz do erro de log quadrático médio (RMSLE, sigla em Inglês). Ainda é mostrado que ambas as técnicas utilizadas obtiveram um valor com baixo (RMSLE).

O modelo utilizado por (SING; YANG JESSE, 2020) para o mercado de Cingapura, obtém um valor com menor erro devido ao fato de ter utilizado um banco de dados maior, cujos registros superam a marca de 30 mil movimentações.

Os modelos mesmo em diferentes localizações geográficas podem ser utilizados para prever preços de imóveis (SING; YANG JESSE, 2020), uma vez que o mercado imobiliário possui características semelhantes ao redor do mundo, porém os resultados terão variância de acordo com o mercado inserido, devido à suas variáveis (GOMES; MACIEL; KUWAHARA, 2012).

Investiga-se a eficiência de diferentes modelos de aprendizado de máquina na previsão de preços de imóveis. Os autores testaram vários modelos, incluindo regressão linear, árvores de decisão, florestas aleatórias e redes neurais, e compararam seus resultados

com o objetivo de identificar o modelo mais preciso. Eles concluíram que o modelo de rede neural obteve os melhores resultados na previsão de preços de imóveis, apresentando uma acurácia significativamente maior em comparação aos outros modelos testados. Além disso, os autores destacam a importância da seleção cuidadosa dos dados de treinamento e da validação para garantir resultados precisos na previsão de preços de imóveis (GUPTA, 2019).

Examina-se a aplicabilidade de técnicas de aprendizado de máquina para prever preços de imóveis (NANDA; KUMAR; BHARTI, 2019). Os autores coletaram dados de uma grande cidade indiana e os utilizaram para treinar e avaliar diferentes modelos de aprendizado de máquina, incluindo regressão linear, árvore de decisão, floresta aleatória e rede neural. Eles descobriram que a floresta aleatória foi o modelo mais preciso na previsão de preços de imóveis, seguido de perto pela rede neural. Além disso, os autores destacam a importância de considerar variáveis relevantes, como localização, tamanho e idade do imóvel, na previsão de preços. Eles concluem que as técnicas de aprendizado de máquina podem ser eficazes na previsão de preços de imóveis, mas que a escolha do modelo adequado e a consideração de todas as variáveis relevantes são fundamentais para garantir resultados precisos.

Não existe um método universalmente aceito de aprendizado de máquina como o melhor para prever preços imobiliários. O melhor método pode variar de - dependendo de muitos fatores, incluindo a quantidade e a qualidade dos dados disponíveis, características específicas do mercado imobiliário em questão e objetivos específicos previsões. Alguns métodos comumente usados incluem regressão linear, z-tree tomada de decisão, floresta aleatória e redes neurais. Qualquer um desses métodos pode ser apropriado em diferentes circunstâncias e a escolha do melhor método pode ser determinada por experimentos cuidadosos e avaliação dos resultados obtidos. É importante enfatizar que nenhum método é perfeito e a precisão da previsão pode ser afetada por muitos fatores além da escolha do método, incluindo a qualidade e quantidade dos dados utilizados.

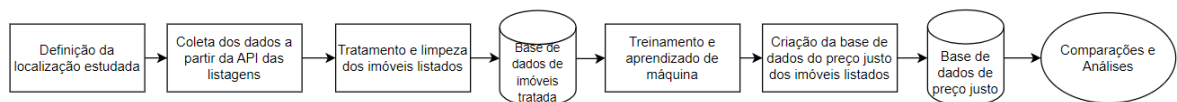
3 Metodologia

No presente estudo, foram comparadas a utilização de 3 métodos de aprendizado de máquina e, por fim, utilizou-se o modelo com menor RMSE para o prosseguimento das análises de preços para a cidade de Brasília - DF e São Paulo - SP.

A pesquisa pode ser determinada como um conjunto de ações as quais têm a finalidade de encontrar a solução de um problema por meio de procedimentos racionais e sistemáticos (OLIVEIRA, 2011). A natureza de uma pesquisa é aplicada para gerar conhecimento para aplicação prática do estudo, o qual é voltado para solução de problemas. Nesse viés, a metodologia descrita tem como objetivo a solução do problema da ausência de ferramentas em sites de listagens para determinação do preço de listagem coeso.

A Figura 11, apresenta o diagrama de etapas para um projeto de avaliação em *machine learning*. Destaca-se que o primeiro passo foi a definição da localização estudada, em seguida realizou-se a coleta de dados em sites de listagens de imóveis. O passo seguinte foi realizar a limpeza e tratamento para se estabelecer a base de dados funcional a ser trabalhada. Em seguida realizou-se o treinamento e aprendizado de máquina juntamente com a definição de qual modelo a ser adotado. A partir dos resultados, montou-se uma base de dados do preço justo calculado e então utilizou-se deles para compará-los com os dados encontrados no mercado:

Figura 11 – Diagrama de etapas da metodologia para a precificação de imóveis



Fonte: De autoria própria

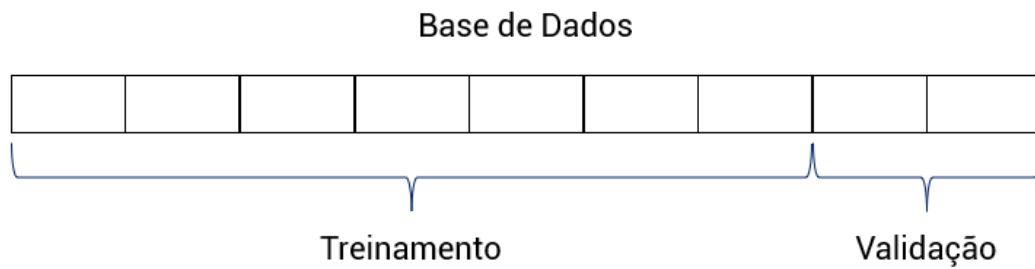
3.1 Procedimento de Coleta de Dados

A coleta de dados foi realizada por meio de uma API do site de listagens Quinto Andar. As informações estruturadas foram armazenadas em uma base de dados no formato CSV para serem tratadas, embora poderiam ser exportadas para outros formatos.

3.2 Tratamento de Dados

De acordo com (MULLER, 2017), ao trabalhar com um novo conjunto de dados em geral é uma boa ideia começar com modelos simples e ver o quão longe é possível chegar.

Figura 12 – Diagrama de execução da base de dados



Fonte: De autoria própria

Depois de entender mais sobre os dados, pode-se mover para modelos mais complexos.

Para a manipulação da base de dados, será utilizado o Python. Devido à sua grande biblioteca e adeptos, se torna cada vez mais utilizado e líder na utilização dentre as linguagens (KDNUGETS, 2020).

A Figura 13 apresenta a base de dados que serão gerada a partir dos dados obtidos das listagens de imóveis, assim possuindo diferentes métricas dentro de um mesmo imóvel:

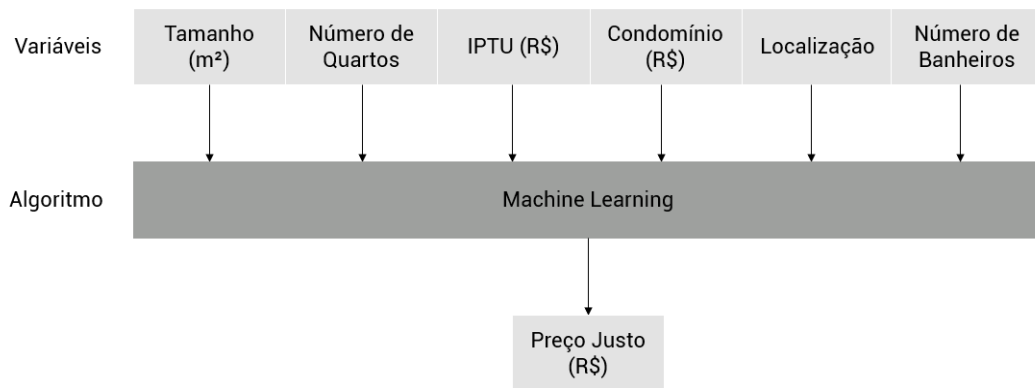
Figura 13 – Tabela representando a base de dados

	Imóvel 1	Imóvel 2	Imóvel 3	Imóvel 4	Imóvel 5
Preço (R\$)	2.600	2.200	950	2.700	3.500
Tamanho (m ²)	82	113	29	93	140
Número de Banheiros	1	2	1	2	4
IPTU (R\$)	752	801	833	904	896
Condomínio (R\$)	321	500	400	453	820
Localização	Asa Sul	Guará	Asa Norte	Lago Norte	Sudoeste
Número de Quartos	2	3	1	2	3

Fonte: De autoria própria

Cada imóvel será utilizado para a determinação de um novo valor que corresponderá ao seu preço justo de acordo com o modelo utilizado. A Figura 14, apresenta-se o diagrama de como as variáveis irão interagir com o modelo para gerar o resultado do preço justo

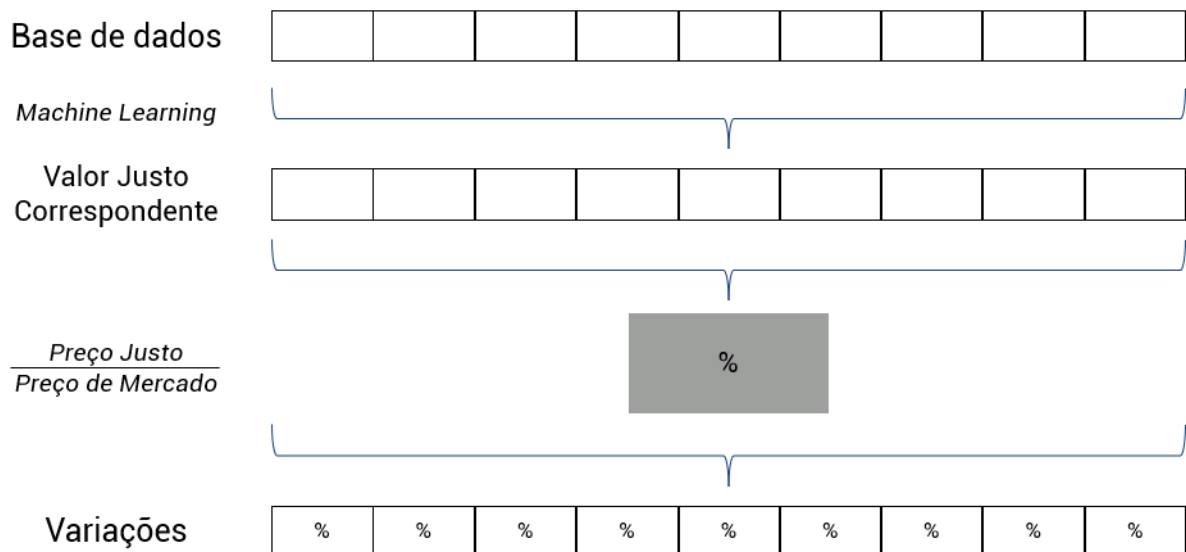
Figura 14 – Diagrama representando a criação do preço justo



Fonte: De autoria própria

A Figura 15 apresenta o diagrama de etapas para que os dados obtidos previamente sejam transformados em uma nova base de dados, agora com o seu valor justo correspondente. Para exemplificar, ao encontrar o valor de um imóvel, será gerado um preço justo correspondente às suas informações como tamanho, número de quartos e outros. A partir desse valor gerado, compara-se com o preço de listagem e assim é possível verificar a variação dos preços na cidade em comparação com o preço justo do modelo.

