

Modelos Estimadores de Preços de Imóveis em São José dos Campos, SP

Antônio Magno Lima Espescht
CC226 – Introdução à Análise de Padrões

18 de junho de 2008

Resumo

Utilizou-se regressão linear múltipla para obter equações que descrevem o mercado imobiliário de São José dos Campos, a partir dos dados obtidos do site www.redesaojose.com.br, nos dias 09 de abril de 2008 e 04 de maio de 2008, georeferenciados com coordenadas obtidas do site maps.google.com.

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 2 | Objetivo | 1 |
| 3 | Banco de Dados | 1 |
| 3.1 | Obtenção dos Dados Imobiliários | 2 |
| 3.2 | Preparação dos Dados Imobiliários | 3 |
| 3.3 | Georeferenciamento | 5 |
| 3.4 | Estrutura do Banco de Dados | 9 |
| 4 | Análise Exploratória dos Dados | 11 |
| 4.1 | Mapas Georeferenciados | 13 |
| 4.2 | Vendedores | 19 |
| 4.3 | Área | 23 |
| 4.3.1 | Área dos 164 Apartamentos à Venda | 23 |
| 4.3.2 | Área dos 25 Apartamentos para Locação | 25 |
| 4.3.3 | Área das 197 Casas à Venda | 26 |
| 4.3.4 | Área das 24 Casas para Locação | 26 |
| 4.3.5 | Área dos 21 Terrenos à Venda | 30 |
| 4.4 | Preço em Reais | 32 |
| 4.4.1 | Preço em Reais dos 164 Apartamentos à Venda | 36 |
| 4.4.2 | Preço em Reais dos 25 Apartamentos para Locação | 40 |
| 4.4.3 | Preço em Reais das 197 Casas à Venda | 44 |
| 4.4.4 | Preço em Reais das 24 Casas para Locação | 48 |
| 4.4.5 | Preço em Reais dos 21 Terrenos à Venda | 52 |
| 5 | Regressão Múltipla | 56 |
| 5.1 | Modelos de Estimativa de Preço para os 164 Apartamentos à Venda | 56 |
| 5.2 | Modelos de Estimativa de Preço para os 25 Apartamentos para Locação | 58 |
| 5.3 | Modelos de Estimativa de Preço para as 197 Casas à Venda | 59 |
| 5.4 | Modelos de Estimativa de Preço para as 24 Casas para Locação | 60 |
| 5.5 | Modelos de Estimativa de Preço para os 21 Terrenos à Venda | 60 |
| 6 | Conclusão | 61 |
| 7 | Referências Bibliográficas | 62 |

| | | |
|-----|------------------|----|
| 7.1 | Sites | 62 |
| 7.2 | Papers | 62 |

Lista de Figuras

| | | |
|----|--|----|
| 1 | Vista parcial da interface de Pesquisa de Imóveis do site www.redesaojose.com.br . | 3 |
| 2 | Página para georeferenciamento dos dados. | 5 |
| 3 | Página para georeferenciamento dos dados exibindo <i>links</i> para alguns bairros. | 6 |
| 4 | Página para georeferenciamento dos dados logo após o usuário ter clicado no link URBANOVA. Nota-se que, além do posicionamento de um balão informativo sobre o bairro, a interface também exibe as coordenadas $-23.196148, -45.935812$ no topo da página. | 6 |
| 5 | Cada marca neste mapa corresponde a um bairro do Banco de Dados. | 7 |
| 6 | Visão dos bairros mais próximos do centro da cidade. | 7 |
| 7 | Coordenadas do bairro JARDIM ORIENTE (ponto central) e coordenadas dos 5 registros do Banco de Dados (pontos em volta do ponto central) que pertencem ao bairro JARDIM ORIENTE. | 8 |
| 8 | Bairros de São José dos Campos nos dados da Análise Exploratória. | 14 |
| 9 | Posição no mapa dos 431 registros usados na Análise Exploratória. | 14 |
| 10 | Zoom sobre o mapa da Figura 9 onde se pode notar a disposição em forma circular dos registros de um mesmo bairro. | 15 |
| 11 | Posicionamento dos 189 apartamentos analisados. | 15 |
| 12 | Posicionamento dos 164 apartamentos para venda. | 16 |
| 13 | Posicionamento dos 25 apartamentos para locação. | 16 |
| 14 | Posicionamento das 221 casas analisadas. | 17 |
| 15 | Posicionamento das 197 casas à venda. | 17 |
| 16 | Posicionamento das 24 casas para locação. | 18 |
| 17 | Posicionamento dos 21 terrenos à venda. | 18 |
| 18 | Vendedores de Todos os 431 Imóveis. | 20 |
| 19 | Vendedores dos 164 Apartamentos à Venda. | 20 |
| 20 | Vendedores dos 25 Apartamentos para Locação. | 21 |
| 21 | Vendedores das 197 Casas à Venda. | 21 |
| 22 | Vendedores das 24 Casas para Locação. | 22 |
| 23 | Vendedores dos 21 Terrenos à Venda. | 22 |
| 24 | Histograma da Área de Todos os 416 Imóveis com Área menor ou igual a $500m^2$. | 23 |
| 25 | Histograma da Área de Todos os 15 Imóveis com Área maior que $500m^2$. | 24 |
| 26 | Área dos 431 Imóveis. | 24 |
| 27 | Histograma da Área dos 164 Apartamentos à Venda. | 25 |
| 28 | Área dos 164 Apartamentos à Venda. | 25 |
| 29 | Histograma da Área dos 25 Apartamentos para Locação. | 26 |

| | | |
|----|--|----|
| 30 | Área dos 25 Apartamentos para Locação. | 27 |
| 31 | Histograma da Área das 197 Casas à Venda. | 27 |
| 32 | Área das 197 Casas à Venda. | 28 |
| 33 | Histograma da Área das 24 Casas para Locação. | 28 |
| 34 | Área das 24 Casas para Locação. | 29 |
| 35 | Histograma da Área dos 21 Terrenos à Venda. | 30 |
| 36 | Área dos 21 Terrenos à Venda. | 31 |
| 37 | Histograma do Preço em Reais de Todos os 431 Imóveis. | 32 |
| 38 | Preço dos 431 Imóveis. | 33 |
| 39 | Histograma do Logaritmo Decimal do Preço em Reais de Todos os 431 Imóveis. | 33 |
| 40 | Logaritmo Decimal do Preço dos 431 Imóveis. | 34 |
| 41 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 431 imóveis. | 34 |
| 42 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 431 imóveis, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 35 |
| 43 | Histograma do Preço em Reais dos 164 Apartamentos à Venda. | 36 |
| 44 | Preço dos 164 Apartamentos à Venda. | 37 |
| 45 | Histograma do Logaritmo Decimal do Preço em Reais dos 164 Apartamentos à Venda. | 37 |
| 46 | Logaritmo Decimal do Preço dos 164 Apartamentos à Venda. | 38 |
| 47 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 164 Apartamentos à Venda. | 38 |
| 48 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 164 Apartamentos à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 39 |
| 49 | Histograma do Preço em Reais dos 25 Apartamentos para Locação. | 40 |
| 50 | Preço dos 25 Apartamentos para Locação. | 41 |
| 51 | Histograma do Logaritmo Decimal do Preço em Reais dos 25 Apartamentos para Locação. | 41 |
| 52 | Logaritmo Decimal do Preço dos 25 Apartamentos para Locação. | 42 |
| 53 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 25 Apartamentos para Locação. | 42 |
| 54 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 25 Apartamentos para Locação, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 43 |
| 55 | Histograma do Preço em Reais das 197 Casas à Venda. | 44 |

| | | |
|----|---|----|
| 56 | Preço das 197 Casas à Venda. | 45 |
| 57 | Histograma do Logaritmo Decimal do Preço em Reais das 197 Casas à Venda. | 45 |
| 58 | Logaritmo Decimal do Preço das 197 Casas à Venda. | 46 |
| 59 | Valor do Imóvel em função da Latitude e da Longitude, para todas as 197 Casas à Venda. | 46 |
| 60 | Valor do Imóvel em função da Latitude e da Longitude, para todas as 197 Casas à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 47 |
| 61 | Histograma do Preço em Reais das 24 Casas para Locação. | 48 |
| 62 | Preço das 24 Casas para Locação. | 49 |
| 63 | Histograma do Logaritmo Decimal do Preço em Reais das 24 Casas para Locação. | 49 |
| 64 | Logaritmo Decimal do Preço das 24 Casas para Locação. | 50 |
| 65 | Valor do Imóvel em função da Latitude e da Longitude, para todas as 24 Casas para Locação. | 50 |
| 66 | Valor do Imóvel em função da Latitude e da Longitude, para todas as 24 Casas para Locação, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 51 |
| 67 | Histograma do Preço em Reais dos 21 Terrenos à Venda. | 52 |
| 68 | Preço dos 21 Terrenos à Venda. | 53 |
| 69 | Histograma do Logaritmo Decimal do Preço em Reais dos 21 Terrenos à Venda. | 53 |
| 70 | Logaritmo Decimal do Preço dos 21 Terrenos à Venda. | 54 |
| 71 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 21 Terrenos à Venda. | 54 |
| 72 | Valor do Imóvel em função da Latitude e da Longitude, para todos os 21 Terrenos à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos. | 55 |
| 73 | Resíduos calculados para o Modelo 3. Estão destacados em vermelho os <i>outliers</i> | 57 |
| 74 | Resíduos calculados para o Modelo 3 com 156 Apartamentos à Venda. Estão destacados em vermelho os <i>outliers</i> | 58 |

Lista de Tabelas

| | | |
|----|---|----|
| 1 | Estimativa do número de Andares em função do campo Sub-Tipo | 4 |
| 2 | Obtenção do Campo Tipo Imóvel em função do campo Sub-Tipo | 4 |
| 3 | Segmentação do Banco de Dados em relação ao tipo de negócio imobiliário. . | 11 |
| 4 | Registros Utilizados na Análise Exploratória. | 12 |
| 5 | Modelos de Estimativa de Preço para os 164 Apartamentos à Venda | 56 |
| 6 | Estatísticas para os 164 Apartamentos à Venda | 57 |
| 7 | Modelos de Estimativa de Preço para 156 Apartamentos à Venda | 58 |
| 8 | Modelos de Estimativa de Preço para os 25 Apartamentos para Locação . . . | 59 |
| 9 | Modelos de Estimativa de Preço para os 23 Apartamentos para Locação . . . | 59 |
| 10 | Modelos de Estimativa de Preço para as 197 Casas à Venda | 59 |
| 11 | Modelos de Estimativa de Preço para as 190 Casas à Venda | 60 |
| 12 | Modelos de Estimativa de Preço para as 24 Casas para Locação | 60 |
| 13 | Modelos de Estimativa de Preço para 23 Casas para Locação | 60 |
| 14 | Modelos de Estimativa de Preço para os 21 Terrenos à Venda | 61 |
| 15 | Modelos de Estimativa de Preço para 19 Terrenos à Venda | 61 |
| 16 | Melhores Equações para Cada Modelo | 61 |

1 Introdução

O preço de um imóvel é definido por diversos fatores tais como: localização, área, idade, condições do mercado, infra-estrutura disponível, criminalidade na vizinhança etc.

