

# oBoticário

## Análise de Dados e Predição com Python e Pandas

### Escopo do Documento:

- Motivação
- Base de dados
- Tratamento da base de dados
- Análise Descritiva
- Visualização dos Dados
- Modelagem de Regressão Linear
- Conclusões
- Tecnologias utilizadas



## Sumário

Histórico do Documento.....	3
Motivação.....	4
Base de dados.....	5
Tratamento da base de dados.....	6
Análise Descritiva.....	7
Visualização dos Dados.....	10
Modelagem de Regressão Linear.....	15
Conclusões.....	16
Tecnologias utilizadas.....	17



## Histórico do Documento

Data	Autor	Nome do Arquivo	Detalhe
31/05/2024	Pedro Lima do Nascimento	Projeto2_Boticario.pdf	Criação do Documento.

## Motivação

Realizar uma análise completa usando Python e a biblioteca Pandas numa base de dados de preços de aluguéis em São Paulo. Utilizar regressão linear para modelar uma relação entre variáveis do conjunto de dados.

O desafio completo está presente nesse [link](#).

## Base de dados

Foi fornecida uma [base de dados](#) relativos a endereços e dados específicos relativos a aluguéis de imóveis na cidade de São Paulo. Essa base, em formato 'csv', possui 11658 registros, e 6 colunas:

Coluna	Definição
address	Logradouro
district	Bairro
area	Metragem do imóvel
bedrooms	Quantidade de quartos
garage	Quantidade de garagens ou vagas na garagem
type	Tipo do apartamento: Apartamento, Studio/kitnet, Casa em condomínio e Casa
rent	Preço do aluguel
total	Preço do aluguel + despesas

*Tabela 1: Colunas da base de dados*

## Tratamento da base de dados

Após a obtenção da base de dados através da biblioteca Pandas do Python, a mesma foi submetida a uma série de tratamentos de dados. Foram eles:

- 1) Através da API Geocoding do Google, foi adicionado a coluna CEP à base de dados, passando como parâmetro os campos 'address' e 'district'.
- 2) Foram excluídos da base de dados os imóveis com área menor que 10 m<sup>2</sup>. Essa reportagem, onde uma construtora alega ter feito o menor apartamento da América Latina, foi usada como base para esse parâmetro. Como resultado, 7 registros foram excluídos.
- 3) Algumas buscas da API não encontraram CEPs ativos com base nos endereços apresentados. Para esses casos, o campo CEP foi atualizado com o CEP de outro endereço no mesmo bairro.
- 4) Após esse tratamento, 5 registros continuaram sem CEP e foram excluídos da base de dados.
- 5) Tanto os registros excluídos no passo 2 quanto no passo 4 foram salvos nessa [base de dados](#) para posterior análise.
- 6) Com base nas faixas de CEP de São Paulo obtidas neste [site](#), foi criada uma nova coluna na base de dados, chamada 'Zona', destinada a agrupar cada registros em regiões diferentes: Central, Leste, Oeste, Norte e Grande SP.

Todos os tratamentos citados acima estão disponíveis neste [link](#).

## Análise Descritiva

Com o uso do Pandas, é possível realizar inúmeras análises com a base de dados. Para este trabalho, vamos analisar algumas:

### 1) 50 bairros com mais imóveis:

	bairro	qtd_imoveis
0	Bela Vista	352
1	Vila Mariana	232
2	Jardim Paulista	220
3	Centro	178
4	Pinheiros	159
5	Consolação	157
6	Santana	152
7	Mooca	147
8	Santo Amaro	146
9	Brás	142
10	Liberdade	141
11	Cerqueira César	140
12	Vila Andrade	129
13	República	124
14	Campos Elíseos	117
15	Vila Suzana	114
16	Perdizes	111
17	Campo Belo	108
18	Cambuci	104
19	Tatuapé	98
20	Vila Prudente	96
21	Vila Buarque	90
22	Paraíso	89
23	Vila Olímpia	88
24	Barra Funda	83
25	Vila Regente Feijó	82
26	Vila Clementino	80
27	Indianópolis	79
28	Vila Nova Conceição	77
29	Ipiranga	76
30	Centro Histórico de São Paulo	73
31	Vila Formosa	64
32	Sé	63
33	Jabaquara	55
34	Vila Santa Catarina	53
35	Vila da Saúde	52
36	Itaim Bibi	51
37	Aclimação	50
38	Taboão	49
39	Cidade Monções	49
40	Planalto Paulista	47
41	Santa Cecília	47
42	Brooklin Paulista	46
43	Butantã	45
44	Vila Bertioga	44
45	Água Branca	44
46	Parque da Mooca	43
47	Santa Efigênia	43
48	Vila Mascote	42
49	Penha de França	40

Tabela 2: Lista dos 50 bairros com mais imóveis.