Figura 15 – Diagrama representando ações



Fonte: De autoria própria

3.3 Linguagem Utilizada

A linguagem de programação Python (3.11.2) foi utilizada e as bibliotecas *pandas*, para o tratamento e análises, *Scikit-learn* para o aprendizado de máquina, *numpy* para análises e tratamentos e *matplotlib* para a plotagem dos gráficos foram utilizadas. Para implementar os algoritmos de aprendizado de máquina.

3.4 Máquina Virtual

Para a elaboração do aprendizado de máquina, utilizou-se da máquina virtual no VirtualBox, executando o sistema operacional Linux, para criar um ambiente de aprendizado de máquina com o objetivo de determinar o preço justo do aluguel de imóveis.

Após a sua criação, configura-se a máquina virtual, instalando as bibliotecas e pacotes necessários para o desenvolvimento do modelo de aprendizado de máquina. Isso inclui *frameworks*, como o *Pandas*, *scikit-learn* e *numpy*, que oferecem uma ampla gama de funcionalidades para a construção e treinamento de modelos preditivos.

3.5 Containers

Containers são ambientes isolados que permitem empacotar um software e suas dependências em uma unidade portátil, fornecendo uma maneira consistente e confiável de executar aplicativos em diferentes sistemas operacionais. No caso do aprendizado de máquina, os containers podem ser usados para criar ambientes de trabalho consistentes, onde as configurações de software, bibliotecas e pacotes são mantidos de forma padronizada.

Os containers têm se mostrado uma solução eficiente para mitigar erros no sistema, oferecendo maior segurança e estabilidade para a execução de aplicativos, incluindo aqueles voltados ao aprendizado de máquina na determinação de preços de imóveis.

Uma das principais razões pelas quais os containers são menos suscetíveis a erros no sistema é devido ao seu isolamento. Cada container é uma unidade isolada que empacota o aplicativo e suas dependências de forma independente. Isso significa que qualquer modificação ou erro dentro de um container não afetará os outros containers ou o sistema operacional hospedeiro.

Essa abordagem isolada evita conflitos entre bibliotecas e dependências. No contexto do aprendizado de máquina, onde diferentes bibliotecas e pacotes são frequentemente necessários, a utilização de containers permite manter ambientes de trabalho consistentes e evita problemas de compatibilidade entre diferentes versões de bibliotecas ou pacotes. Dessa forma, reduz-se a probabilidade de erros relacionados a dependências, pois cada container tem suas próprias versões específicas e isoladas.

Além disso, os containers oferecem um ambiente controlado e reproduzível. Os containers são criados a partir de imagens pré-configuradas que contêm todas as dependências necessárias. Essas imagens são criadas com base em arquivos de definição específicos, como o Dockerfile, que especificam exatamente quais bibliotecas e pacotes devem ser instalados no container. Isso garante que o ambiente de trabalho seja reproduzido de maneira consistente, evitando configurações não intencionais ou erros de instalação durante a criação do ambiente.

3.6 Docker

O Docker é uma plataforma de código aberto que permite empacotar aplicativos e suas dependências em containers leves e portáteis. Esses containers contêm tudo o que é necessário para executar o software, incluindo bibliotecas, pacotes e configurações específicas do sistema. No contexto do aprendizado de máquina, o Docker pode ser usado para criar ambientes isolados e consistentes, onde os modelos e algoritmos podem ser executados de forma padronizada.

Além disso, o Docker facilita o gerenciamento e a manutenção dos ambientes de trabalho. As imagens Docker podem ser facilmente atualizadas ou revertidas para versões anteriores, garantindo a consistência do ambiente e facilitando a experimentação com diferentes configurações e bibliotecas. Isso é especialmente útil no aprendizado de máquina, onde diferentes versões de bibliotecas e *frameworks* podem ter um impacto significativo nos resultados do modelo.

Em resumo, a utilização do Docker oferece benefícios como facilidade de criação e distribuição de imagens, portabilidade, escalabilidade e facilidade de gerenciamento. Essa abordagem permitiu um desenvolvimento mais eficiente, contribuindo para avanços mais rápidos na área de precificação imobiliária.

3.7 Métricas de Avaliação

Por fim, os valores serão calculados para gerar mapas de calor e indicadores de rentabilidade para cada imóvel (DEBASTIANI; RUSSO, 2008), assim gerado o valor de índice de preço dos imóveis.

Para o cálculo dos imóveis acima do valor de mercado, se dá por seu valor justo ser maior do que o valor de mercado, dessa forma, caracteriza-se como menos atrativo, como mostrado na equação 3.1:

$$\text{Imóveis Supervalorizados} = \text{Preço justo} < \text{Preço de mercado} \quad (3.1)$$

Por outro lado, quando o valor calculado do preço justo é maior do que o preço de mercado, o imóvel é subvalorizado, como mostrado em 3.2:

$$\text{Imóveis Subvalorizados} = \text{Preço justo} > \text{Preço de mercado} \quad (3.2)$$

Outra análise a ser feita se dá pelo cálculo do Yield, ou seja, o valor percentual recebido do aluguel comparado com preço de venda de mercado, como na equação 3.3:

$$\text{Yield} = \frac{\text{Aluguel} \times 12}{\text{Valor de venda}} \quad (3.3)$$

Além disso, é possível analisar o potencial de Yield ao utilizar o preço justo calculado, como mostrado na equação 3.4:

$$\text{Yield Projetado} = \frac{\text{Valor justo de aluguel} \times 12}{\text{Valor de venda}} \quad (3.4)$$

3.8 Implementação e Testes

Após o desenvolvimento e ajuste do modelo, será realizada a implementação em um sistema ou aplicação que possa receber os dados de um imóvel e prever o preço justo. Testes extensivos serão conduzidos para avaliar a precisão e o desempenho do modelo em um ambiente de produção. Dentre as métricas utilizadas para avaliar o modelo estão o RMSE, MSE e variação dos preços.

Para verificar o modelo a ser utilizado, todos os modelos passaram por testes de estresse para resultar na melhor metodologia a ser utilizada, dentre os testes e variações de modelagem estão:

- Variáveis Utilizadas;
- Tamanho da amostra;
- Parâmetros para regressão linear múltipla;
- Número de *hidden layers*;
- Número de iterações;

4 Resultados

Este capítulo se inicia com os resultados das simulações para a previsão do preço justo dos aluguéis dos imóveis do Distrito Federal e São Paulo. Para isso, realiza-se a comparação da utilização dos 3 diferentes modelos de *Machine Learning*: Regressão Linear, Random Forest e Redes Neurais.

Em primeira análise, apresenta-se a base de dados utilizada para o aprendizado de máquina. Foram extraídas as listagens de imóveis tanto de aluguel quanto de venda para duas localizações, Brasília - DF e São Paulo - SP, como apresentado na Tabela 2:

Tabela 2 – Número de imóveis em Brasília - DF e São Paulo - SP

	Brasília - DF	São Paulo - SP
Imóveis para alugar	312	415
Imóveis para vender	0	164

Fonte: De autoria própria

Destaca-se que a aquisição de imóveis listados para venda em Brasília pela plataforma, resultou em zero imóveis, dessa forma, as análises de venda serão realizadas para São Paulo, visto que possui imóveis listados a venda.

Para efeito de comparação, realiza-se o cálculo do RMSE, onde para cada unidade apresentada do erro, representa um desvio de R\$ 1,00 com relação ao valor do imóvel. Assim, apresenta-se na Tabela 3 que o menor RMSE foi do modelo de Regressão Linear, dessa forma, utilizar-se-á dele para o prosseguimento das análises.

Tabela 3 – Comparação dos modelos de aprendizado de máquina

Modelo	MSE	RMSE
Linear Regression	7.048	83
Random Forest	109.349	330
Neural Network	573.205	757

Fonte: De autoria própria

Os modelos de aprendizado são comparados um ao outro. Evidencia-se o modelo de regressão linear, apresenta uma menor dispersão dos valores analisados, em seguida apresenta-se o modelo de Random Forest, que possui uma divergência dos valores encontrados apresenta-se o modelo de Redes Neurais, cujo resultado, assim como o anterior, representa uma maior dispersão quando comparado com a Regressão Linear.

4.1 Brasília

O primeiro resultado avaliado são as correlações das variáveis obtidas, na Figura 16. dentre as variáveis pode-se listar:

- Latitude: latitude que se encontra o imóvel;
- Longitude: longitude que se encontra o imóvel;
- Quartos: número de quartos que o imóvel possui;
- Garagem: número de vagas de garagem que o imóvel possui;
- Área: tamanho do imóvel em m²;
- Aluguel: valor do aluguel;
- Custo Total: soma de todos os custos do imóvel.

Assim, realiza-se uma análise para determinar qual é a variável que exerce maior influência nos resultados do preço justo. Dessa forma, constata-se que a variável Custo Total é a que possui maior correlação com o valor do imóvel, seguido da área do imóvel. Além dessas variáveis é apresentado a matriz de correlação Pearson, ela quantifica a relação linear entre a variável *rent* e as demais.