Alguns destes fatores estão correlacionados de uma forma muito intrincada ou de difícil mensuração. É o caso, por exemplo, de imóveis que tenham valor histórico ou que tenham sido criados por arquitetos de renome.

Há fatores que seriam mensuráveis mas para os quais não há dados disponíveis. Presume-se, por exemplo, que regiões de trânsito congestionado tendem a ter seus imóveis desvalorizados. Porém, nem sempre há dados de trânsito disponíveis que permitam analisar numericamente este efeito.

Deve-se considerar também, a variação destes parâmetros no tempo, sendo comum uma área nobre tornar-se muito populosa para em seguida sofrer a consequente queda nos preços dos imóveis devido à falência da infra-estrutura da região.

Há ainda as considerações que variam de bairro para bairro. Numa área recém construída, espera-se imóveis homogêneos quanto à idade, o que melhora a correlação entre preço e área, por exemplo. Analogamente, em bairros mais antigos, a correlação entre área e preço pode ser bem pior pois a diferença entre idade (e estado de conservação) dos imóveis passa a influir mais no preço.

Abstraindo estas dificuldades, nossa proposta é tentar obter equações de preço a partir de dados que podem ser facilmente obtidos de sites na internet tais como: área, bairro em que o imóvel está localizado, número de quartos, quantidade de vagas na garagem etc.

Para tornar o estudo mais interessante, associamos coordenadas geográficas (longitude e latitude) aos bairros e, pela adição de um ruído estatístico, transferimos estas coordenadas para os imóveis de cada bairro.

A partir dos dados disponíveis na internet, criamos um banco de dados com o qual, por regressão linear, definimos equações matemáticas que descrevem o preço de imóveis em São José dos Campos.

Na análise exploratória, definimos diversos gráficos que descrevem a variabilidade dos dados coletados e, com a validação cruzada, escolhemos os melhores modelos analisados.

2 Objetivo

Definir equações para preço de imóveis de São José dos Campos-SP, utilizando regressão linear com os parâmetros disponíveis no site www.redesaojose.com.br.

3 Banco de Dados

Os dados utilizados nas análises foram coletados no site www.redesaojose.com.br em 09 de abril e em 04 de maio de 2008.

Após serem transferidos para uma planilha Excel, estes dados foram tratados para remover registros espúrios e criar novos campos.

Em seguida, os dados foram georeferenciados a coordenadas obtidas do site `maps.google.com` resultando no Banco de Dados final.

3.1 Obtenção dos Dados Imobiliários

Os dados imobiliários foram coletados a partir da interface de Pesquisa de Imóveis (Figura 1) do site `www.redesaojose.com.br` obtendo-se dados da cidade de São José dos Campos que atendem aos critérios:

- i) Apartamentos à venda.
- ii) Apartamentos para locação.
- iii) Casas à venda.
- iv) Casa para locação.
- v) Terrenos à venda.
- vi) Imóveis comerciais à venda.
- vii) Imóveis comerciais para locação.

Removeu-se a formatação HTML dos dados coletados, consolidando os registros em uma planilha Excel com os campos:

- i) **Referência:** identifica univocamente cada registro. Ex.: PI-00708.
- ii) **Tipo:** preenchido com “Venda” ou “Locação” conforme o tipo de negócio imobiliário considerado.
- iii) **Sub-Tipo:** preenchido com um dentre os seguintes valores, de acordo com o tipo de imóvel a ser vendido ou alugado: Apartamento, Área, Casa, Casa Comercial, Casa Condomínio, Cobertura, Flat, Fundo de Comércio, Galpão / Depósito, Prédio Inteiro, Sala Comercial, Salão Comercial, Sobrado, Sobrado Condomínio, Terreno ou Terreno Condomínio.
- iv) **Bairro:** preenchido com o nome do bairro onde o imóvel está localizado. Foram listados 127 diferentes bairros.
- v) **Área:** preenchido com a área do imóvel em metros quadrados.
- vi) **Dormitórios:** número de dormitórios do imóvel (exceto suíte).
- vii) **Suítes:** número de suítes (dormitórios com banheiro anexo) do imóvel.
- viii) **Vagas:** número de vagas na garagem.
- ix) **Preço:** preço de venda ou locação mensal do imóvel, em reais.

PESQUISA DE IMÓVEIS

| | |
|---|--|
| Finalidade: | <input checked="" type="radio"/> Venda <input type="radio"/> Locação |
| Cidade: | Sao Jose Dos Campos |
| Tipo do Imóvel: | Apartamento |
| Sub-Tipo: | Indiferente |
| Bairro: | Indiferente |
| Dormitórios: | Indiferente |
| Vagas de Garagem: | Indiferente |
| Valor Mínimo: | Indiferente |
| Valor Máximo: | Indiferente |
| <input type="checkbox"/> Mostrar somente imóveis com foto | |
| Código de referência do imóvel: <input type="text"/> | |
| <input type="button" value="Buscar"/> | |

| IMÓVEIS A VENDA - Total: 167 | | | | | | | | | |
|------------------------------|-------------|---------------|-----|---|---|---|-----------|--|--------------------------|
| Ref. | Sub-Tipo | Bairro | | | | | Preço R\$ | | |
| NF-01042 | Apartamento | CENTRO | - | 1 | - | 1 | 45.000,00 | | conferir |
| PI-00695 | Apartamento | JD. AQUARIUS | 102 | 3 | - | 2 | 48.000,00 | | conferir |
| IN-00250 | Apartamento | JD. PORTUGAL | 46 | 2 | - | 1 | 54.000,00 | | conferir |
| AX-02021 | Apartamento | MONTE CASTELO | 54 | 2 | - | 1 | 55.000,00 | | conferir |
| AX-02003 | Apartamento | V. INDUSTRIAL | 42 | 2 | - | - | 56.000,00 | | conferir |

Figura 1: Vista parcial da interface de Pesquisa de Imóveis do site www.redesaojose.com.br.

3.2 Preparação dos Dados Imobiliários

Foram coletados dados em duas datas: 09 de abril e 04 de maio de 2008.

Considerou-se que os dados que estavam no site na primeira coleta e não estavam na segunda foram vendidos. Portanto, foi criado o campo **Vendido**, preenchido com 1 para os dados existentes em 09 de abril e inexistentes em 04 de maio de 2008 e com 0 para os demais registros.

Foram removidos da base os registros com áreas e/ou preços iguais a zero. Em alguns casos, foi possível obter os valores da área e/ou do preço consultando o site a partir da referência do imóvel. Porém, houve 35 casos em que os registros foram removidos.

Restaram 535 registros.

Foi criado o campo **ID**, preenchido com números de 1 a 535, que permite identificar univocamente cada registro a partir de um valor numérico ao invés do campo **Referência** que é alfanumérico.

Foi criado o campo **Vendedor** associando-se números de 1 a 12 para cada imóvel de acordo com as duas primeiras letras do código de referência. Por exemplo: os imóveis de referência 'AL-00049' e 'AL-00062' foram associados ao número 1 enquanto que os imóveis 'AX-00038'

e ‘AX -01507’ foram associados ao número 2. Cada número corresponde a uma empresa vendedora dentre o conjunto de empresas que anunciam no site www.redesaojose.com.br.

Foi criado o campo **VendaLoca** associando-se 1 para venda e 2 para locação do imóvel. Deste modo, o campo alfanumérico **Tipo** que existia na base original foi removido.

Foi criado o campo **Comercial** preenchido com 1 para os imóveis comerciais (isto é, que pertencem aos sub-tipos: Área, Casa Comercial, Fundo de Comércio, Galpão / Depósito, Prédio Inteiro, Sala Comercial e Salão Comercial) e com 0 para os demais sub-tipos.

Foi criado o campo **Condomínio** preenchido com 1 para os imóveis em que o acesso dos visitantes é controlado por portaria (isto é, que pertencem aos sub-tipos: Apartamento, Casa Condomínio, Cobertura, Sobrado Condomínio e Terreno Condomínio) e com 0 para os demais sub-tipos.

Foi criado o campo **Andares** preenchido com o número de andares do imóvel o qual foi estimado a partir do campo **Sub-Tipo** (Tabela 1).

Tabela 1: Estimativa do número de **Andares** em função do campo **Sub-Tipo**

| Andares | Sub-Tipo |
|----------------|--|
| 0 | Área, Terreno, Terreno Condomínio |
| 1 | Apartamento, Casa, Casa Comercial, Casa Condomínio, Cobertura, Flat, Fundo de Comércio, Galpão / Depósito, Sala Comercial, Salão Comercial |
| 2 | Sobrado, Sobrado Condomínio |
| 10 | Prédio Inteiro |

Foi criado o campo **TipoImovel** que reclassifica os registros de acordo com o campo **Sub-Tipo** (Tabela 2).

Tabela 2: Obtenção do Campo **Tipo Imóvel** em função do campo **Sub-Tipo**

| Tipo Imóvel | Sub-Tipo |
|--------------------|--|
| 1 | Apartamento, Cobertura |
| 2 | Área, Terreno, Terreno Condomínio |
| 3 | Casa, Casa Comercial, Casa Condomínio, Flat, Sobrado, Sobrado Condomínio |
| 4 | Fundo de Comércio |
| 5 | Galpão / Depósito |
| 6 | Prédio Inteiro |
| 7 | Sala Comercial |

Foi criado o campo **LogPreço** com o logaritmo decimal do valor do preço. Este campo é útil para tornar mais linear a variação exponencial dos preços.

3.3 Georeferenciamento

Utilizando uma API do Google Maps e a lista dos bairros, foi criado a página web (Figura 2): <http://www.ita90.com.br/CC226/map/map2.htm> na qual pode-se clicar sobre um mapa de São José dos Campos e obter as coordenadas (longitude e latitude) referentes ao ponto clicado.

Nesta página também é possível procurar por um endereço e visualizar pontos encontrados via Google Maps.

Uma vez encontradas as coordenadas do bairro o mesmo foi iterativamente listado no código fonte da página acima, na forma de um link (Figura 3).

Publicando novamente a página no site e clicando-se no link, o ponto referente ao bairro (Figura 4) é mostrado no mapa o que permite verificar se a coordenada está corretamente associada ao respectivo bairro.

Após muitas iterações e com o auxílio do Google Earth e de uma lista telefônica, foram obtidas as coordenadas de todos os bairros listados nos 535 registros da base (Figura 5), conforme pode ser visto no link: <http://www.ita90.com.br/CC226/map/map.htm>.