## 2) Mínimo, média e máximo de valores (aluguel e total):

	rent	total	bairro
min	R\$ 500.00	R\$ 509.00	Vila Nivi
mean	R\$ 3,251.68	R\$ 4,081.20	None
max	R\$ 25,000.00	R\$ 28,700.00	Planalto Paulista

Tabela 3: Maiores e menores valores, seus bairros e médias gerais.

## 3) Mínimo, média e máximo de área por tipo de imóvel:

	min	mean	max
type			
Apartamento	10	73.332221	568
Casa	10	136.372134	580
Casa em condomínio	16	120.138075	560
Studio e kitnet	10	31.764493	130

Tabela 3: Mínimo, média e máximo de m<sup>2</sup> por tipo de imóvel.

## 4) Contagem de imóveis por tipo de imóvel:

Apartamento	7191
Casa	2835
Studio e kitnet	1380
Casa em condomínio	239

Tabela 4: Quantidade de imóveis por tipo de imóvel.



## 5) Média, mínimo e máximo por região:

Zona	rent			total		
	min	mean	max	min	mean	max
Central	R\$ 650.00	R\$ 3,225.54	R\$ 15,000.00	R\$ 782.00	R\$ 4,196.52	R\$ 22,990.00
Grande SP	R\$ 500.00	R\$ 2,957.03	R\$ 15,000.00	R\$ 556.00	R\$ 3,576.45	R\$ 28,700.00
Leste	R\$ 500.00	R\$ 2,465.02	R\$ 15,000.00	R\$ 623.00	R\$ 2,983.37	R\$ 19,760.00
Norte	R\$ 500.00	R\$ 2,376.60	R\$ 15,000.00	R\$ 509.00	R\$ 2,863.00	R\$ 19,160.00
Oeste	R\$ 600.00	R\$ 3,987.09	R\$ 15,000.00	R\$ 625.00	R\$ 5,082.89	R\$ 23,180.00
Sul	R\$ 500.00	R\$ 3,943.37	R\$ 25,000.00	R\$ 521.00	R\$ 4,983.77	R\$ 26,710.00

Tabela 5: Mínimo, média e máximo por região(aluguel e total)