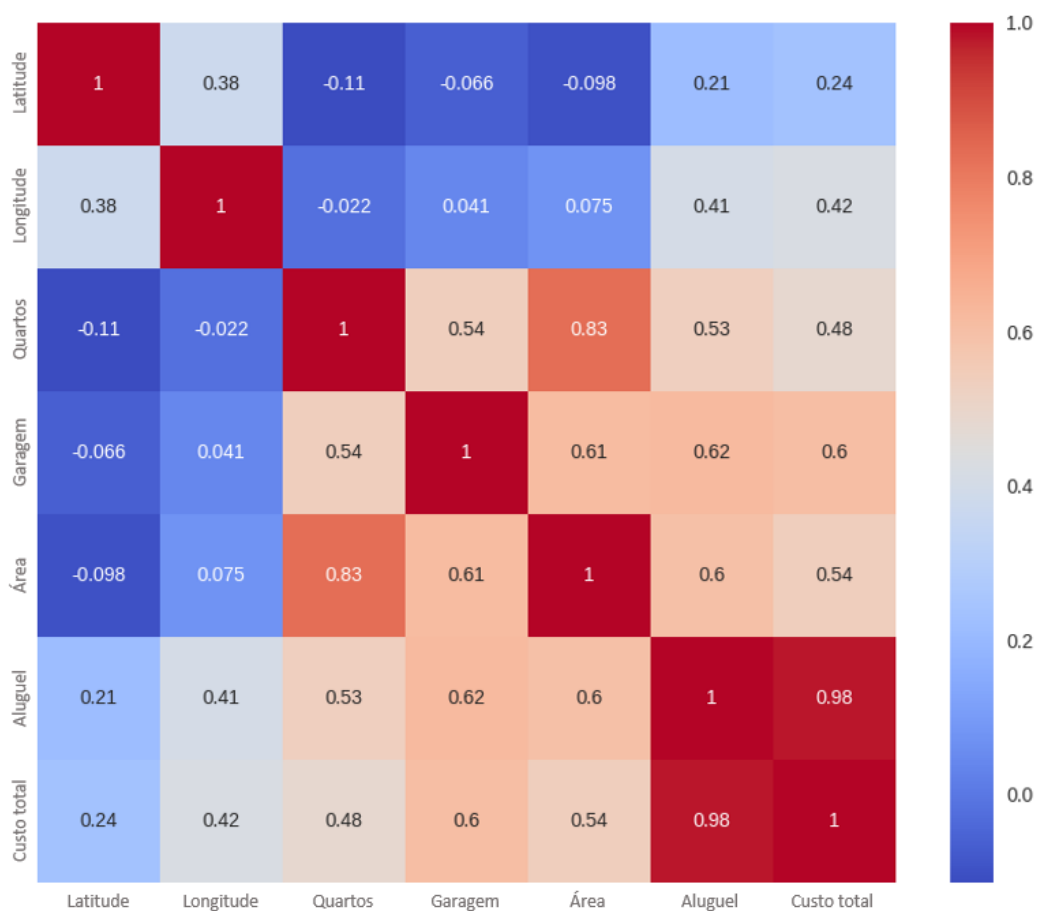
Figura 16 – Correlação de Pearson em Brasília



Fonte: De autoria própria

Seguindo na análise, na Figura 17 realiza-se a construção da Matriz de Spearman que possui uma relação monotônica entre duas variáveis. Ela é uma alternativa à matriz de correlação de Pearson quando a relação entre as variáveis não é necessariamente linear. Os resultados encontrados em ambas as matrizes são semelhantes e convergem para a a mesma ordem de correlação entre as variáveis:

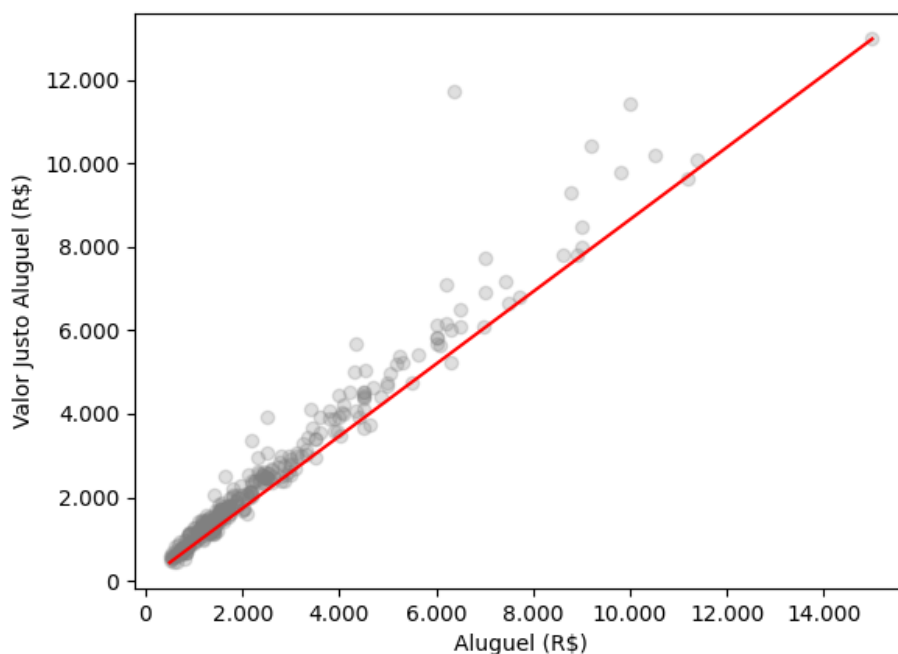
Figura 17 – Correlação de Spearman em Brasília



Fonte: De autoria própria

Em seguida, é apresentado o resultado do modelo aprendido de máquina de Regressão Linear para Brasília. Cada ponto na Figura 18 representa um imóvel, e a reta é a regressão linear entre os valores de mercado com os valores calculados.

Figura 18 – Reta da Regressão Linear: Aluguel vs Valor justo aluguel (RL) em Brasília



Fonte: De autoria própria

Na Tabela 4, é apresentado a lista das 10 maiores variações entre o preço justo de aluguel comparado com o preço do aluguel de mercado, destaca-se para a maior variação no imóvel no Jardim Botânico, com 83,6% de variação:

Tabela 4 – Maiores Variações de Valor Justo em Brasília

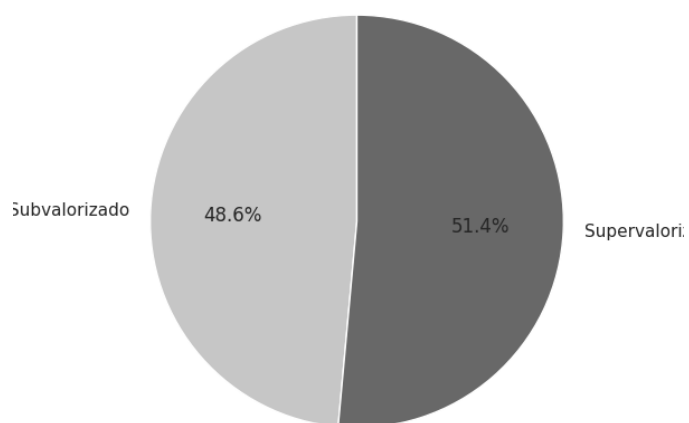
Aluguel(R\$)	Valor Justo (R\$)	Variação (%)	Bairro
6.375	11.706	83,6	Jardim Botânico
2.500	3.922	56,9	Asa Sul
2.200	3.364	52,9	Asa Norte
1.650	2.514	52,3	Asa Norte
1.400	2.047	46,2	Asa Norte
818	542	-33,7	São Sebastião
607	811	33,7	Ceilândia
646	437	-32,2	São Sebastião
4.350	5.666	30,2	SH Nova Colina
900	1.168	29,8	Asa Norte

Fonte: De autoria própria

Dessa forma, a Figura 19 apresenta as 2 faixas abordadas, onde a 48,6% do total de imóveis estão abaixo do valor de mercado, ou seja, o valor de mercado está abaixo do preço justo. Isso se configura como uma possível oportunidade para o aluguel, uma vez

que o modelo apresenta que se está pagando menos para receber um imóvel com valor que deveria ser cobrado mais.

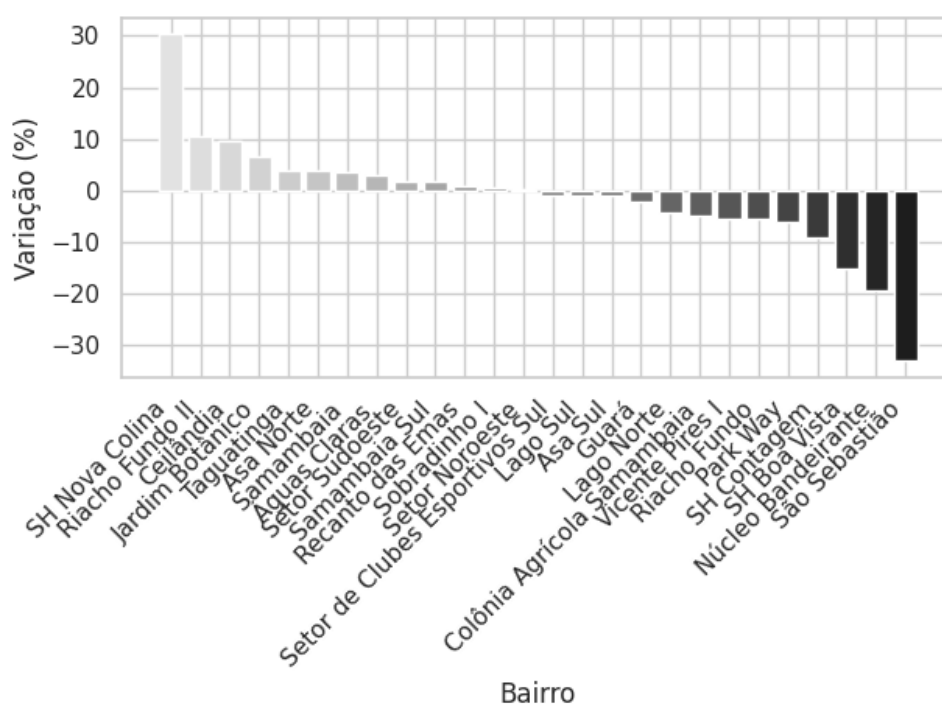
Figura 19 – Distribuição da variação dos valores em Brasília



Fonte: De autoria própria

Ademais, realiza-se a análise de quais bairros estão apresentando as maiores variações dos preços justos frente aos preços de mercado. Assim, a Figura 20 evidencia-se que os bairros na esquerda do gráfico caracterizam-se por possuir maior quantidade de imóveis com preços justos acima do valor de mercado, ou seja, são locais que estão supervalorizados. De forma inversa ocorre quando se caminha no gráfico para a direita, mostrando os bairros com subvalorizados.

Figura 20 – Bairros com maiores variações em Brasília



Fonte: De autoria própria

Nesse viés, os bairros que apresentam maior quantidade de imóveis supervalorizados são SH Nova Colina e Riacho Fundo II. Em contrapartida, os bairros que apresentam imóveis subvalorizados, ou seja, uma oportunidade para quem precisa alugar, são São Sebastião e Núcleo Bandeirante.

4.2 São Paulo

Para São Paulo, buscou-se utilizar uma quantidade maior de imóveis listados e de maior proximidade, dessa forma, o bairro de Pinheiros e seu entorno foi definido como fronteira de busca de listagens, assim resultando em 415 imóveis. Isso foi realizado de forma empírica, a fim de diminuir o erro e auxiliar na classificação de imóveis cuja localização é próxima uma da outra. A Figura 21 representa os limites utilizados para a nova base de dados.