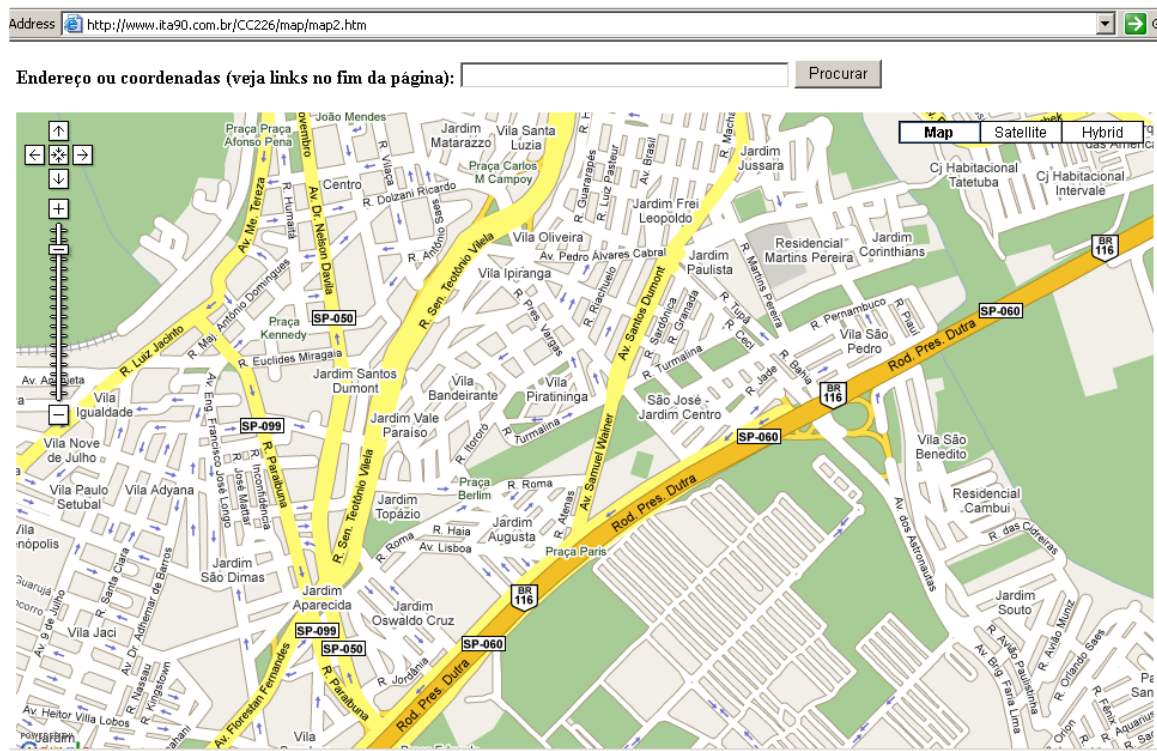
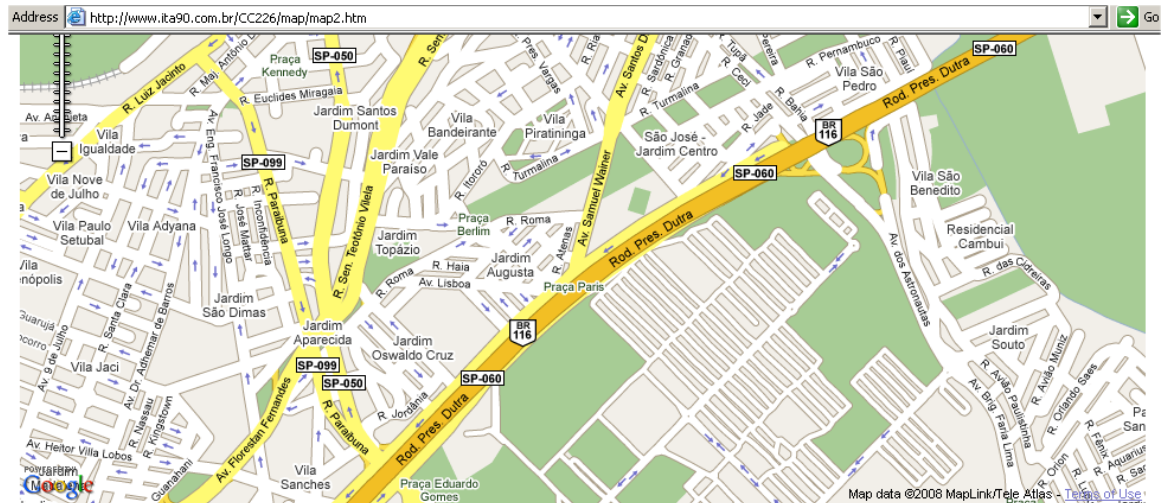


Figura 2: Página para georeferenciamento dos dados.



Endereços de Exemplo:

[31 DE MARÇO](#)

[ÁGUA SOCA](#)

[ALTO DA PONTE](#)

[ALTOS DA SERRA I](#)

[BOM RETIRO](#)

[BOM SUCESSO](#)

[BOSQUE DOS EUCALIPTOS](#)

[BUQUIRINHA I](#)

[BUQUIRINHA II](#)

[CAMPO DE SÃO JOSÉ](#)

Figura 3: Página para georeferenciamento dos dados exibindo *links* para alguns bairros.

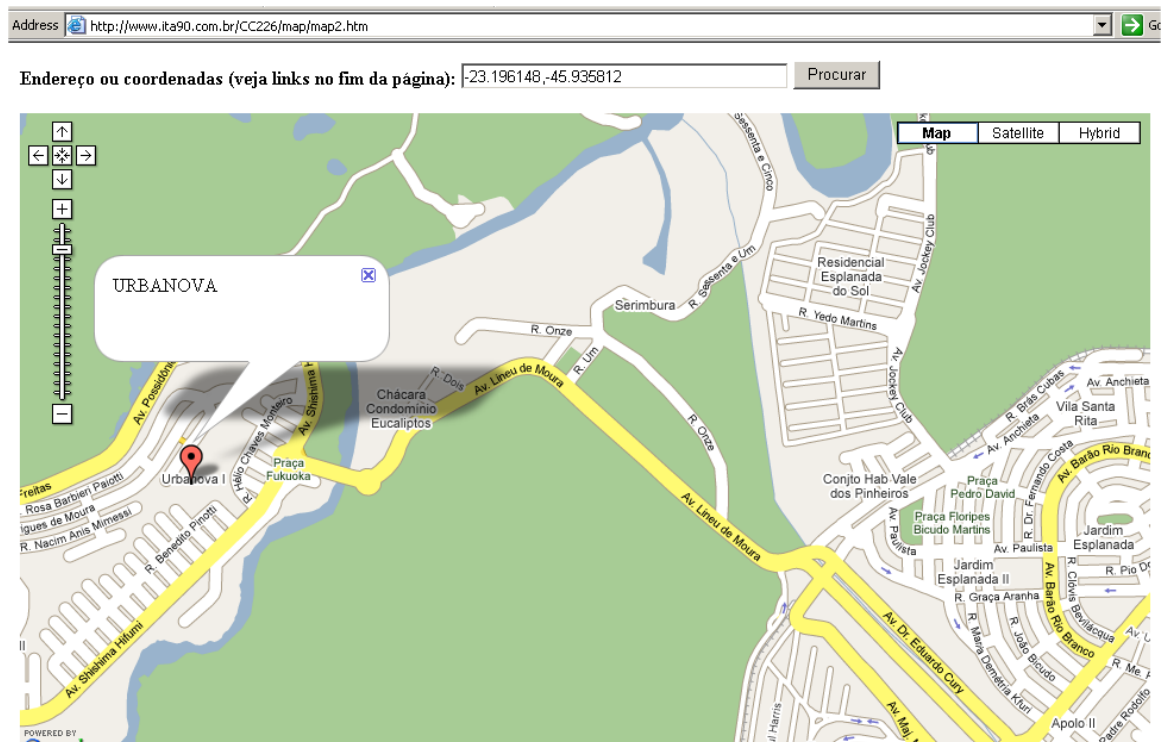


Figura 4: Página para georeferenciamento dos dados logo após o usuário ter clicado no link URBANOVA. Nota-se que, além do posicionamento de um balão informativo sobre o bairro, a interface também exibe as coordenadas $-23.196148, -45.935812$ no topo da página.

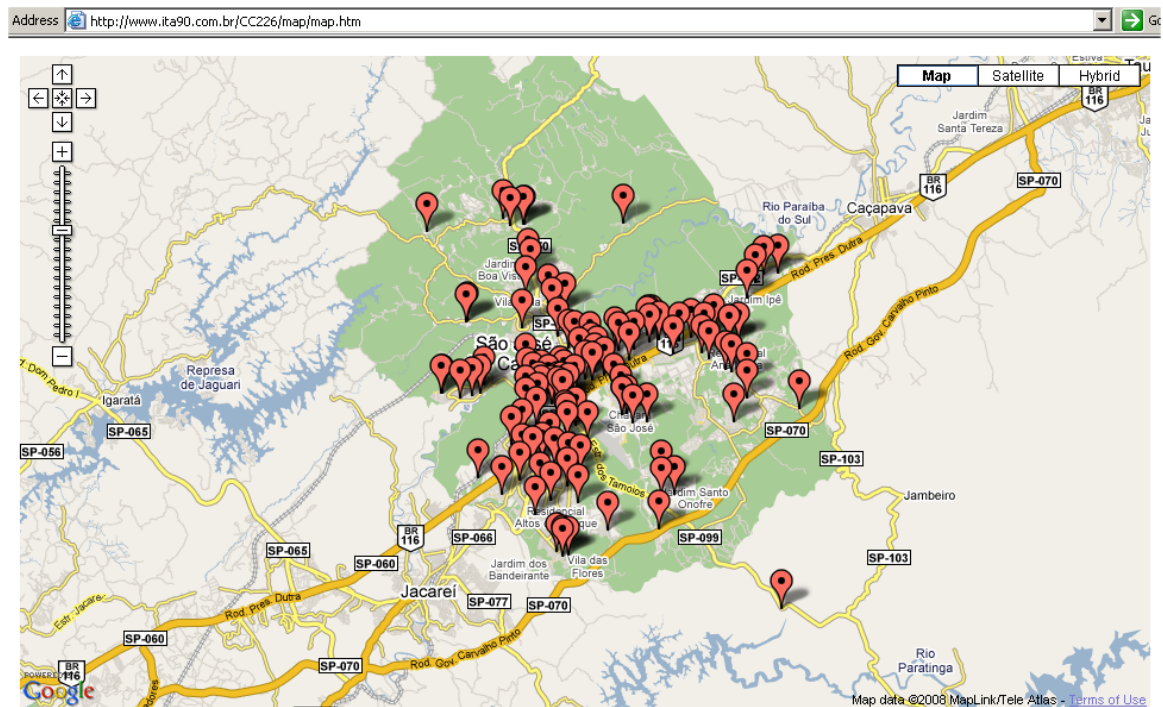


Figura 5: Cada marca neste mapa corresponde a um bairro do Banco de Dados.