## 6) Preço por metro quadrado de cada região:

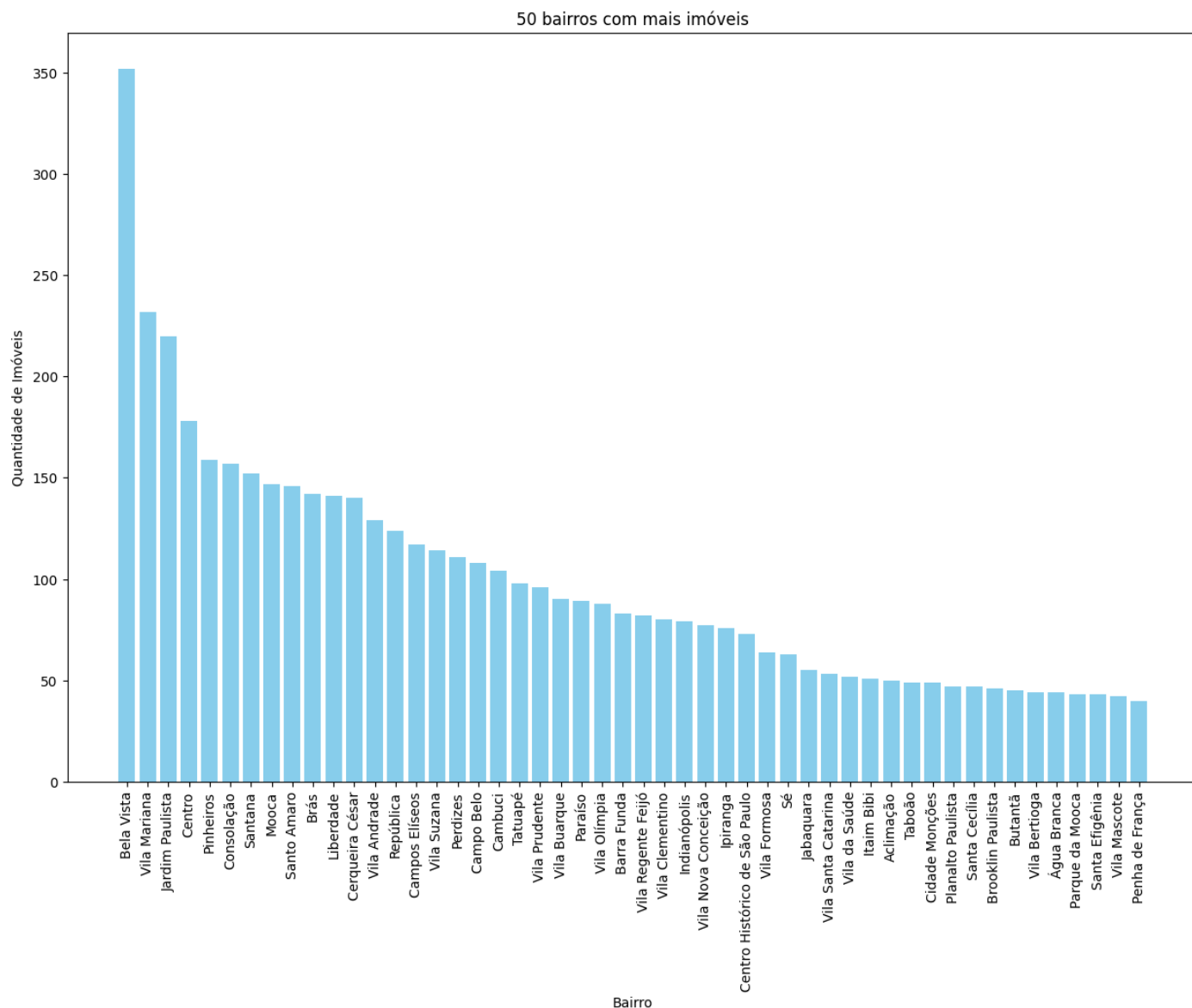
Preço por Metro Quadrado por Zona:		
Zona	rent_per_sqm	total_per_sqm
Central	R\$ 56.45	R\$ 72.86
Sul	R\$ 51.75	R\$ 65.52
Oeste	R\$ 48.46	R\$ 61.32
Leste	R\$ 39.10	R\$ 47.03
Grande SP	R\$ 35.06	R\$ 42.94
Norte	R\$ 31.95	R\$ 38.81

Tabela 6: Preço por m<sup>2</sup> quadrado de cada região, aluguel e total.