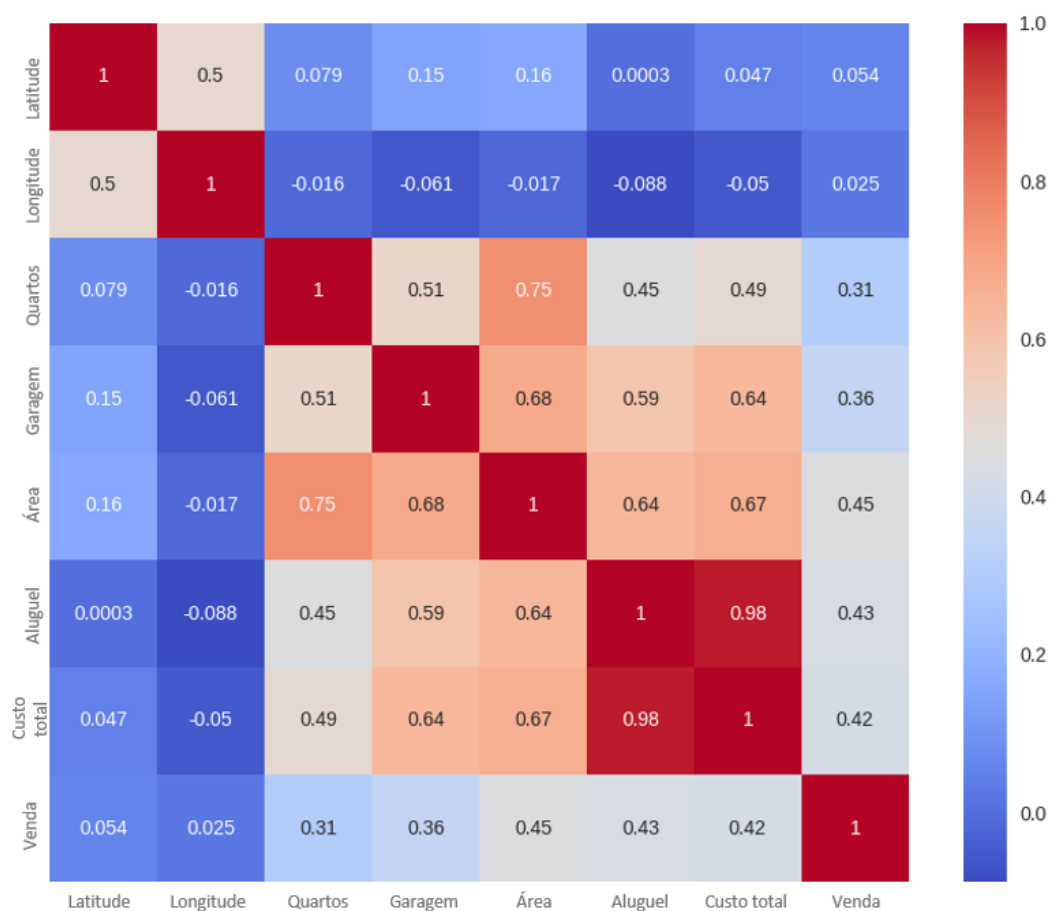
Figura 21 – Mapa do bairro de Pinheiros



Fonte: Google Maps

Com posse da base de dados, realiza-se uma análise de correlação para determinar qual é a variável que exerce maior influência nos resultados do preço justo de aluguel e venda, como apresentado anteriormente, porém para São Paulo, utilizou-se a variável de Venda, como mostrado na Figura 22:

Figura 22 – Correlação de Pearson em São Paulo

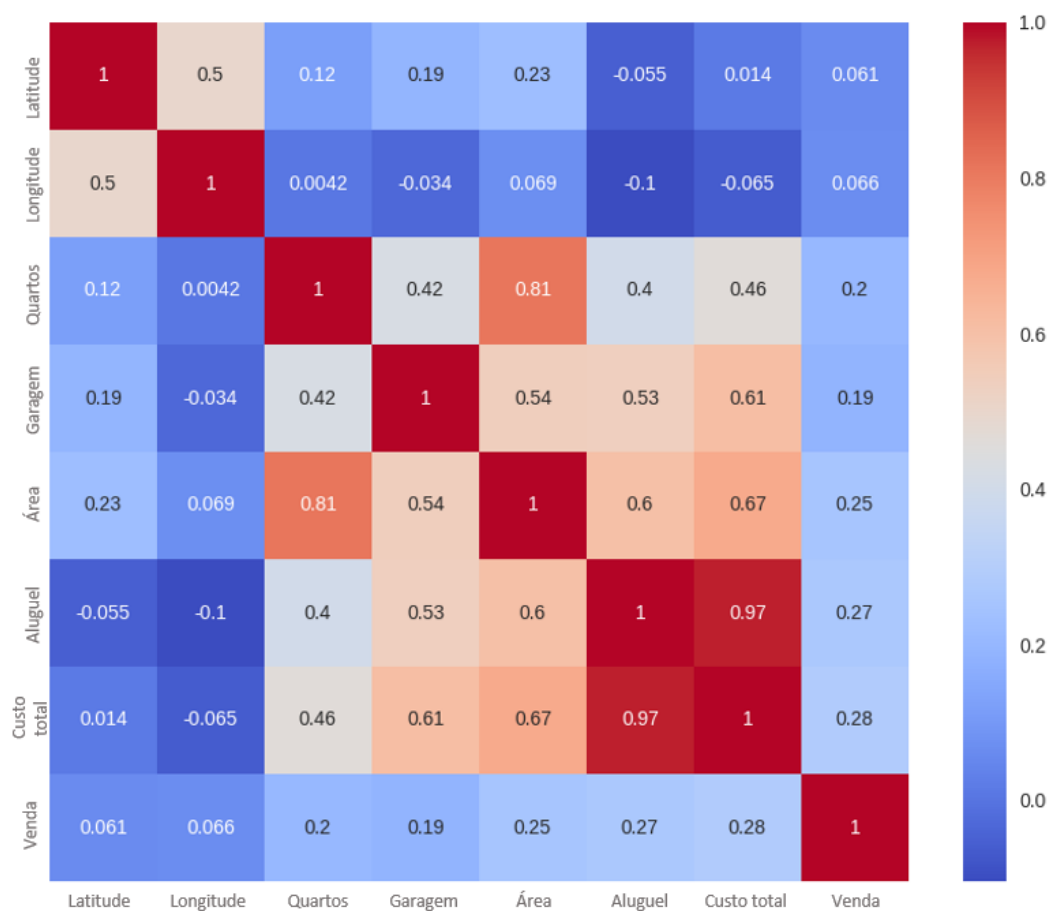


Fonte: De autoria própria

Como na matriz para Brasília, a variável Custo Total foi a que mais influenciou o resultado do modelo.

Seguindo na análise, na Figura 23 realiza-se a construção da Matriz de Spearman, como na matriz de Brasília, novamente Custo Total possui maior influência no valor do aluguel:

Figura 23 – Correlação de Spearman em São Paulo

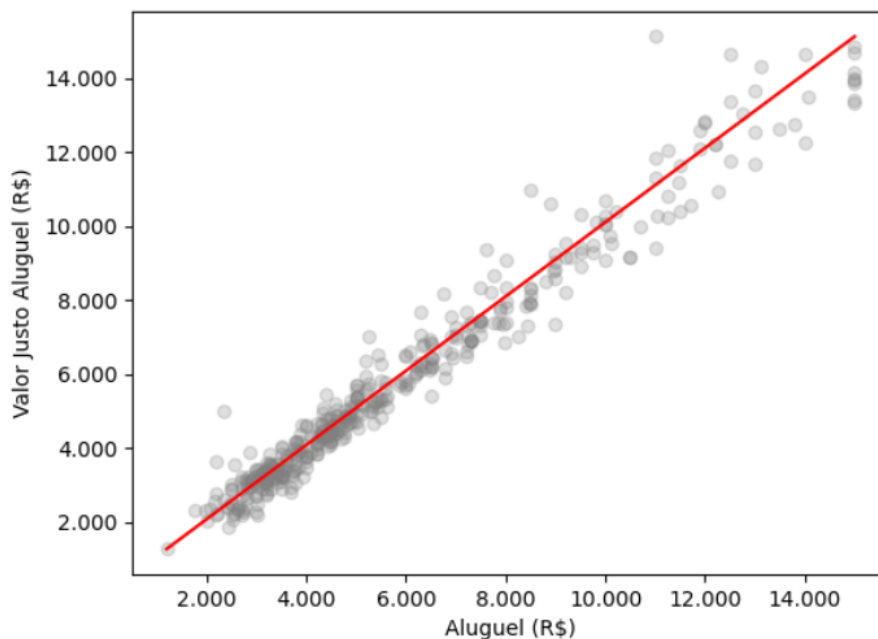


Fonte: De autoria própria

4.2.1 Imóveis listados para aluguel em São Paulo

A partir da base de dados de imóveis de São Paulo, identifica-se quais estão listados para aluguel, assim calculado os valores justos de aluguel, onde são apresentados na Figura 24, onde será apresentado a dispersão de todos a amostra e, por fim, realizada uma regressão linear.

Figura 24 – Reta da Regressão Linear: Aluguel vs Valor Justo Aluguel



Fonte: De autoria própria

Na Tabela 5, é apresentado a lista das 10 maiores variações entre o preço justo comparado com o preço de mercado, destaca-se que dois imóveis possuem variação acima de 30% e 4 imóveis de pelo menos -20% de variação:

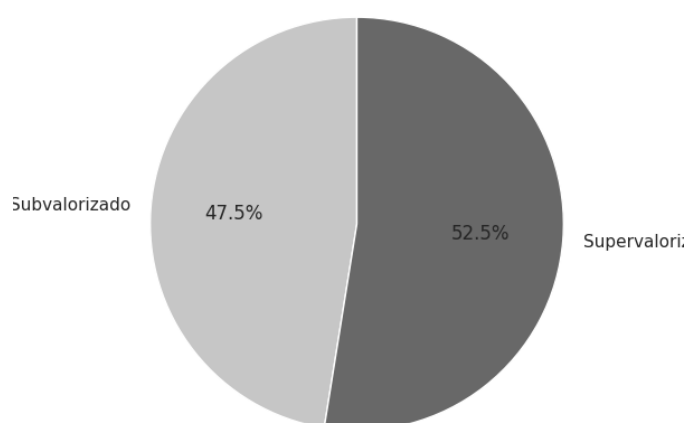
Tabela 5 – Maiores Variações de Valor Justo de aluguel em São Paulo

Aluguel(R\$)	Valor Justo Aluguel (R\$)	Variação (%)	Bairro
2.550	3.552	39,3	Pinheiros
2.860	3.908	36,6	Pinheiros
2.200	2.796	27,1	Vila Madalena
3.010	2.219	-26,2	Pinheiros
4.400	5.480	24,5	Pinheiros
3.000	2.289	-23,6	Pinheiros
3.700	2.843	-23,1	Pinheiros
2.430	1.872	-22,9	Vila Madalena
6.300	7.676	21,8	Vila Madalena
6.750	8.177	21,1	Pinheiros

Fonte: De autoria própria

Em seguida, a Figura 25 apresenta as 2 faixas de classificação, onde a 47,5% do total de imóveis estão abaixo do valor de mercado, ou seja, o valor de mercado está abaixo do preço justo. Isso se configura como uma possível oportunidade para o aluguel, uma vez que o modelo apresenta que se está pagando menos para receber um imóvel com valor que deveria ser cobrado mais.