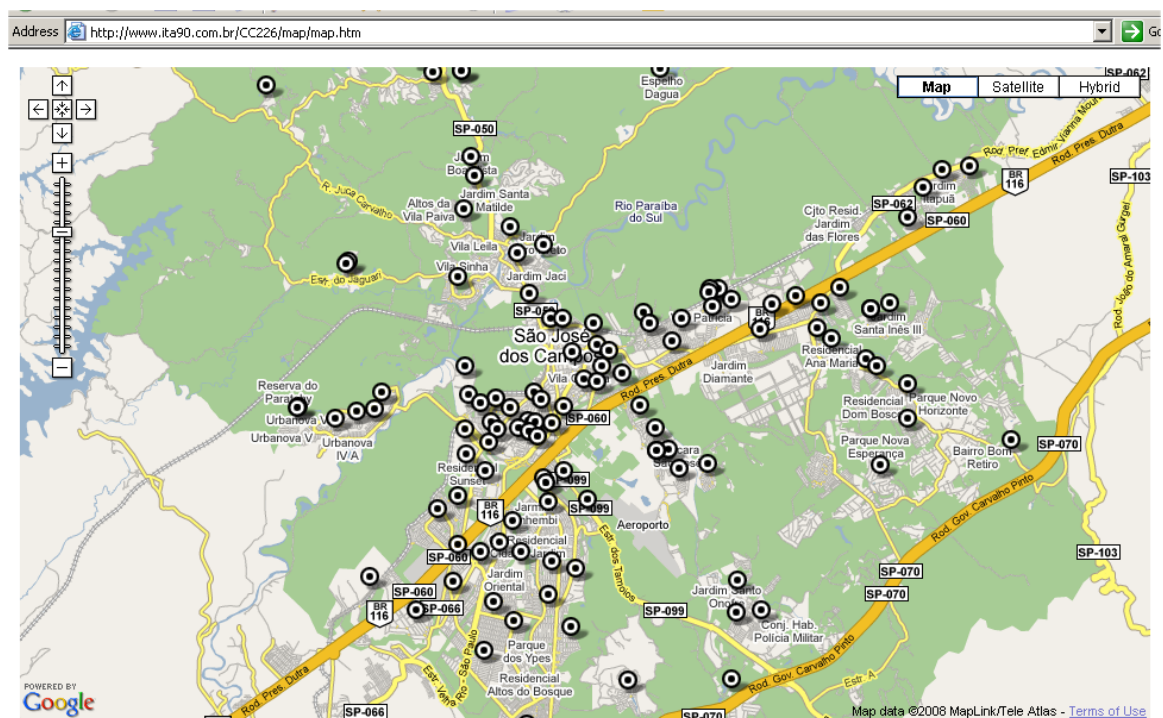


Figura 6: Visão dos bairros mais próximos do centro da cidade.

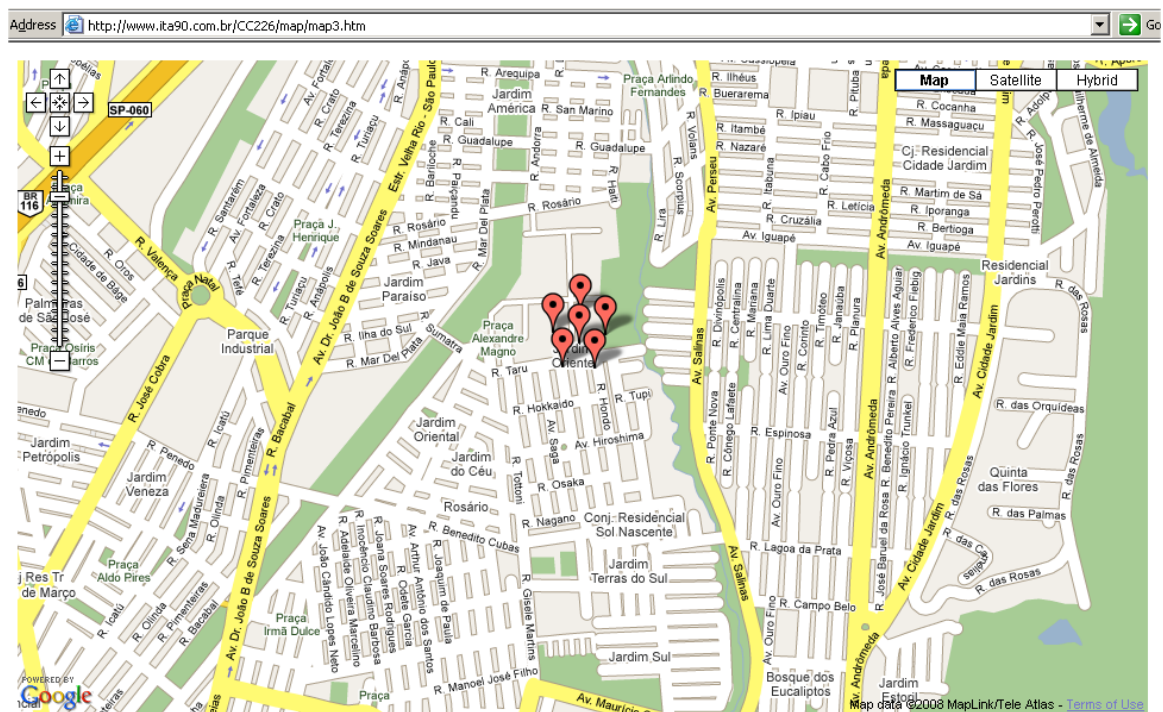


Figura 7: Coordenadas do bairro JARDIM ORIENTE (ponto central) e coordenadas dos 5 registros do Banco de Dados (pontos em volta do ponto central) que pertencem ao bairro JARDIM ORIENTE.

No Banco de Dados, foram criados os campos **Longitude** e **Latitude** para armazenar as coordenadas de cada registro.

Foi criada uma macro em Excel para atribuir coordenadas a cada registro a partir da coordenada do bairro ao qual o registro pertence. Esta macro define um círculo de raio igual a 0.001° cujo centro está na coordenada do bairro e em cujo perímetro se distribuem, igualmente espaçados, os registros referentes àquele bairro. Quando o bairro possui apenas um registro este é situado no centro do círculo.

Na Figura 7, por exemplo, temos os 5 registros do Jardim Oriente situados em torno do ponto central que é a coordenada do bairro em nosso banco de dados.

Dete modo, evita-se que todos os registros de um determinado bairro fiquem associados à coordenada daquele bairro, o que afetaria a forma como estes dados são visualizados nos mapas e como influenciam as equações de regressão.

A macro adiciona um “ruído” às coordenadas do bairro obtendo as coordenadas do registro. Os valores de ruído para longitude e latitude, respectivamente, estão nos campos **RuidoLon** e **RuidoLat**. Nota-se que, quando se soma os valores destes campos para um dado bairro, obtém-se zero.

As coordenadas de cada ponto, após adição do ruído, estão nos campos **LongitudeR** e **LatitudeR**.

3.4 Estrutura do Banco de Dados

Após obter, preparar e georeferenciar os dados obtidos nos sites www.redesaojose.com.br e maps.google.com foi criado um banco de dados com 535 registros e os seguintes campos:

- i) **ID:** identifica numérica e univocamente cada registro. O banco foi ordenado pelo campo **Referência** e, em seguida, cada registro recebeu um número na sequência de 1 a 535.
- ii) **Referência:** identifica alfanumérica e univocamente cada registro. Ex.: PI-00708.
- iii) **Vendedor:** identifica numericamente o vendedor do imóvel de acordo com as duas primeiras letras da **Referência**. Cada vendedor é associado a um número de 1 a 12.
- iv) **VendaLoca:** preenchido com 1 para “Venda” e 2 para “Locação” conforme o tipo de negócio imobiliário considerado.
- v) **Vendido:** preenchido com 1 se o imóvel estava no site em 09 de abril de 2008 e não estava no site em 04 de maio de 2008. Vale 0, caso contrário. Quando é igual a 1, indica que o imóvel foi removido do site, provavelmente por ter sido vendido.
- vi) **Comercial:** preenchido com 0 se o imóvel se destina a uso residencial ou 1 se o imóvel se destina à instalação de algum tipo de comércio.
- vii) **Condomínio:** preenchido com 1 se o imóvel se localiza em região cujo acesso requer que o visitante passe por uma portaria ou 0 caso contrário.

- viii) **Andares:** número estimado de andares do imóvel. É preenchido com 0 para terrenos, 1 para casas térreas e apartamentos, 2 para sobrados e 10 para prédios inteiros.
- ix) **TipoImovel:** separa os dados em 7 grandes grupos dentre os quais se destacam: 1=Apartamentos, 2=Terrenos e 3=Casas. Os demais tipos resultam em um número muito pequeno de registros para que se possa fazer uma análise estatisticamente significativa.
- x) **Sub-Tipo:** preenchido com um dentre os seguintes valores, de acordo com o imóvel a ser vendido ou alugado: Apartamento, Área, Casa, Casa Comercial, Casa Condomínio, Cobertura, Flat, Fundo de Comércio, Galpão / Depósito, Prédio Inteiro, Sala Comercial, Salão Comercial, Sobrado, Sobrado Condomínio, Terreno ou Terreno Condomínio.
- xi) **Bairro:** preenchido com o nome do bairro onde o imóvel está localizado. Foram listados 127 diferentes bairros.
- xii) **Área:** preenchido com a área do imóvel, em metros quadrados.
- xiii) **Dormitórios:** número de dormitórios do imóvel (exceto suíte).
- xiv) **Suítes:** número de suítes (dormitórios com banheiro anexo) do imóvel.
- xv) **Vagas:** número de vagas na garagem.
- xvi) **Preço:** preço de venda ou locação mensal do imóvel, em reais.
- xvii) **Log Preço:** logaritmo decimal do preço de venda ou locação mensal do imóvel, em reais.
- xviii) **Longitude:** Longitude do bairro do imóvel.
- xix) **Latitude:** Latitude do bairro do imóvel.
- xx) **RuidoLon:** Ruído adicionado à longitude do bairro do imóvel de modo a estimar a longitude do imóvel.
- xxi) **RuidoLat:** Ruído adicionado à latitude do bairro do imóvel de modo a estimar a latitude do imóvel.
- xxii) **LongitudeR:** Estimativa da longitude do imóvel, adicionando-se **RuidoLon** a **Longitude**.
- xxiii) **LatitudeR:** Estimativa da longitude do imóvel, adicionando-se **RuidoLat** a **Latitude**.

Tabela 3: Segmentação do Banco de Dados em relação ao tipo de negócio imobiliário.

| Negócio Imobiliário | VendaLoca | Tipo Imóvel | Registros |
|-------------------------|-----------|-------------|-----------|
| Venda de apartamentos | 1 | 1 | 197 |
| Venda de terrenos | 1 | 2 | 27 |
| Venda de casas | 1 | 3 | 229 |
| Outras vendas | 1 | 4, 5, 6, 7 | 16 |
| TOTAL de vendas | - | - | 469 |
| Aluguel de apartamentos | 2 | 1 | 25 |
| Aluguel de casas | 2 | 3 | 25 |
| Outros aluguéis | 2 | 5, 6, 7 | 16 |
| TOTAL de aluguéis | - | - | 66 |
| TOTAL GERAL | - | - | 535 |

4 Análise Exploratória dos Dados

Os registros do Banco de Dados foram segmentados por tipo de negócio imobiliário conforme mostrado na Tabela 3.