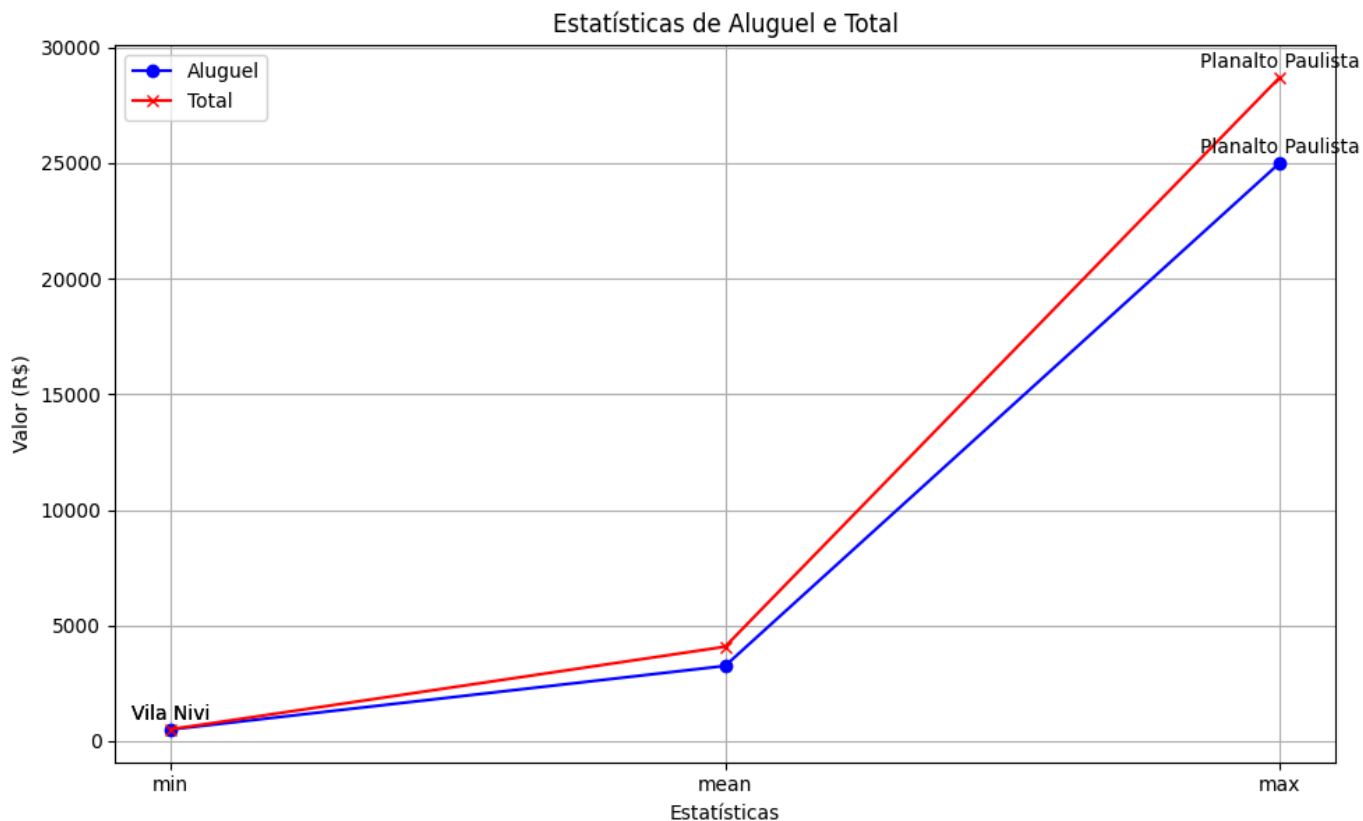
## Visualização dos Dados

Com o uso do Matplotlib foi possível a plotagem de gráficos relativos às tabelas e dados coletados e citados em capítulos anteriores. Vamos a eles:

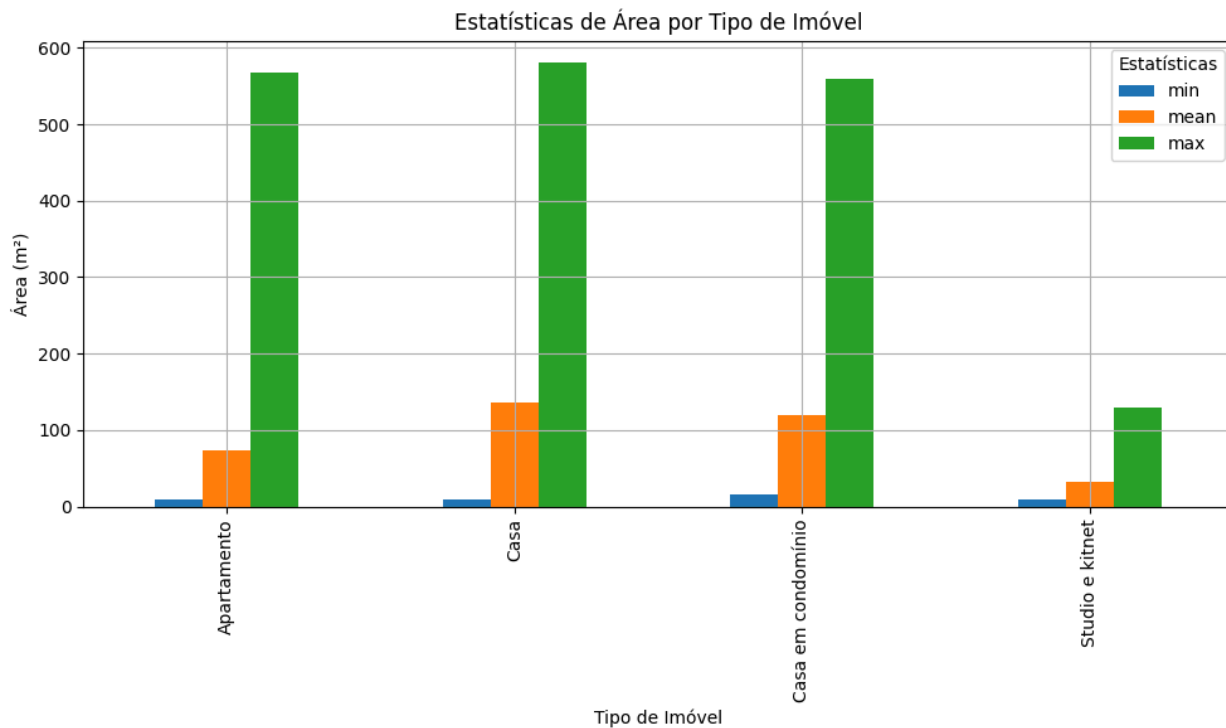
### 1. 50 bairros com mais imóveis:



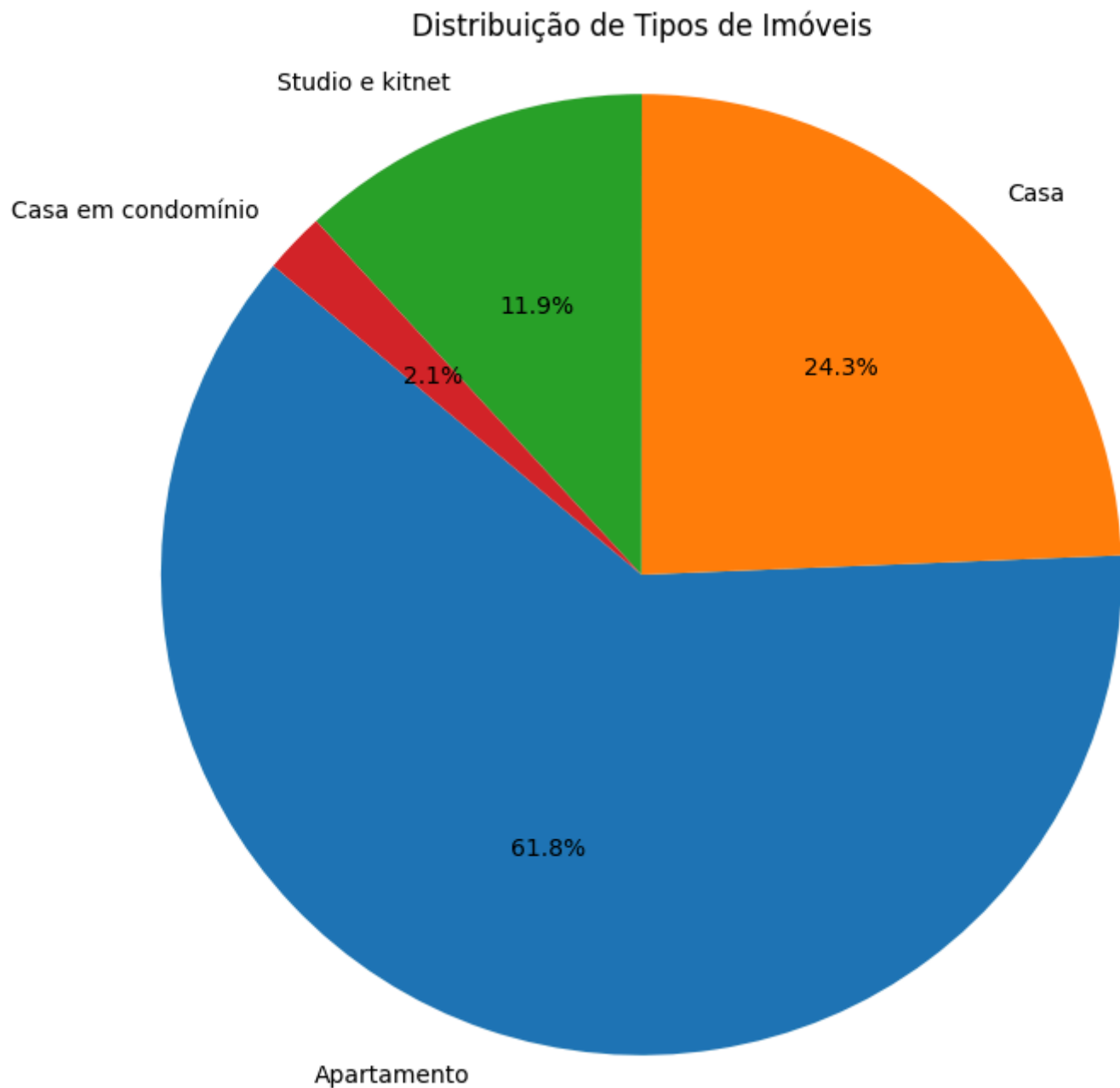
## 2. Mínimo, média e máximo de valores (aluguel e total):