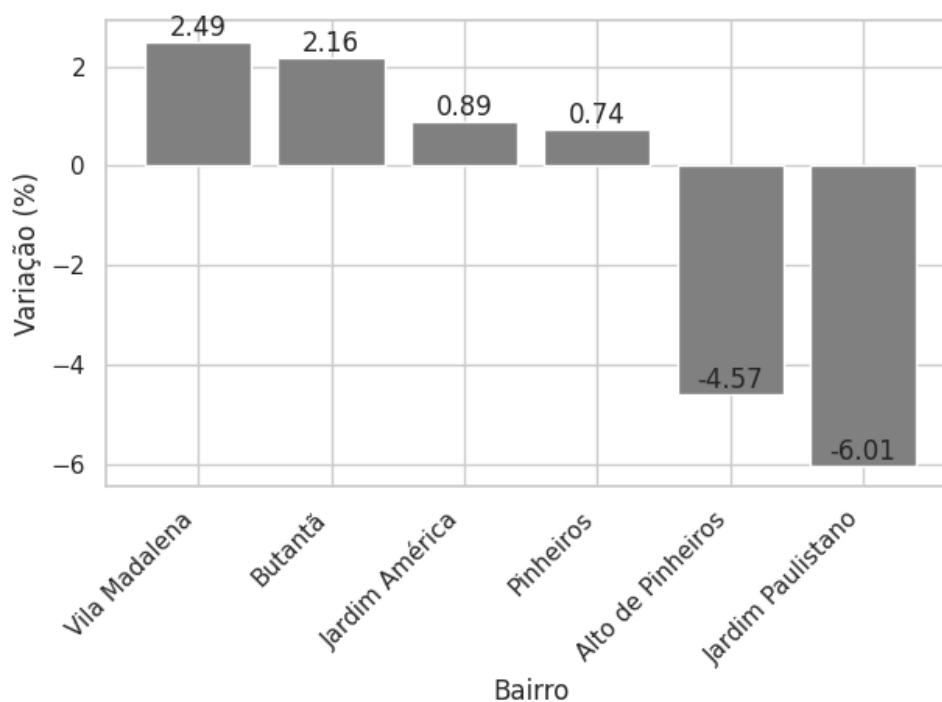
Figura 25 – Distribuição da variação dos valores em São Paulo



Fonte: De autoria própria

Ademais, realiza-se uma análise para determinar quais bairros estão apresentando as maiores variações dos preços justos de aluguel comparados aos preços de mercado. Assim, a Figura 26 evidencia-se que os bairros na esquerda do gráfico, como Vila Madalena e Butantã, caracterizam-se por possuir maior quantidade de imóveis com preços justos acima do valor de mercado, ou seja, são locais que estão supervalorizados. De forma inversa ocorre quando se caminha no gráfico para a direita, mostrando os bairros com subvalorizados, Jardim Paulista e Alto de Pinheiros.

Figura 26 – Variação de preço do aluguel justo vs preço de mercado por bairro em São Paulo

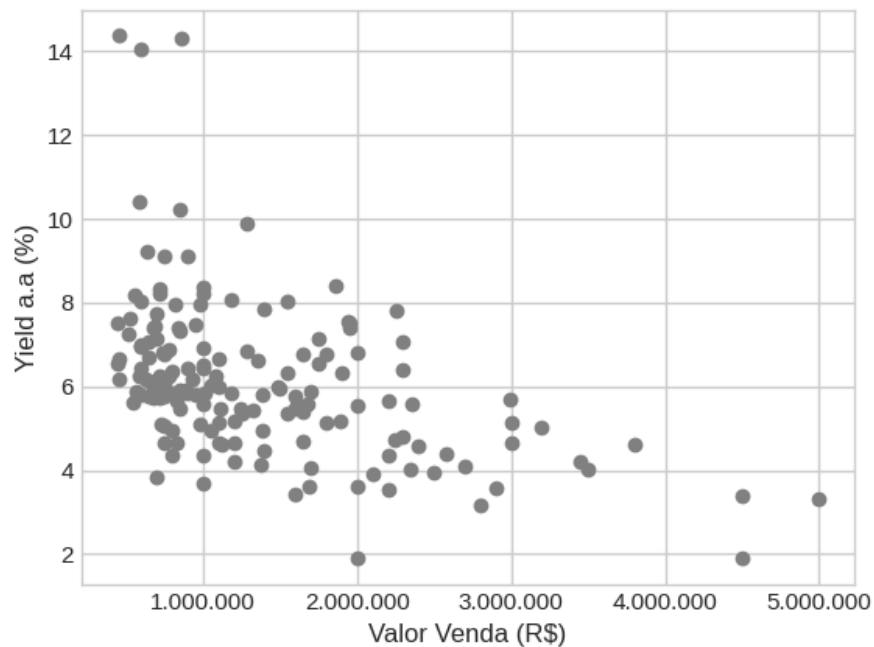


Fonte: De autoria própria

Em análise final, são verificadas possíveis oportunidades de imóveis a serem comprados para rentabilizar-se do aluguel recebido. Para isso, calcula-se a taxa de Yield como apresentado na fórmula 3.3 e comparado com o valor do imóvel listado. A Figura 27 apresenta a dispersão de todos os imóveis, destaca-se que imóveis no 2º quadrante, ou seja, no topo esquerdo, são as melhores oportunidades, visto que possuem maior yield e menor valor.

Dessa forma, seu investimento é de menor valor, e possui maior retorno percentual, ou seja, seu retorno ocorre de forma mais rápida e necessitando de menos capital.

Figura 27 – Distribuição Yield projetado vs valor Venda em São Paulo



Fonte: De autoria própria

Nesse viés, apresenta-se as informações dos 10 imóveis classificados como melhores oportunidades para esse tipo de investimento, destaque para 5 imóveis com Yield de mais de 14%, valor acima da renda fixa, conforme apresentado na Tabela 6:

Tabela 6 – Maiores Yields em São Paulo

Venda(R\$)	Valor Justo Aluguel (R\$)	Yield a.a. (%)	Bairro
450.000	5.000	14,37	Pinheiros
860.000	11.250	14,31	Pinheiros
600.000	8.250	14,05	Vila Madalena
590.000	5.440	10,40	Pinheiros
850.000	7.000	10,22	Pinheiros
1.282.000	11.700	9,88	Vila Madalena
639.000	4.650	9,22	Butantã
899.000	6.500	9,13	Pinheiros
750.000	6.000	9,11	Pinheiros
1.860.000	12.750	8,41	Pinheiros

Fonte: De autoria própria

Como resultado da análise, é apresentado o imóvel líder no ranking de Yield, ou seja, a melhor oportunidade de investimento analisada, que pode ser encontrado no site de listagens Quinto Andar, como mostrado na Figura 33:

Figura 28 – Imóvel listado com maior Yield

quintoandar.com.br/imovel/892847072/alugar/apartamento-3-quartos-morro-dos-ingleses-sao-paulo?from_route=search_results&search_id=c2d8911822d011eeb8a382f6ad5aab99?&sea...

Rua, bairro ou código

Buscar Imóveis Para proprietários Para Imobiliárias Indicar imóveis Sobre Ajuda

Início > São Paulo > Morro dos Ingleses > Rua Ingleses > Imóvel 147072

Rua Ingleses
Morro dos Ingleses, São Paulo

120 m²

3 quartos (1 suite)

2 banheiros

1 vaga

Até 3º andar

Não aceita

Mobiliado

Metrô próx.

Imóvel 147072 Publicado há 9 meses

Aluguel R\$ 5.000
Condomínio R\$ 1.100
IPTU R\$ 292
Seguro incêndio R\$ 64
Taxa de serviço R\$ 129
Total R\$ 6.585

Agendar visita

Fazer proposta

Faça sua avaliação

E descubra em dois minutos se você está pré-aprovado para alugar esse imóvel.

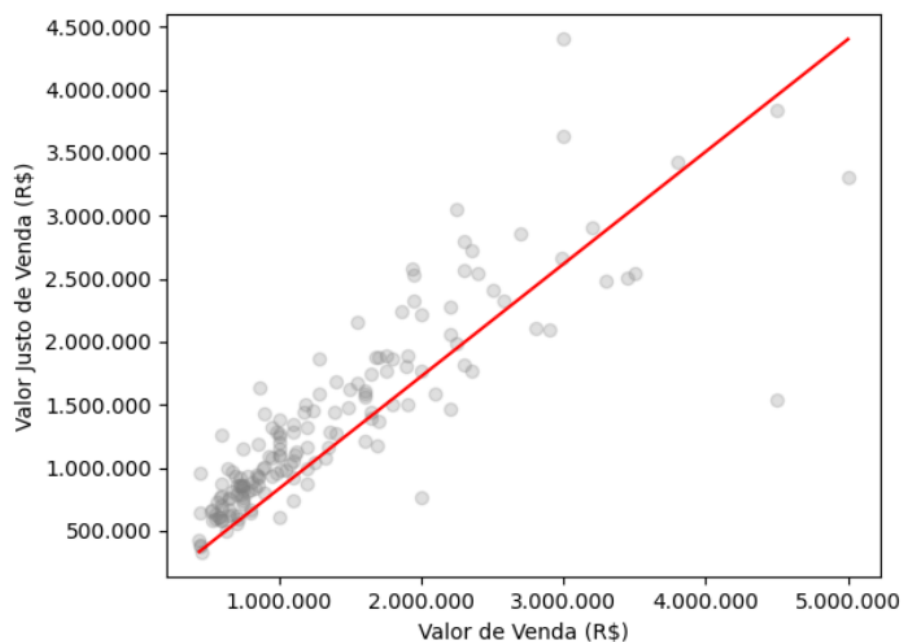
Favoritar Compartilhar

Fonte: Quinto Andar

4.2.2 Imóveis listados para venda

A partir da base de dados de imóveis de São Paulo, identifica-se quais estão listados para venda, assim calculado os valores justos de venda, onde são apresentados na Figura 29, onde será apresentado a dispersão de todos a amostra e, por fim, realizada uma regressão linear.