Pela análise desta tabela, nota-se que há poucos imóveis comerciais (apenas 32 registros subdivididos em 7 categorias). Portanto, optou-se por analisar apenas estes negócios imobiliários:

- i) Apartamentos à venda (**Tipo Imóvel** = 1 e **VendaLoca** = 1).
- ii) Apartamentos para locação (**Tipo Imóvel** = 1 e **VendaLoca** = 2).
- iii) Terrenos à venda (**Tipo Imóvel** = 2 e **VendaLoca** = 1).
- iv) Casas à venda (**Tipo Imóvel** = 3 e **VendaLoca** = 1).
- v) Casa para locação (**Tipo Imóvel** = 3 e **VendaLoca** = 2).

Na análise exploratória foram considerados apenas os registros existentes no site www.redesaojose.com.br no dia 04 de maio de 2008.

Restaram portanto 431 registros segmentados conforme a Tabela 4.

Tabela 4: Registros Utilizados na Análise Exploratória.

| Negócio Imobiliário | VendaLoca | Tipo Imóvel | Registros |
|----------------------------|------------------|--------------------|------------------|
| Venda de apartamentos | 1 | 1 | 164 |
| Venda de terrenos | 1 | 2 | 21 |
| Venda de casas | 1 | 3 | 197 |
| Outras vendas | 1 | 4, 5, 6, 7 | 0 |
| TOTAL de vendas | - | - | 382 |
| Aluguel de apartamentos | 2 | 1 | 25 |
| Aluguel de casas | 2 | 3 | 24 |
| Outros aluguéis | 2 | 5, 6, 7 | 0 |
| TOTAL de aluguel | - | - | 49 |
| TOTAL GERAL | - | - | 431 |

4.1 Mapas Georeferenciados

Os 431 registros analisados cobrem os principais bairros de São José dos Campos.

Na Figura 8, cada um dos 127 bairros é representado por um ponto. Na Figura 9, cada ponto representa um dos 431 registros analisados. Conforme esperado, as figuras são bastante semelhantes, mas na Figura 9 nota-se uma mancha escura nos bairros onde há maior número de registros (Urbanova, por exemplo). Esta mancha é causada pela disposição circular dos dados em torno de cada bairro, o que é ainda mais visível na Figura 10.

Dos 431 imóveis analisados:

- i) 189 são apartamentos (Figura 11). Destes, há 164 apartamentos à venda (Figura 12) e apenas 25 apartamentos para locação (Figura 13).
- ii) 221 são casas (Figura 14). Destas, há 197 casas à venda (Figura 15) e apenas 24 casas para locação (Figura 16).
- iii) 21 são terrenos (Figura 17).

Nota-se que há bem menos imóveis para locação do que para venda.

Pela análise dos mapas georeferenciados, nota-se que:

- i) Os apartamentos para venda (Figura 12) estão localizados em áreas próximas ao centro e paralelas à Dutra.
- ii) Os apartamentos para locação (Figura 13) estão localizados ao norte da Via Dutra.
- iii) Há casas à venda (Figura 15) em praticamente todos os bairros de São José dos Campos, principalmente nos mais afastados do centro e ao norte da Via Dutra.
- iv) De modo análogo ao que acontece com os apartamentos, as casas para locação (Figura 16) estão em menor número que as casas à venda e situadas, em sua maioria, ao norte da Via Dutra.
- v) Há terrenos à venda (Figura 17) apenas nas regiões mais afastadas do centro, em especial nas áreas de expansão da cidade como Urbanova (região oeste) e Espelho D'Água (ao norte).
- vi) Há grandes *vazios* sem nenhuma oferta de imóveis o que pode ser explicado pela existência do CTA, da Petrobrás, do Banhado e de outras áreas de preservação ambiental.

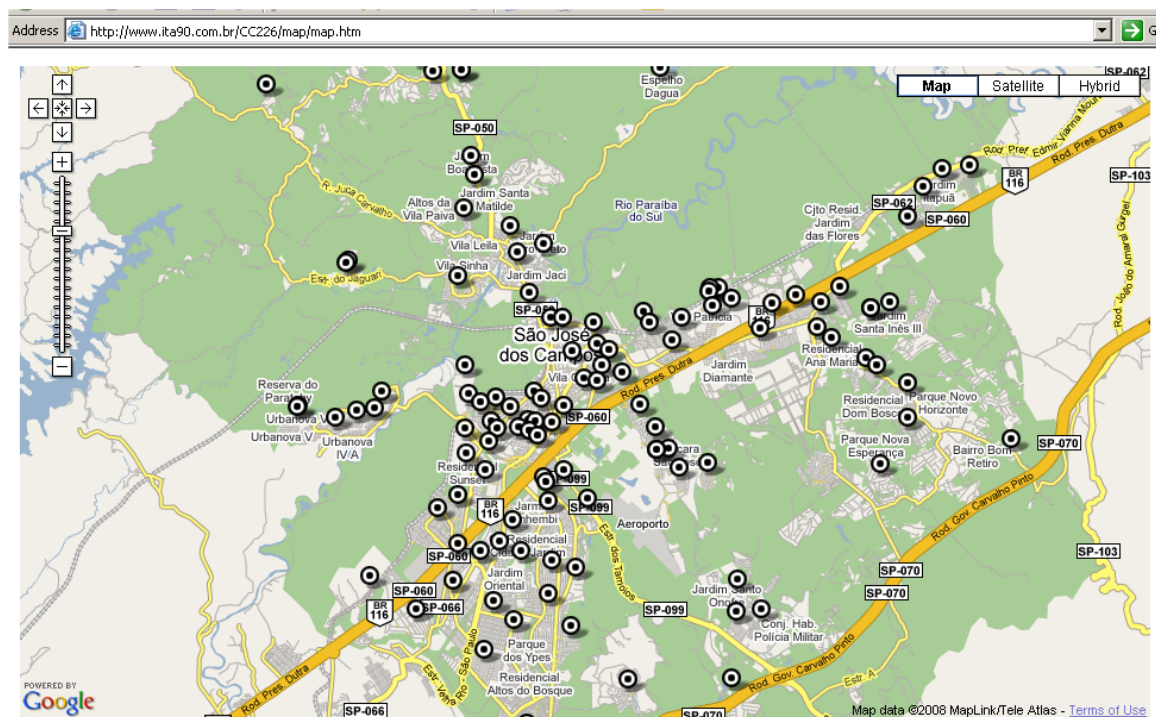


Figura 8: Bairros de São José dos Campos nos dados da Análise Exploratória.

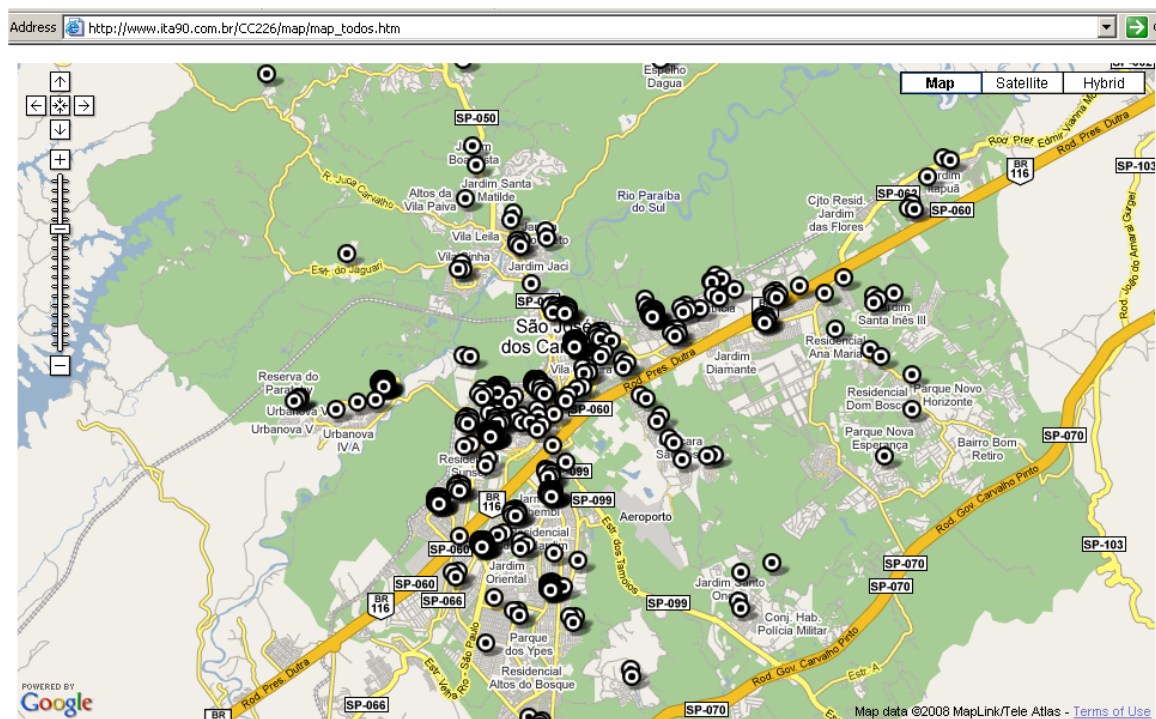


Figura 9: Posição no mapa dos 431 registros usados na Análise Exploratória.

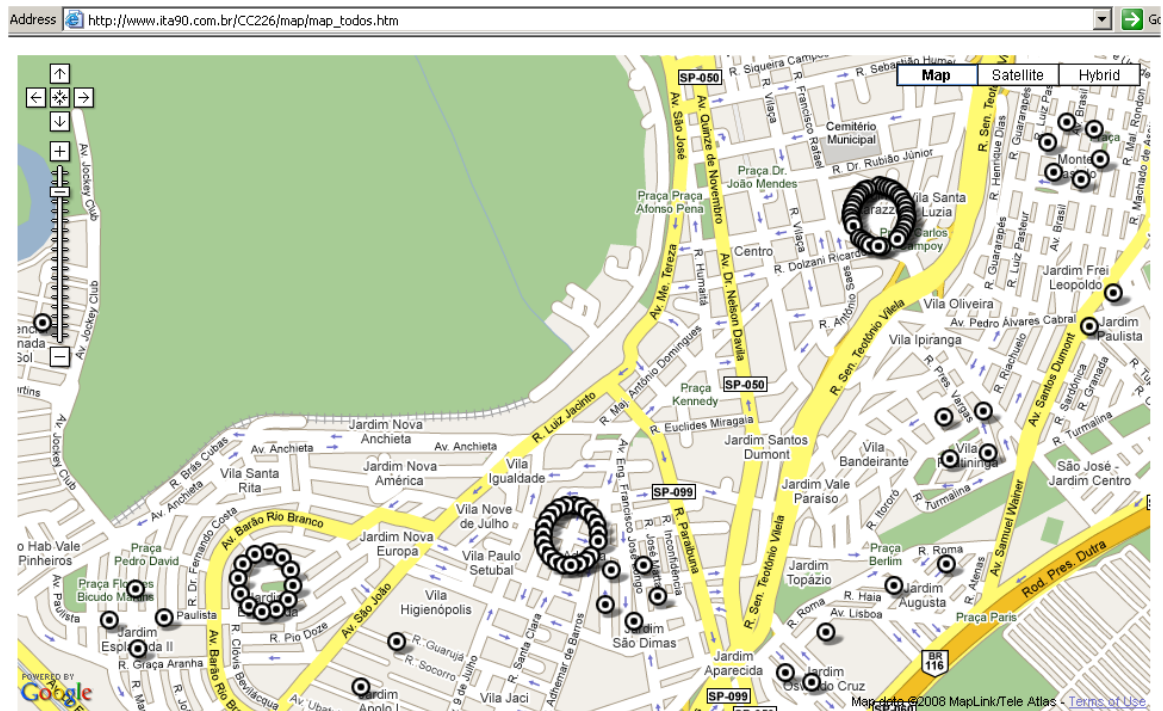


Figura 10: Zoom sobre o mapa da Figura 9 onde se pode notar a disposição em forma circular dos registros de um mesmo bairro.