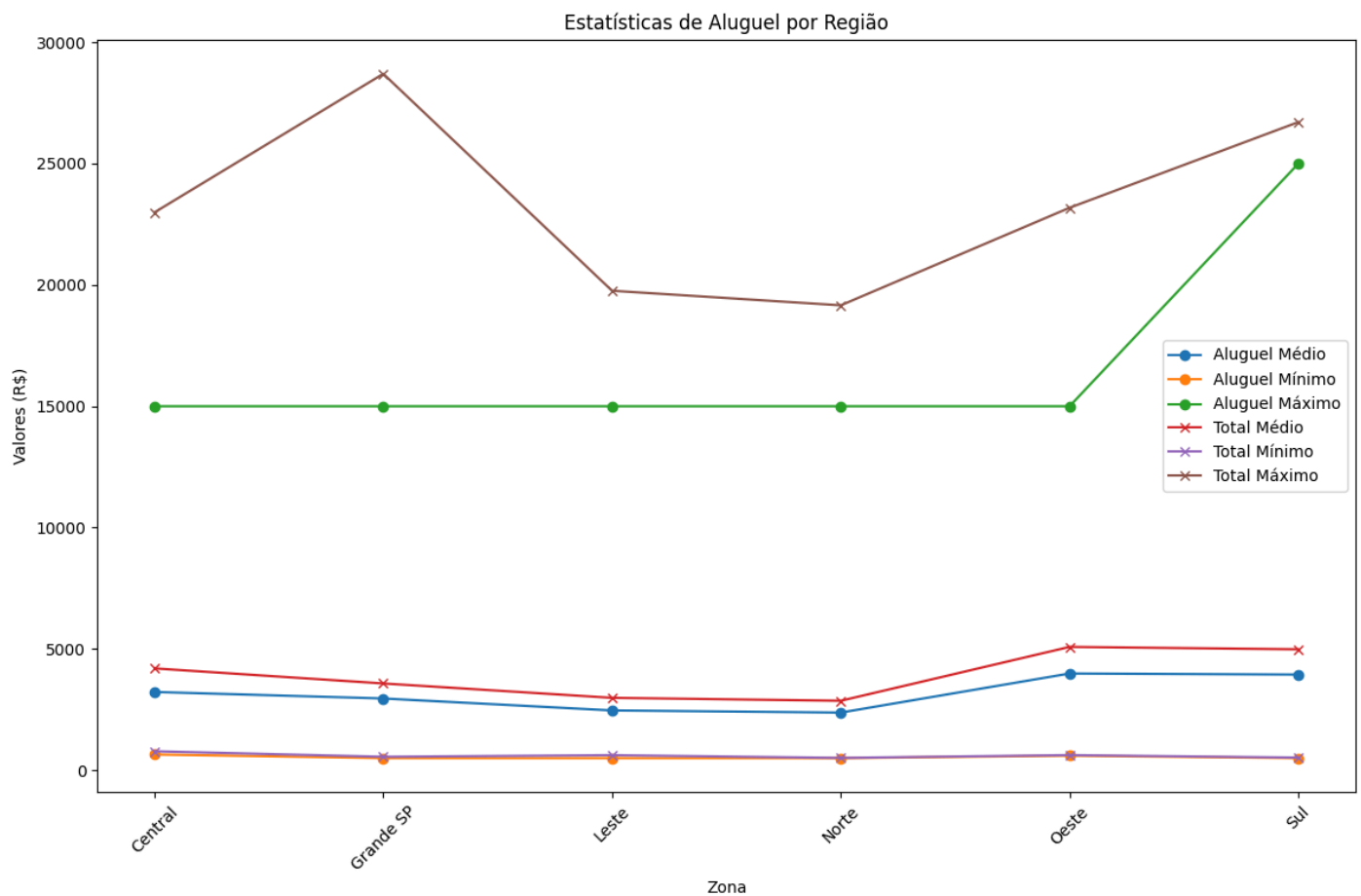
## 3. Mínimo, média e máximo de área por tipo de imóvel:



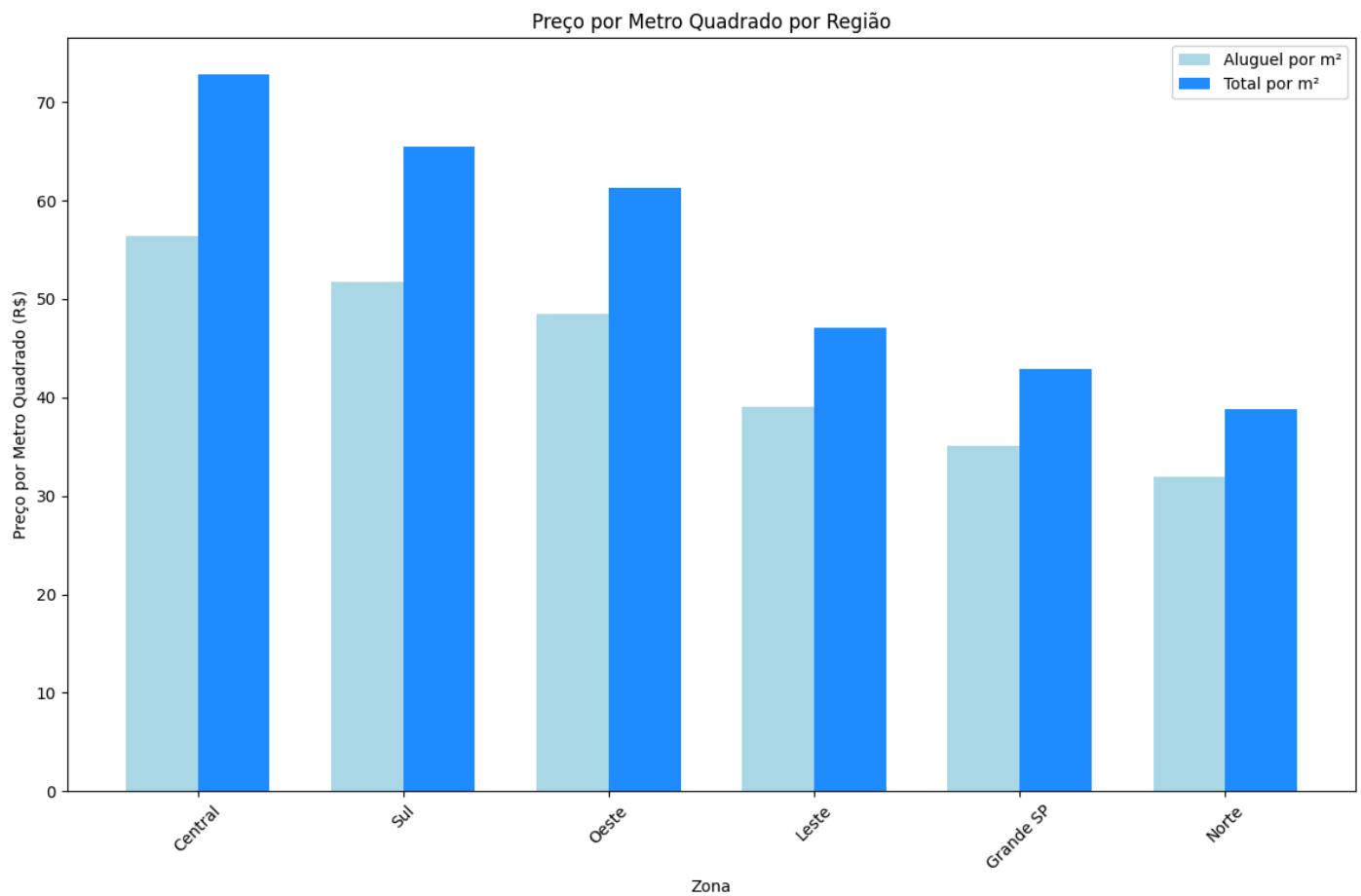
## 4. Contagem de imóveis por tipo de imóvel



## 5. Média, mínimo e máximo por região:



## 6. Preço por metro quadrado de cada região:



O código relativo a criação dos gráficos e também das tabelas do capítulo anterior está disponível [aqui](#).