Figura 29 – Reta da Regressão Linear: Venda vs Valor Justo Venda em São Paulo



Fonte: De autoria própria

Na Tabela 7, é apresentado a lista das 10 maiores variações entre o preço justo de venda comparado com o preço de mercado, destaca-se que a maior variação no preço justo de venda comparado com o de mercado foi de 112% e o com maior variação negativa foi de -65,59%:

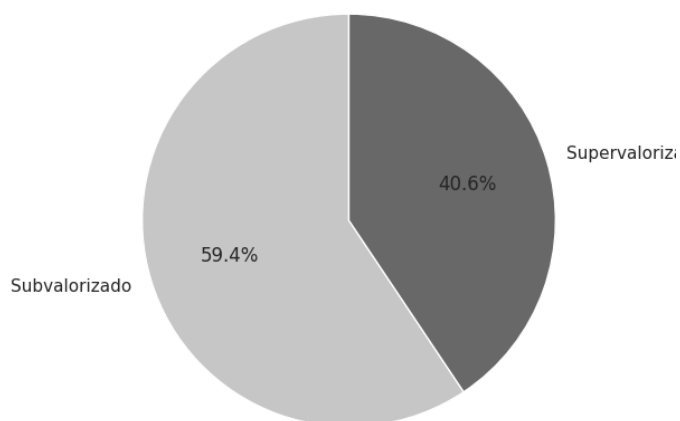
Tabela 7 – Maiores Variações de Valor Justo de venda em São Paulo

Venda(R\$)	Valor Justo Venda (R\$)	Variação (%)	Bairro
450.000	956.423	+112, 54	Pinheiros
600.000	1.248.768	+108, 13	Vila Madalena
860.000	1.603.579	+86, 46	Pinheiros
4.500.000	1.548.306	-65, 59	Pinheiros
2.000.000	765.222.8	-61, 74	Pinheiros
899.000	1.416.532	+57, 57	Pinheiros
639.000	1.005.674	+57, 38	Butantã
750.000	1.161.579	+54, 88	Pinheiros
675.000	994.546	+47, 34	Pinheiros
450.000	658.711	+46, 38	Butantã

Fonte: De autoria própria

Em seguida, a Figura 30 apresenta as 2 faixas de classificação, onde a 59,4% do total de imóveis estão abaixo do valor de mercado, ou seja, o valor de mercado está abaixo do preço justo. Isso se configura como uma possível oportunidade para a compra, uma vez que o modelo apresenta que se está pagando menos para receber um imóvel com valor que deveria ser cobrado mais.

Figura 30 – Representatividade da variação dos valores de venda em São Paulo

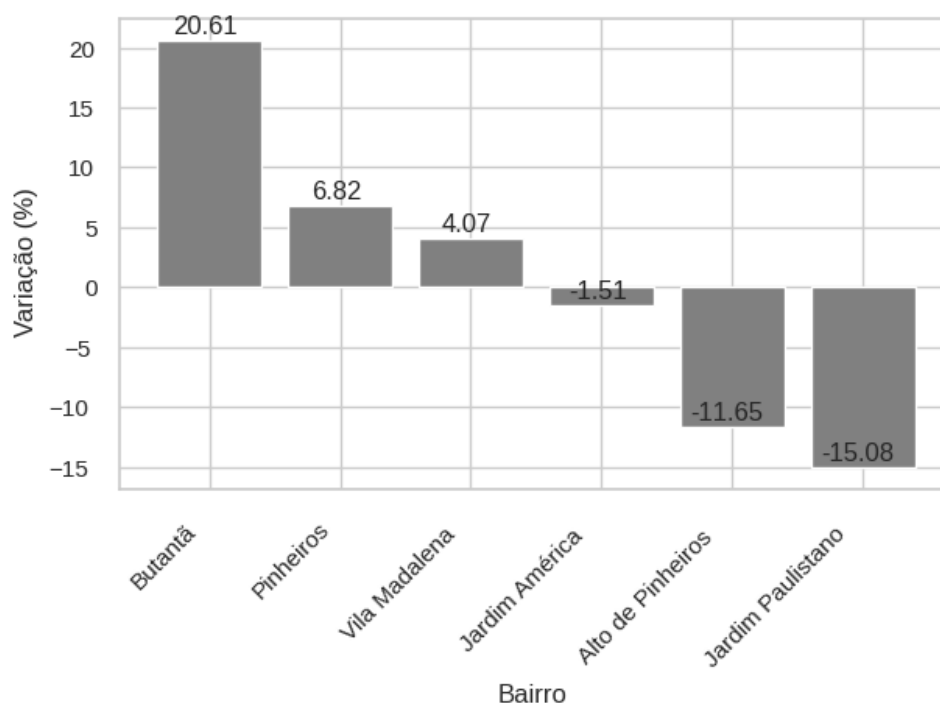


Fonte: De autoria própria

Ademais, realiza-se uma análise para determinar quais bairros estão apresentando as maiores variações dos preços justos de aluguel comparados aos preços de mercado.

Assim, a Figura 31 evidencia-se que os bairros na esquerda do gráfico caracterizam-se por possuir maior quantidade de imóveis com preços justos acima do valor de mercado, ou seja, são locais que estão supervalorizados. De forma inversa ocorre quando se caminha no gráfico para a direita, mostrando os bairros com subvalorizados.

Figura 31 – Bairros com maiores variações em São Paulo

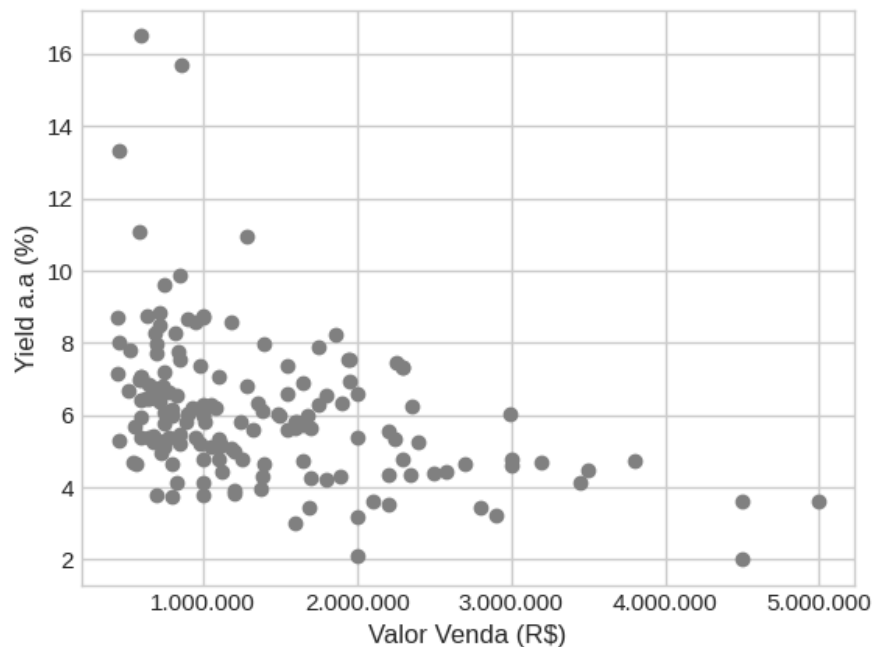


Fonte: De autoria própria

Em análise final, é analisado possíveis oportunidades de imóveis serem comprados para rentabilizar-se do aluguel recebido. Para isso, calcula-se a taxa de Yield como apresentado na fórmula 3.3 e comparado com o valor do imóvel listado. A Figura 32 apresenta a dispersão de todos os imóveis, destaca-se que imóveis no 2º quadrante, ou seja, no topo esquerdo, são as melhores oportunidades, visto que possuem maior yield e menor valor.

Dessa forma, seu investimento é de menor valor, e possui maior retorno percentual, ou seja, seu retorno ocorre de forma mais rápida e necessitando de menos capital. Destaca-se que o cálculo foi realizado com o valor justo do aluguel, dessa forma, apresenta-se um valor projetado.

Figura 32 – Distribuição Yield vs valor Venda em São Paulo



Fonte: De autoria própria

Nesse viés, apresenta-se as informações dos 10 imóveis classificados como melhores oportunidades para esse tipo de investimento, conforme apresentado na Tabela 8, destacasse 5 imóveis com Yield de pelo menos 10%:

Tabela 8 – Maiores Yields Projetados São Paulo

Yield Projetado a.a. (%)	Valor Venda (R\$)	Aluguel (R\$)	Bairro
16,5	600.000	8.250	Pinheiros
15,7	860.000	11.250	Pinheiros
13,3	450.000	5.000	Pinheiros
11	590.000	5.440	Pinheiros
10,1	1.282.000	11.700	Vila Madalena
9,9	850.000	7.000	Pinheiros
9,6	750.000	6.000	Pinheiros
8,8	720.000	5.300	Vila Madalena
8,8	999.000	7.300	Pinheiros
8,7	639.000	4.650	Butantã

Fonte: De autoria própria

Como resultado da análise, é apresentado o imóvel líder no ranking de Yield Projetado, ou seja, a melhor oportunidade de investimento analisada, que pode ser encontrado no site de listagens Quinto Andar, como mostrado na Figura 33:

Figura 33 – Imóvel listado com maior Yield Projetado

The screenshot shows a real estate listing on the Quinto Andar website. The listing is for an apartment located at Rua Ingleses, Morro dos Ingleses, São Paulo. The listing ID is 147072 and it was published 9 months ago. The apartment features 120 m² of area, 3 bedrooms (1 suite), 2 bathrooms, and 1 parking space. It is located on the 3rd floor, is furnished, and has a nearby subway station. The total monthly cost is R\$ 6,585, which includes a rent of R\$ 5,000, a condominium fee of R\$ 1,100, IPTU of R\$ 292, fire insurance of R\$ 64, and a service fee of R\$ 129. The listing includes buttons for 'Agendar visita' (Schedule visit) and 'Fazer proposta' (Make offer), as well as options to 'Favoritar' (Favorite) and 'Compartilhar' (Share).

Fonte: Quinto Andar

5 Conclusão

A modelagem de regressão linear apresentou menor erro em relação às demais abordagens utilizadas. Isso significa que, para o conjunto de dados analisado, a regressão linear foi capaz de estimar o valor justo do aluguel e de venda de forma mais consistente e próxima aos valores reais.

Os resultados obtidos destacam a importância da escolha adequada do modelo de regressão, levando em consideração as características dos dados e a relação entre as variáveis. A regressão linear pode ser uma opção viável quando os dados possuem uma relação linear e não apresentam padrões complexos.