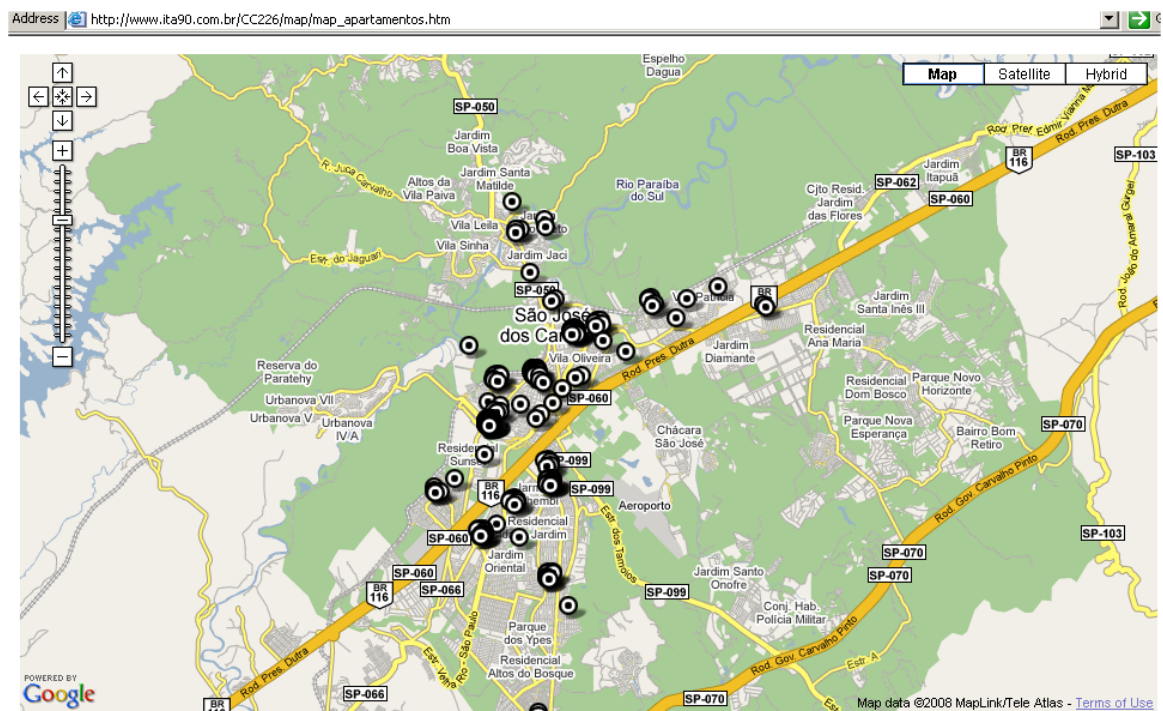
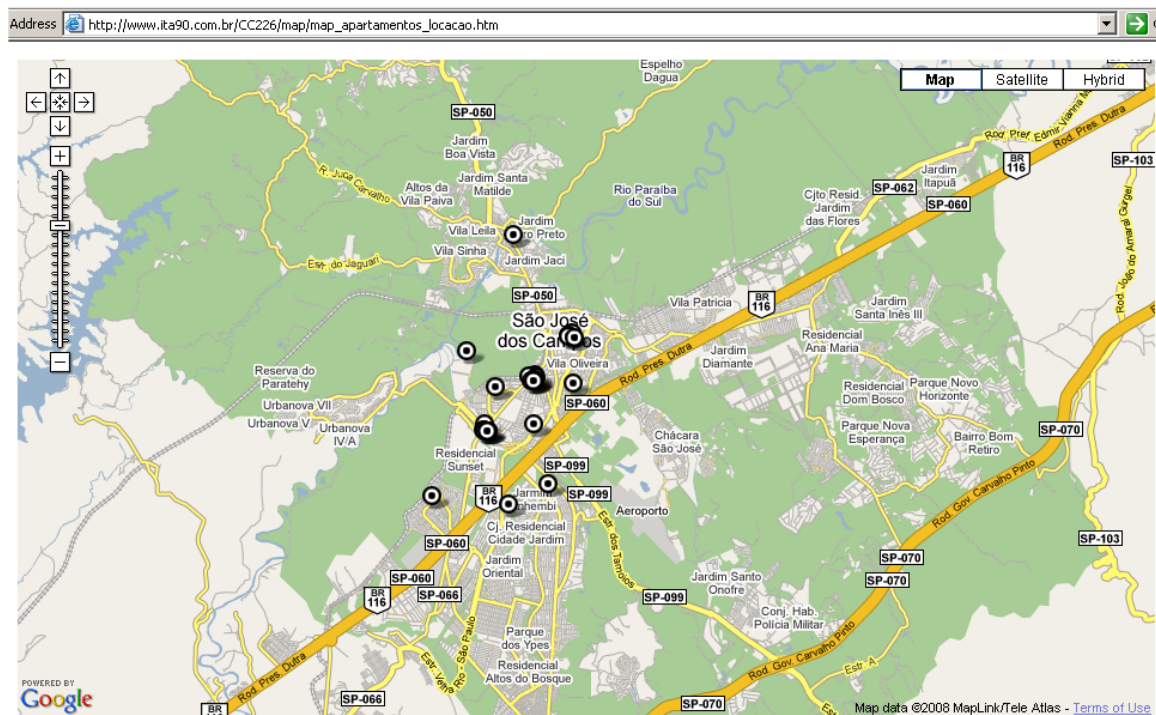
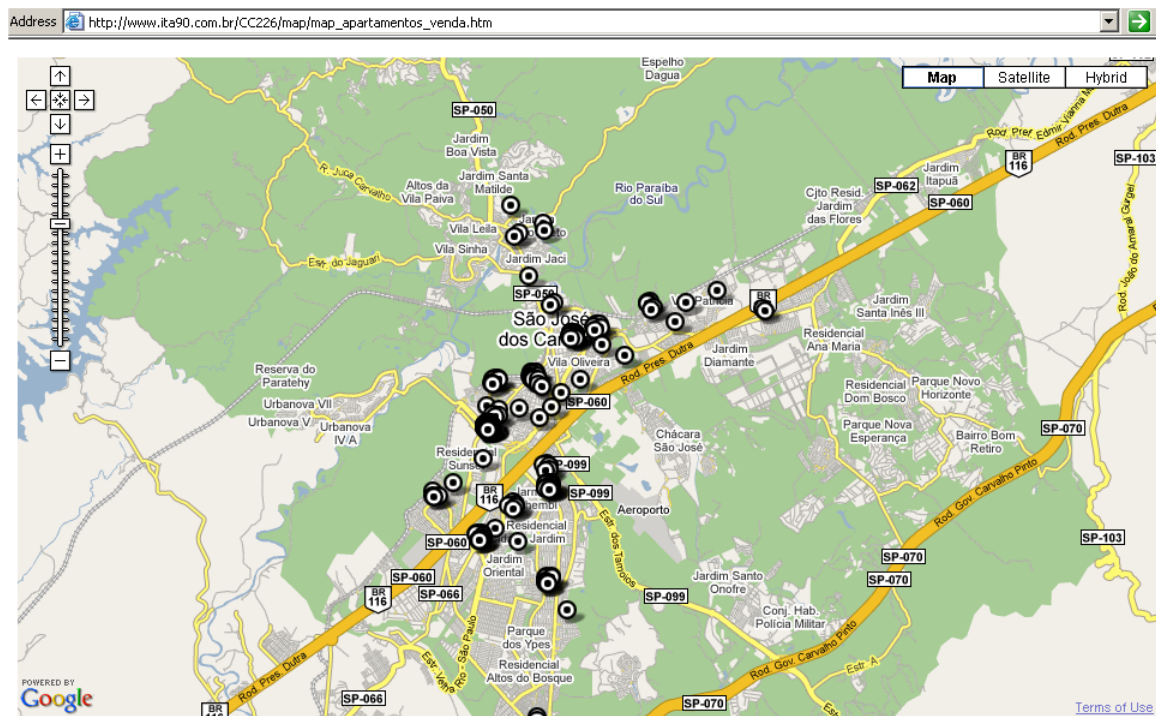


Figura 11: Posicionamento dos 189 apartamentos analisados.



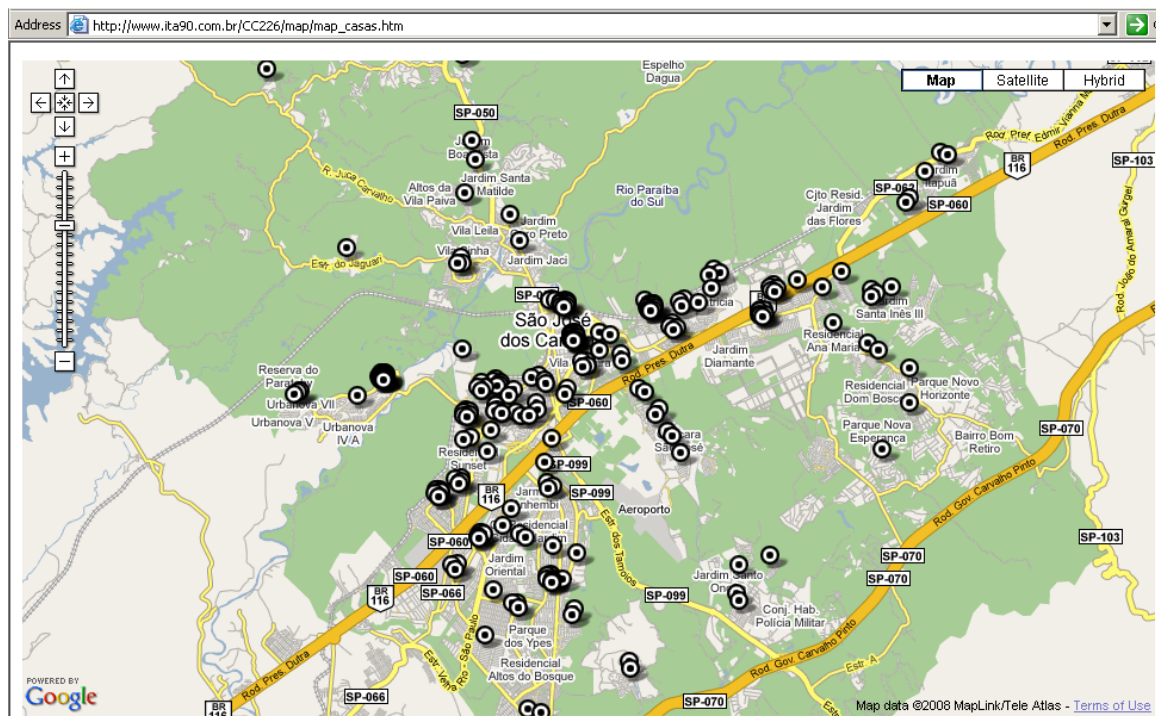


Figura 14: Posicionamento das 221 casas analisadas.