## Modelagem de Regressão Linear

Foi elaborado um modelo de regressão linear para prever o preço do aluguel na Zona Sul de São Paulo. A região foi escolhida por possuir 2.885 registros na base de dados tratada, correspondendo a 24,77% do total de imóveis, sendo, portanto, a região com a maior quantidade de dados.

Para a aplicação do modelo, foram considerados dois dados: a área (m<sup>2</sup>) e o preço do aluguel (R\$). O modelo resultou em um coeficiente de correlação de aproximadamente 0,64, indicando uma [correlação moderada](#).

Com base na aplicação desse modelo, chegamos às seguintes projeções:

Área (m <sup>2</sup> )	Projeção de Aluguel (R\$)
10	R\$ 1.931,87
30	R\$ 2.407,33
50	R\$ 2.882,80
70	R\$ 3.358,26
90	R\$ 3.833,73
110	R\$ 4.309,19
130	R\$ 4.784,66
150	R\$ 5.260,13

*Tabela 7: Projeções do preço do aluguel*

A análise completa do modelo de regressão linear está disponível neste [link](#).

## Conclusões

- Os bairros com maior quantidade de imóveis, como Bela Vista, Vila Mariana e Jardim Paulista, são conhecidos por serem áreas centrais e valorizadas. Isso pode indicar uma alta demanda por imóveis nessas regiões, tanto para aluguel quanto para compra. Juntos, representam cerca de 6,9% dos imóveis da base de dados.
- A diferença significativa entre o valor mínimo (R\$ 500,00) e o valor máximo (R\$ 25.000,00) de aluguel destaca a ampla gama de preços de imóveis na cidade. Isso reflete uma grande diversidade no mercado imobiliário, atendendo diferentes níveis socioeconômicos.
- Existe uma ampla variedade de tamanhos de imóveis disponíveis, desde unidades muito pequenas (10 m<sup>2</sup>) até imóveis muito grandes (580 m<sup>2</sup>). Isso reflete a diversidade do mercado, atendendo a diferentes necessidades e preferências dos moradores.
- A predominância de apartamentos sugere que muitas pessoas preferem a conveniência e a praticidade associadas a esse tipo de imóvel, apesar da popularização recente dos kitnets. A urbanização e a tendência de morar perto do local de trabalho e de serviços podem ser fatores importantes.
- As Zonas Central, Oeste e Sul apresentam valores médios e máximos mais altos, indicando que são áreas mais valorizadas e procuradas, possivelmente devido à proximidade de centros comerciais, empresariais e melhores infraestruturas. Enquanto as Zonas Leste e Norte têm os menores valores médios, sugerindo que são áreas mais acessíveis para moradores com menor renda.
- Há uma diferença clara nos preços por metro quadrado entre as zonas, com a Zona Central sendo a mais cara e a Zona Norte a mais acessível. Essas diferenças refletem a valorização das áreas em termos de infraestrutura, proximidade de serviços e centros comerciais.
- A análise cruzada dos dados revela um mercado imobiliário altamente diversificado e segmentado em São Paulo, com claras diferenças regionais e de tipo de imóvel. Essas informações são valiosas para investidores, planejadores urbanos, desenvolvedores e consumidores, permitindo decisões mais estratégicas e informadas no mercado imobiliário.



## Tecnologias utilizadas

Neste projeto algumas tecnologias foram utilizadas. Seja pela elaboração do tratamento dos dados, elaboração dos gráficos, busca de dados relativos aos CEP 's, versionamento e documentação.

Python 3.12

Pandas

Numpy

Seaborn

Statsmodels

Scipy

Matplotlib

Google Colab

API Geocodin

Google Docs

GitHub