No entanto, é importante ressaltar que os resultados obtidos estão sujeitos às limitações da base de dados utilizada. É possível que a base de dados não seja totalmente representativa do mercado imobiliário, visto que o número de listagens analisadas não representa o total de imóveis disponíveis no mercado, o que pode impactar os resultados da modelagem. Recomenda-se, portanto, buscar uma base de dados mais robusta e abrangente em futuros trabalhos, a fim de obter resultados mais generalizáveis e confiáveis.

As análises realizadas também identificaram bairros subvalorizados, nos quais os valores justos estimados pelo modelo estão acima dos valores reais de aluguel. Essa informação pode ser relevante para investidores e tomadores de decisão no mercado imobiliário, pois indica oportunidades de investimento em bairros com potencial de valorização.

Como recomendação, sugere-se explorar a disponibilidade de uma base de dados mais ampla e diversificada, incluindo uma maior variedade de características imobiliárias, como infraestrutura, proximidade de serviços e facilidades, entre outros. Além disso, é importante considerar a inclusão de dados históricos e acompanhar a evolução dos preços ao longo do tempo para obter uma visão mais completa do mercado imobiliário.

Em resumo, os resultados da modelagem de regressão linear indicam que esse método pode ser útil na estimativa do valor justo de aluguéis, porém é necessário considerar as limitações da base de dados e buscar aprimorar a qualidade e a quantidade dos dados disponíveis. Isso permitirá uma análise mais precisa e abrangente do mercado imobiliário, fornecendo informações valiosas para investidores e profissionais do setor.

Referências

- 5ANDAR. 2023. *Quanto cobrar de aluguel*. Disponível em: <<https://mkt.quintoandar.com.br/quanto-cobrar-de-aluguel/>>. Citado na página 13.
- AFONSO, B. et al. Preço de imóveis em são paulo. In: *Housing Prices Prediction with a Deep Learning and Random Forest Ensemble*. [S.l.: s.n.], 2019. Citado na página 25.
- AGGARWAL, C. *Neural Networks and Deep Learning: A Textbook*. 1. ed. [S.l.]: Springer, 2018. Citado na página 25.
- BARTO, S. Aprendizado de máquina por reforço. 2018. Citado na página 21.
- BOURASSA, S.; CANTONI, E.; MARTIN, H. Spatial dependence, housing submarkets, and house prices. 2005. Citado na página 20.
- BREIMAN, L. Florestas aleatórias. In: *Random Forests*. [S.l.: s.n.], 2001. p. 2–4. Citado 2 vezes nas páginas 22 e 23.
- CHAPELLE, O. Semi-supervised learning. 2006. Citado na página 21.
- DANTAS, R. A. *Engenharia de Avaliações: uma introdução à metodologia científica*. 3. ed. [S.l.]: Pini, 2012. Citado na página 14.
- DEBASTIANI; RUSSO. Dividend yield. 2008. Citado na página 32.
- FIPE. 2023. *Relatório FIZEZAP*. Disponível em: <<https://www.fipe.org.br/pt-br/publicacoes/relatorios/#relatorio-fizezap>>. Citado na página 16.
- GOMES, A.; MACIEL, V.; KUWAHARA, M. Determinantes dos preços de imóveis residenciais verticais no município de são paulo. In: . [S.l.: s.n.], 2012. Citado 3 vezes nas páginas 19, 20 e 25.
- GUPTA. Study of machine learning models for real estate price predictio. 2019. Citado na página 26.
- HASTIE, T.; FRIEDMAN. In: *The Elements of Statistical Learning*. [S.l.]: Springer, 2009. Citado 4 vezes nas páginas 20, 21, 22 e 24.
- HAYKIN, S. Ann neuron. In: *Neural Networks and Learning Machines*. [S.l.: s.n.], 2009. Citado na página 24.
- IBGE. 2017. . Disponível em: <http://ibape-nacional.com.br/biblioteca/wp-content/uploads/2017/05/III_Sem_Nac_Eng_Av_4_6_MarceloRossi.pdf>. Citado na página 14.
- IBGE. 2019. *Número de domicílios brasileiros*. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37238-pais-tem-90-milhoes-de-domicilios-34-a-mais-que-em-2010#:~:text=Dos%20recenseados%20em%202022%2C%2090,e%20os%20coletivos%2C%20105%20mil.>> Citado na página 16.

- IBGE. *Conta-satélite do PIB do Brasil 2021*. 2022. <<https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/26568-Conta-satelite-do-PIB.html?=&t=o-que-e>>. Acesso em: 20 fev. 2022. Citado na página 18.
- KDNUGGETS. *Poll Results: Python Overtakes R, Becomes the Leader in Data Science, Machine Learning Platforms*. 2020. <<https://www.kdnuggets.com/2020/05/poll-results-python-overtakes-r-data-science-machine-learning-platforms.html>>. [Acessado em 18 fev. 2023]. Citado na página 28.
- KELLEHER, M. N. e. D. In: *Fundamentals of Machine Learning dor Predictive Data Analytics*. [S.l.: s.n.], 2015. p. 40–52. Citado 2 vezes nas páginas 20 e 21.
- KUHN, M.; JOHNSON, K. *Applied predictive modeling*. 2013. Citado na página 21.
- MULLER, C. 2017. Citado na página 27.
- MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. 1. ed. Cambridge: MIT Press, 2012. Disponível em: <<https://www.cs.ubc.ca/~murphyk/MLbook/>>. Citado na página 20.
- NANDA; KUMAR; BHARTI. Real estate price prediction using machine learning technique. 2019. Citado na página 26.
- NETO, A. P. Curso de engenharia de avaliação imobiliária. In: *Curso de Engenharia de Avaliação Imobiliária*. [S.l.: s.n.], 2003. Citado na página 20.
- NICULESCU-MIZIL, C. Floresta aleatórias. 2006. Citado na página 23.
- OLIVEIRA, D. d. P. R. d. Metodologia científica: um manual para a realização de pesquisas em administração. universidade federal de goiás. catalão-go. 2011. Citado na página 27.
- PANDIMURUGAN, V. Random forest tree classification algorithm for predicating loan. 2021. Citado 2 vezes nas páginas 22 e 23.
- PASSOS, R. C. M.; NOGAMI, O. *Princípios de economia*. 7. ed. [S.l.: s.n.], 2005. Citado na página 14.
- PINDYCK, R.; RUBINFELD, D. *Microeconomia*. 9. ed. São Paulo: Pearson, 2014. Citado na página 14.
- PNADC. 2020. *Pesquisa nacional de amostra de domicílios*. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/habitacao/9276-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?=&t=o-que-e>>. Citado 2 vezes nas páginas 16 e 17.
- SING, T.; YANG JESSE, Y. S. Decision tree and boosting techniques in artificial intelligence based automated valuation models (ai-avm). 2020. Citado na página 25.
- WANG, D.; LI, J. V. Revisão bibliográfica de 100 artigos. In: *Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 24 e 25.
- YU; WUNSCH. Survey of clustering algorithms. 2005. Citado na página 20.

ZHANG H, N. D.; ZHU, Z. Random forest. In: *Regression-Enhanced Random Forests*. [S.l.: s.n.], 2019. p. 2–4. Citado na página [22](#).

Apêndices

APÊNDICE A – Código utilizado para obter a base de dados São Paulo

```
1 import http.client
2 import json
3 import csv
4 import chardet
5 import os
6
7
8 bounds_south = -23.574688
9 bounds_north = -23.551657
10 bounds_east = -46.677270
11 bounds_west = -46.708336
12
13
14 conn = http.client.HTTPSConnection("www.quintoandar.com.br")
15 payload = json.dumps({
16     "filters": {
17         "map": {
18             "bounds_south": bounds_south,
19             "bounds_north": bounds_north,
20             "bounds_east": bounds_east,
21             "bounds_west": bounds_west,
22
23         },
24         "availability": "any",
25         "occupancy": "any",
26         "country_code": "BR",
27         "sorting": {
28             "criteria": "relevance_rent",
29             "order": "desc"
30         },
31         "page_size": 500,
32         "offset": 0
33     },
34     "return": [
35         "id",
36         "coverImage",
37         "rent",
38         "totalCost",
39         "salePrice",
40         "area",
```

```
41         "imageList",
42         "imageCaptionList",
43         "address",
44         "regionName",
45         "city",
46         "visitStatus",
47         "activeSpecialConditions",
48         "type",
49         "forRent",
50         "center",
51         "forSale",
52         "isPrimaryMarket",
53         "bedrooms",
54         "parkingSpaces",
55         "listingTags",
56         "yield",
57         "yieldStrategy",
58         "neighbourhood"
59     ],
60     "user_id": 8031568,
61     "device_id": "411a6cd6-f9c9-4f78-a7c0-4bd77f737975R"
62 })
63 headers = {
64     'Authorization': 'Bearer
65     eyJhbGciOiJSUzI1NiIsImtpZCI6ImY4NzZiNzIxNDAwYmZhZmEyOWQ0MTFmZTYwODE2YmRhZWMy
66     .
67     eyJpc3MiOiJodHRwczovL3NlY3VyZXRva2VuLmdvb2dsZS5jb20vcXVpbnRvYW5kYXItY29tLWJy
68     .
69     QAzug1oavx3IZFb2S90XjMm7nPgkXP_AB1hry8TB6iQ4c2hiQR2ArFr8Yuy7XGAYVO_2zH1VA2SI
70     -
71     XU_Akk2XuWa29dY_ZxR06lsJeIGKquP0pgdcWCbrs10kWKFYfBcDd8JVH4Pf1100XShyCS
72     -PL4kjQpk9xGnzHcyjtT01Yh6XyvFnOKHd09LsD_0XVivTKP -
73     ArH_k690m9QwxTcMe8qmdnwh05dhJmRXUCo3u--
74     WMcwgA55CkuRo7If3ho7HtsA4xt0X6L4zTtkyKzcMsfP0E_jeI9GX5ut3lVLRIPveeZ
75     -aP5olw',
76     'Content-Type': 'application/json'
77 }
78 conn.request("POST", "/api/yellow-pages/v2/search?relax_query=false",
79             payload, headers)
80 res = conn.getresponse()
81 data = res.read()
82 objectJson = json.loads(data.decode("utf-8").encode('latin-1').decode
83                        ("utf-8"))
84
85 #print(objectJson["hits"]["hits"])
86
87 json_list = objectJson["hits"]["hits"]
```

```
75
76 conn = http.client.HTTPSConnection("www.quintoandar.com.br")
77 payload = ''
78 headers = {
79     'Authorization': 'Bearer
    eyJhbGciOiJSUzI1NiIsImtpZCI6ImY4NzZiNzIxNDAwYmZhZmEyOWQ0MTFmZTYwODE2YmRhZWMy
    ',
80     'Cookie': 'X-AB-Test=ab_exposed_filter_sale_desktop:0,
    ab_exposed_filter_sale_v1:1,ab_exposed_filters_pf:0,
    ab_guided_search_neighborhood_nudge:1,ab_load_more_v0:-1,
    ab_new_house_card_v2:0,ab_new_rent_listing_phase2:1,
    ab_profiling_create_alert_filters:0,ab_retrieval_scope_fix:1,
    ab_scheduling_step_by_step_v1:0,ab_show_track_floor_sale:0,
    ab_street_search_retrieval_v1:1'
81 }
82 conn.request("GET", "/api/yellow-pages/v2/search?map%5Bbounds_north%5
    D=" + str(bounds_north) + "&map%5Bbounds_south%5D=" + str(
    bounds_south) + "&map%5Bbounds_east%5D=" + str(bounds_east) + "&
    map%5Bbounds_west%5D=" + str(bounds_west) + "&availability=any&
    occupancy=any&country_code=BR&business_context=RENT&return=
    location&photos=12&relax_query=false", payload, headers)
83 res = conn.getresponse()
84 data = res.read()
85 objectJsonLocation = json.loads(data.decode("utf-8"))["hits"]["hits"]
86
87 locations = {}
88
89 for i in objectJsonLocation:
90     locations[str(i["_id"])] = i["_source"]["location"]
91
92 csv_file = os.path.expanduser("~/Downloads/sp.csv")
93
94 # CSV header row
95 header = ["id", "coverImage", "rent", "totalCost", "salePrice", "area
    ", "imageList", "imageCaptionList", "address", "regionName", "city
    ", "visitStatus", "activeSpecialConditions", "type", "forRent", "
    forSale", "isPrimaryMarket", "bedrooms", "parkingSpaces", "
    neighbourhood", "listingTags", "lon", "lat"]
96
97
98 for i in range(len(json_list)):
99     print(json_list[i]['_source']['lon'])
100    print(json_list[i]['_source']['lat'])
101
102
103 # Open CSV file for writing and write header row
104 with open(csv_file, 'w', newline='', encoding='utf-8') as f:
```

```
105     writer = csv.DictWriter(f, fieldnames=header)
106     writer.writeheader()
107
108     # Write JSON elements as rows in CSV file
109     for json_obj in json_list:
110         writer.writerow(json_obj["_source"])
```