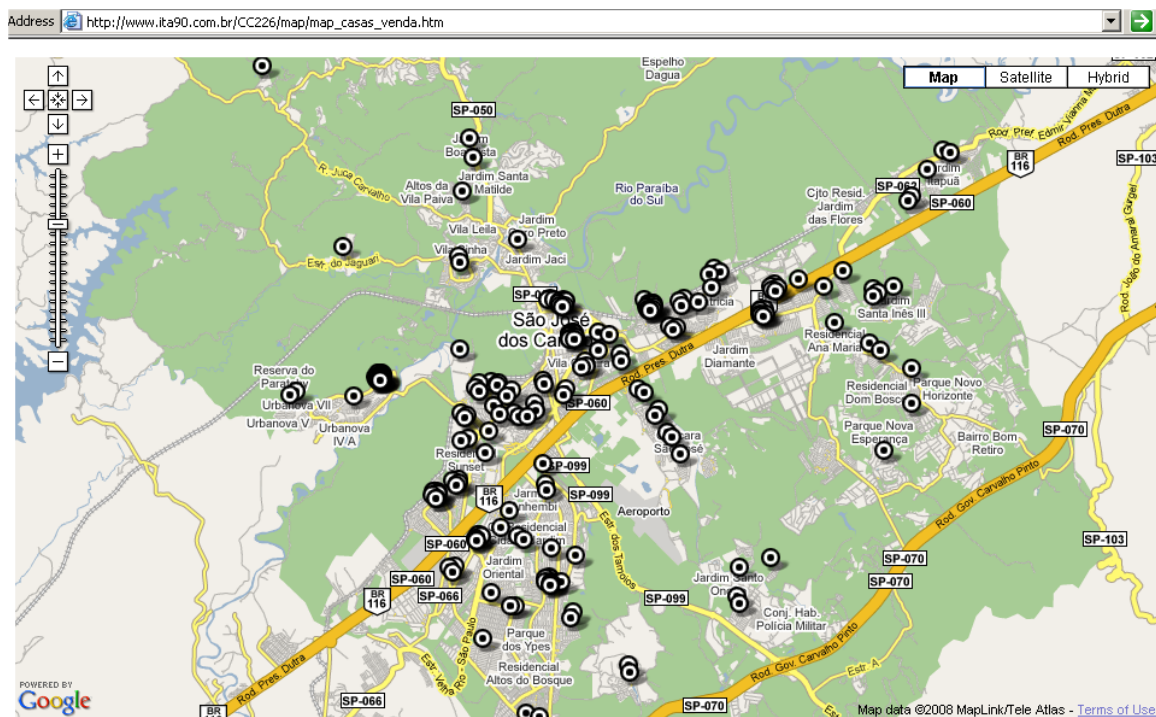


Figura 15: Posicionamento das 197 casas à venda.

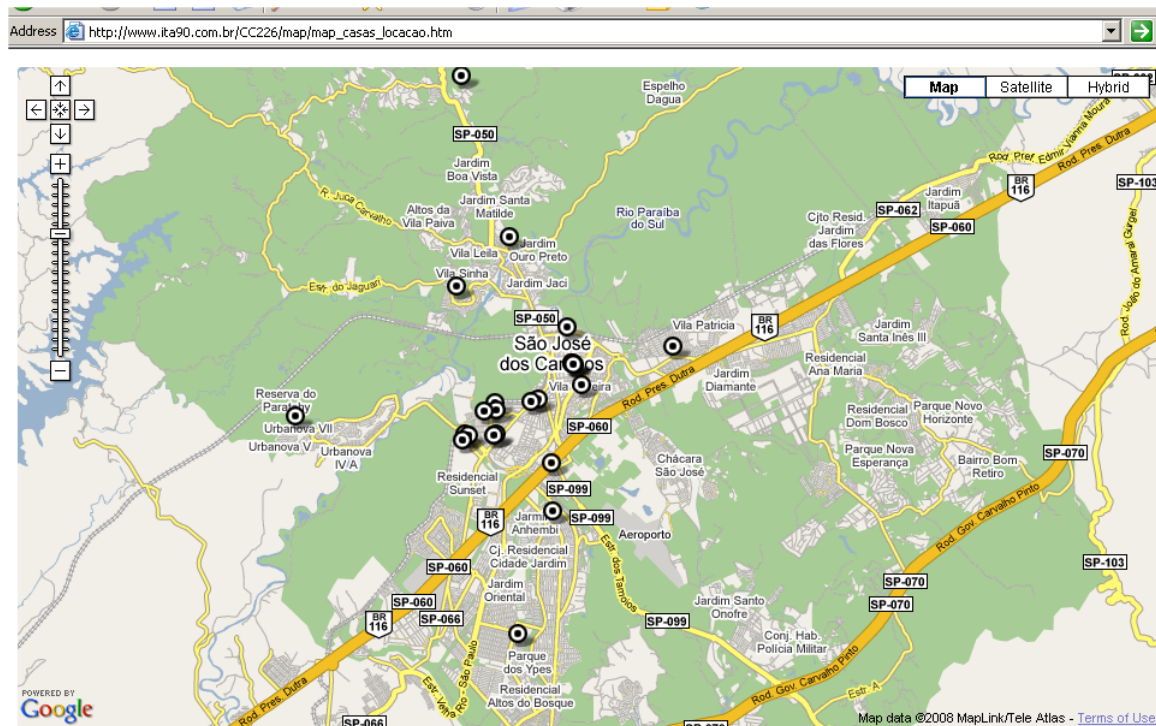


Figura 16: Posicionamento das 24 casas para locação.

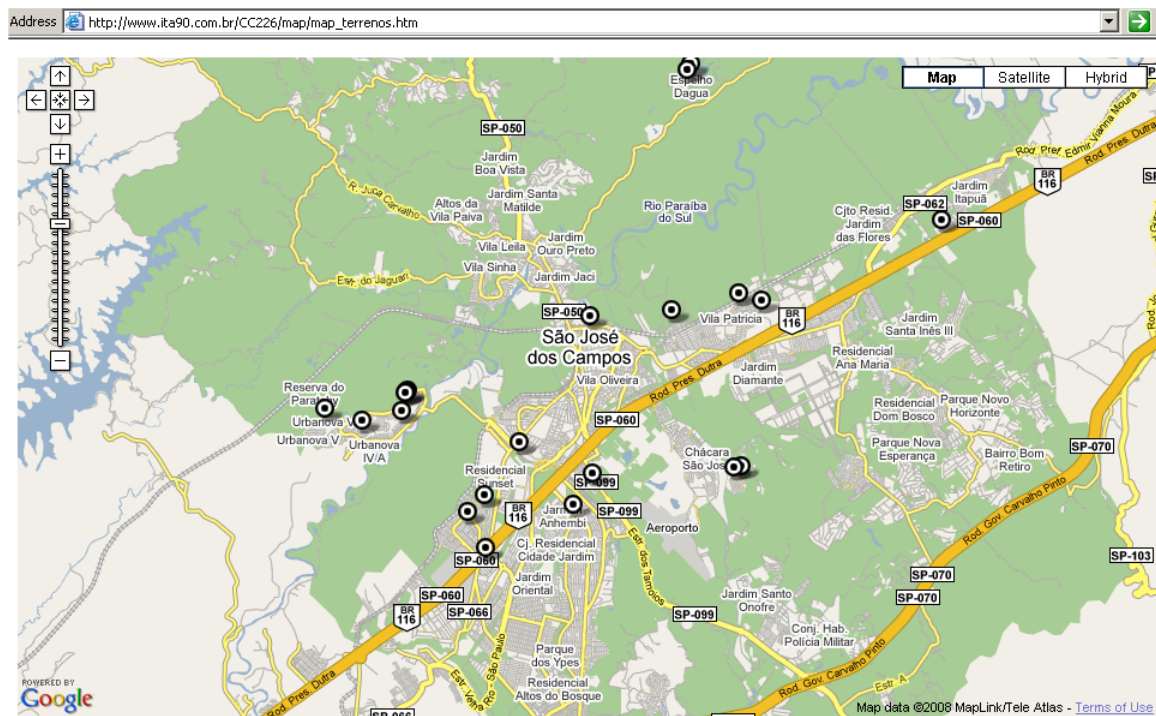


Figura 17: Posicionamento dos 21 terrenos à venda.

4.2 Vendedores

Analisando a quantidade de imóveis por vendedor (Figura 18) nota-se que:

- i) Há uma distribuição relativamente homogênea da quantidade de imóveis por vendedor.
- ii) Os vendedores com maior número de imóveis são os de código 4, 5, 11 e 10.
- iii) O vendedor 9 destaca-se pelo pequeno número de imóveis disponíveis.

Todos os vendedores possuem apartamentos à venda (Figura 19) com destaque para os vendedores 4 (maior quantidade) e 9 (menor quantidade).

Poucos vendedores possuem apartamentos para locação (Figura 20) com destaque para o vendedor 10 que os tem em maior quantidade (8 unidades). Os demais vendedores com apartamentos para locação são: 4, 5, 6, 8, 11 e 12 que possuem entre 2 e 4 unidades.

Todos os vendedores possuem casas à venda (Figura 21) com destaque para os vendedores 2, 3, 4 e 5 (maior quantidade) e 1, 6, 9 e 12 (menor quantidade).

Poucos vendedores possuem casas para locação (Figura 22) com destaque para os vendedores 8, 10 e 12 (mais unidades) e 1, 2, 3, 4, 5 e 6 (menos unidades).

Poucos vendedores possuem terrenos à venda (Figura 23) com destaque para os vendedores 4, 5 e 11 (mais unidades) e 1 (menos unidades). Os vendedores 2, 3, 6, 7, 8 e 9 não possuem terrenos à venda.

Em resumo:

- i) O vendedor 4 é o que possui maior número de imóveis, o maior número de apartamentos à venda e também o maior número de terrenos à venda.
- ii) O vendedor 10 é o que possui o maior número de apartamentos para locação.
- iii) O vendedor 11 é o segundo em número de imóveis e é o que possui o maior número de casas para locação.