Listing A.1 – Código para importar dados de imóveis.

B Código utilizado para obter a base de dados de Brasília

```
1 import http.client
2 import json
3 import csv
4 import chardet
5 import os
6
7
8 bounds_south=-16.15
9 bounds_north=-15.45
10 bounds_east=-47.55
11 bounds_west=-48.60
12 center_lat=-15.8
13 center_lng=-48.42
14
15 conn = http.client.HTTPSConnection("www.quintoandar.com.br")
16 payload = json.dumps({
17     "filters": {
18         "map": {
19             "bounds_south": bounds_south,
20             "bounds_north": bounds_north,
21             "bounds_east": bounds_east,
22             "bounds_west": bounds_west,
23             "center_lat": center_lat,
24             "center_lng": center_lng
25         },
26         "availability": "any",
27         "occupancy": "any",
28         "country_code": "BR",
29         "sorting": {
30             "criteria": "relevance_rent",
31             "order": "desc"
32         },
33         "page_size": 500,
34         "offset": 0
35     },
36     "return": [
37         "id",
38         "coverImage",
39         "rent",
40         "totalCost",
41         "salePrice",
```



```
42     "area",
43     "imageList",
44     "imageCaptionList",
45     "address",
46     "regionName",
47     "city",
48     "visitStatus",
49     "activeSpecialConditions",
50     "type",
51     "forRent",
52     "center",
53     "forSale",
54     "isPrimaryMarket",
55     "bedrooms",
56     "parkingSpaces",
57     "listingTags",
58     "yield",
59     "yieldStrategy",
60     "neighbourhood"
61 ],
62 "user_id": 8031568,
63 "device_id": "411a6cd6-f9c9-4f78-a7c0-4bd77f737975R"
64 })
65 headers = {
66     'Authorization': 'Bearer
67     eyJhbGciOiJSUzI1NiIsImtpZCI6ImY4NzZiNzIxNDAwYmZhZmEyOWQ0MTFmZTYwODE2YmRhZWMy
68     .
69     eyJpc3MiOiJodHRwczovL3N1Y3VyZXRva2VuLmdvb2dsZS5jb20vcXVpbmRvYW5kYXItY29tLWJy
70     .
71     QAzug1oavx3IZFb2S90XjMm7nPgkXP_AB1hry8TB6iQ4c2hiQR2ArFr8Yuy7XGAYVO_2zH1VA2SI
72     -
73     XU_Akk2XuWa29dY_ZxR06lsJeIGKquP0pgdcWCbrs10kWKFYfBcDd8JVH4Pf1100XShyCS
74     -PL4kjQpk9xGnzHcyjtT01Yh6XyvFnOKHd09LsD_0XVivTKP -
75     ArH_k690m9QwxTcMe8qmdnwh05dhJmRXUCo3u --
76     WMcwgA55CkuRo7If3ho7HtsA4xt0X6L4zTtkyKzcMsfP0E_jeI9GX5ut3lVLRIPveeZ
77     -aP5olw',
78     'Content-Type': 'application/json'
79 }
80 conn.request("POST", "/api/yellow-pages/v2/search?relax_query=false",
81             payload, headers)
82 res = conn.getresponse()
83 data = res.read()
84 objectJson = json.loads(data.decode("utf-8").encode('latin-1').decode(
85     "utf-8"))
86
87 #print(objectJson["hits"]["hits"])
```

```
76 json_list = objectJson["hits"]["hits"]
77
78 conn = http.client.HTTPSConnection("www.quintoandar.com.br")
79 payload = ''
80 headers = {
81     'Authorization': 'Bearer
      eyJhbGciOiJIUzI1NiIsImtpZCI6ImY4NzZiNzIxNDAwYmZhZmEyOWQ0MTFmZTYwODE2YmRhZWMy
      ',
82     'Cookie': 'X-AB-Test=ab_exposed_filter_sale_desktop:0,
      ab_exposed_filter_sale_v1:1,ab_exposed_filters_pf:0,
      ab_guided_search_neighborhood_nudge:1,ab_load_more_v0:-1,
      ab_new_house_card_v2:0,ab_new_rent_listing_phase2:1,
      ab_profiling_create_alert_filters:0,ab_retrieval_scope_fix:1,
      ab_scheduling_step_by_step_v1:0,ab_show_track_floor_sale:0,
      ab_street_search_retrieval_v1:1'
83 }
84 conn.request("GET", "/api/yellow-pages/v2/search?map%5Bbounds_north%5
      D=" + str(bounds_north) + "&map%5Bbounds_south%5D=" + str(
      bounds_south) + "&map%5Bbounds_east%5D=" + str(bounds_east) + "&
      map%5Bbounds_west%5D=" + str(bounds_west) + "&availability=any&
      occupancy=any&country_code=BR&business_context=RENT&return=
      location&photos=12&relax_query=false", payload, headers)
85 res = conn.getresponse()
86 data = res.read()
87 objectJsonLocation = json.loads(data.decode("utf-8"))["hits"]["hits"]
88
89 locations = {}
90
91 for i in objectJsonLocation:
92     locations[str(i["_id"])] = i["_source"]["location"]
93
94 csv_file = os.path.expanduser("~/Downloads/bsb.csv")
95
96 # CSV header row
97 header = ["id", "coverImage", "rent", "totalCost", "salePrice", "area
      ", "imageList", "imageCaptionList", "address", "regionName", "city
      ", "visitStatus", "activeSpecialConditions", "type", "forRent", "
      forSale", "isPrimaryMarket", "bedrooms", "parkingSpaces", "
      neighbourhood", "listingTags", "lon", "lat"]
98
99
100 for i in range(len(json_list)):
101     print(json_list[i]['_source']["lon"])
102     print(json_list[i]['_source']["lat"])
103
104
105 # Open CSV file for writing and write header row
```

```
106 with open(csv_file, 'w', newline='', encoding='utf-8') as f:
107     writer = csv.DictWriter(f, fieldnames=header)
108     writer.writeheader()
109
110     # Write JSON elements as rows in CSV file
111     for json_obj in json_list:
112         writer.writerow(json_obj["_source"])
```

Listing B.1 – Código para importar dados de imóveis.

C Código utilizado para aprendizado de máquina

```
1
2
3     import pandas as pd
4     from sklearn.linear_model import LinearRegression
5     import matplotlib.pyplot as plt
6     import numpy as np
7     from sklearn.metrics import mean_squared_error
8     from sklearn.pipeline import Pipeline
9     from sklearn.preprocessing import StandardScaler
10    from sklearn.model_selection import train_test_split
11    import seaborn as sns
12    import folium
13    from folium.plugins import MarkerCluster
14
15
16    # Carregar o arquivo CSV
17    data = pd.read_csv('sp.csv')
18
19    # Separar os dados em características (X) e alvo (y)
20    X = data[['lat', 'lon', 'bedrooms', 'parkingSpaces', 'area', '
21             totalCost']]
22    y = data['rent']
23
24    # Criar o modelo de regressão linear
25    model = LinearRegression()
26
27    # Treinar o modelo com todos os dados
28    model.fit(X, y)
29
30    # Fazer previsões para cada linha do CSV e adicionar a coluna "
31    # valor_justo_aluguel"
32    data['valor_justo_aluguel'] = model.predict(X)
33
34    # Salvar o arquivo CSV atualizado com a nova coluna
35    data.to_csv('sp_com_valor_justo_aluguel.csv', index=False)
36
37    data['variação'] = (data['valor_justo_aluguel'] / data['rent'] -1)
38    *100
```

```
38
39 #feature_names = X.columns
40 #Plotar o gráfico com os pontos menores em cinza
41 plt.scatter(data['rent'], data['valor_justo_aluguel'], color='gray',
42             alpha=0.25)
43 #Plotar a reta da regressão linear em preto
44 plt.plot([data['rent'].min(), data['rent'].max()], [data['
45             valor_justo_aluguel'].min(), data['valor_justo_aluguel'].max()],
46           color='red')
47 plt.xlabel('Aluguel (R$)')
48 plt.ylabel('Valor Justo Aluguel (R$)')
49 #plt.title('Reta da Regressão Linear: Aluguel vs. Valor Justo Aluguel
50             ')
51 plt.ticklabel_format(style='plain', axis='x')
52 plt.gca().get_xaxis().set_major_formatter(plt.FuncFormatter(lambda x,
53             p: format(int(x), ',').replace(',',' ','.')))
54 plt.ticklabel_format(style='plain', axis='y')
55 plt.gca().get_yaxis().set_major_formatter(plt.FuncFormatter(lambda y,
56             p: format(int(y), ',').replace(',',' ','.')))
57 plt.show()
```

Listing C.1 – Código utilizado para aprendizado de máquina.