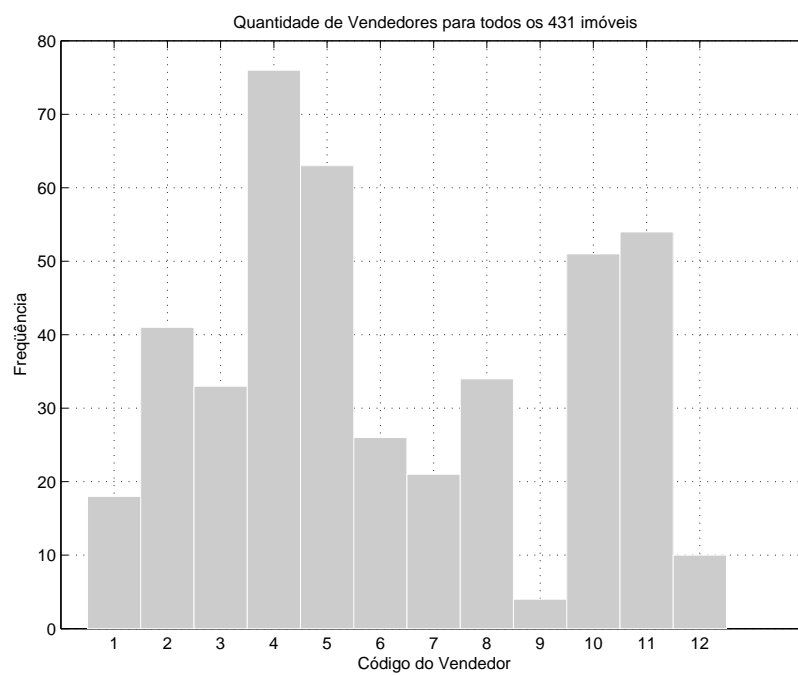


Figura 18: Vendedores de Todos os 431 Imóveis.

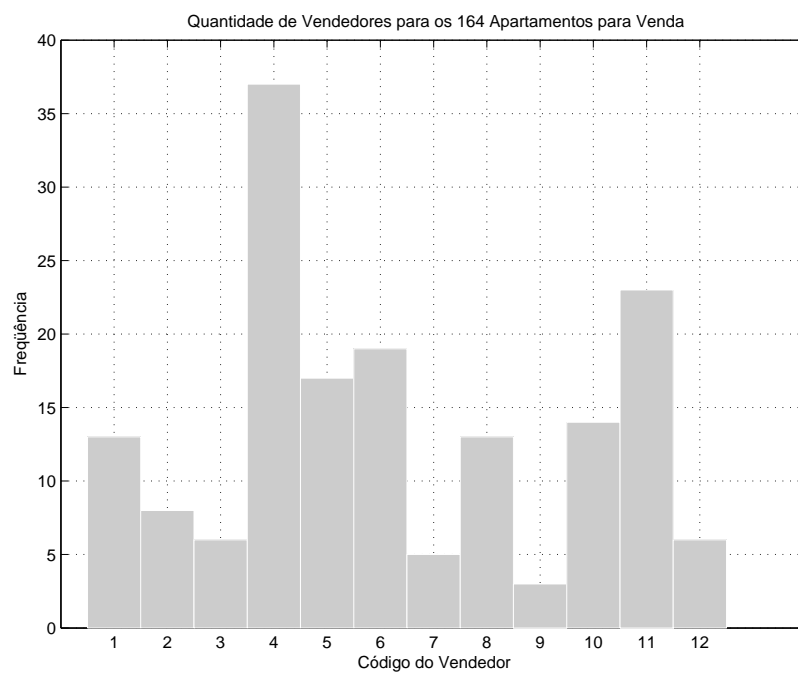


Figura 19: Vendedores dos 164 Apartamentos à Venda.

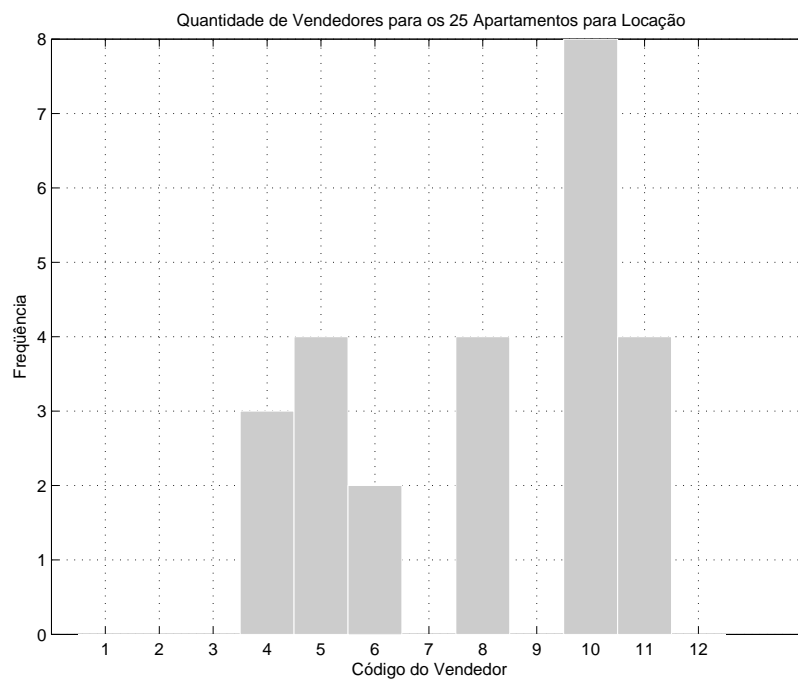


Figura 20: Vendedores dos 25 Apartamentos para Locação.

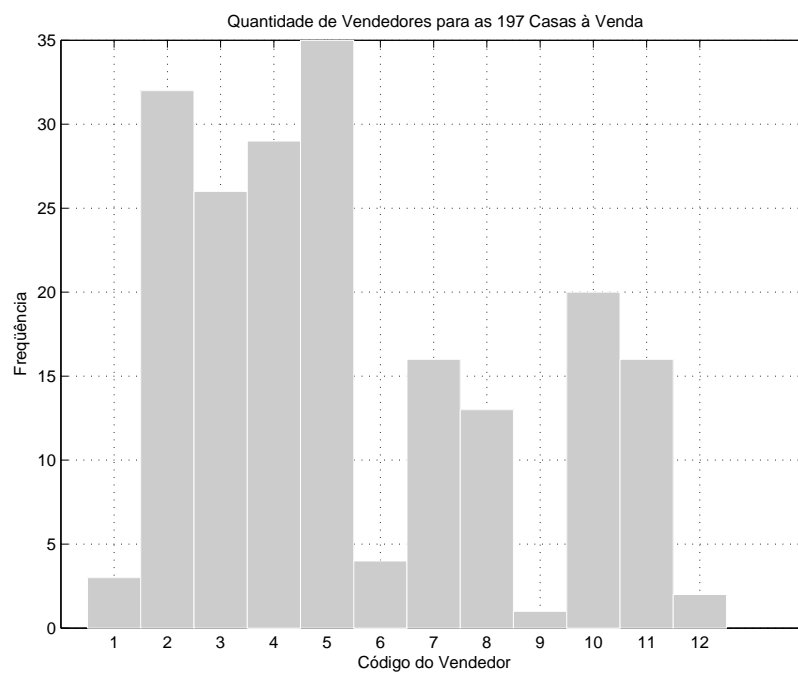


Figura 21: Vendedores das 197 Casas à Venda.

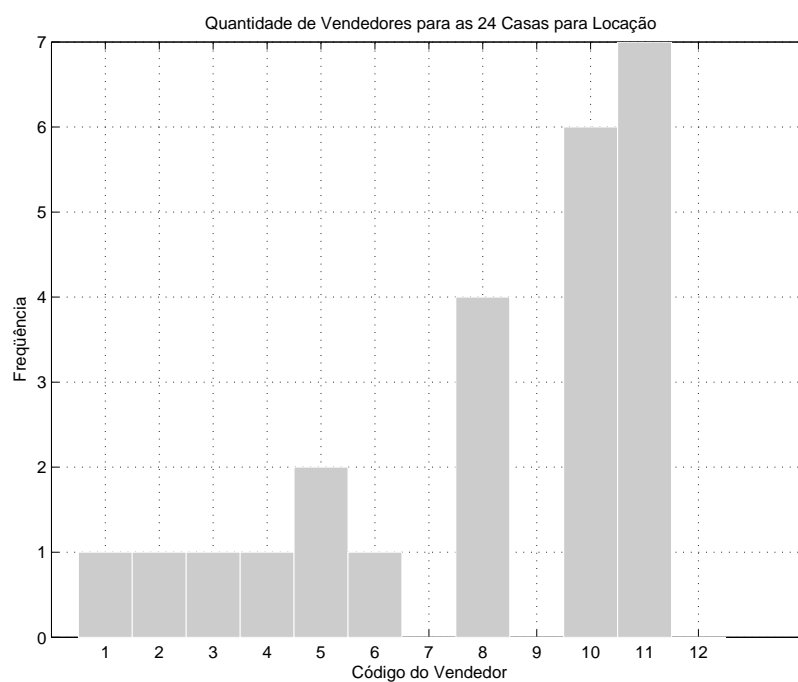


Figura 22: Vendedores das 24 Casas para Locação.

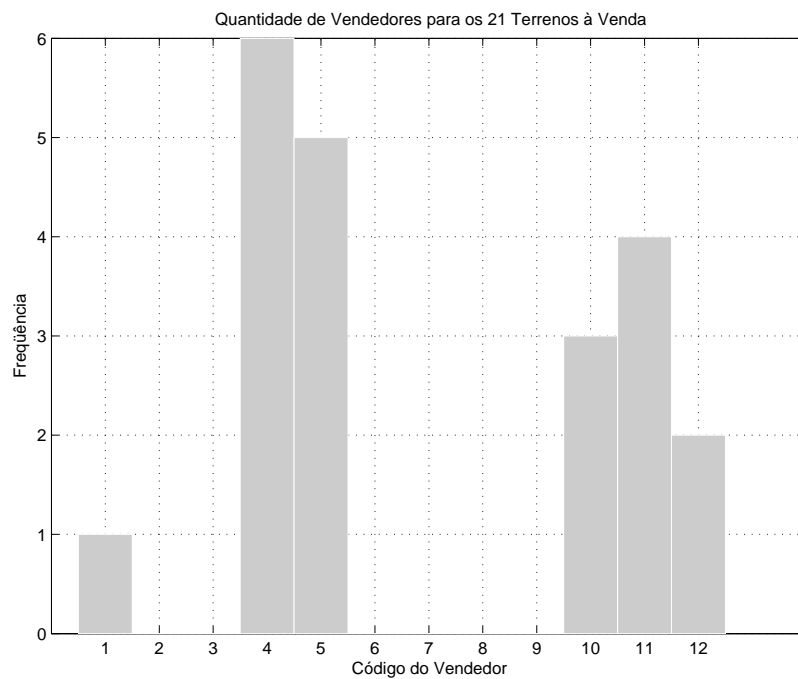


Figura 23: Vendedores dos 21 Terrenos à Venda.

4.3 Área

Analisando a área dos 431 imóveis, constatamos que podemos dividi-los em dois grupos:

- i) 416 imóveis com área menor ou igual a $500m^2$ (Figura 24).
- ii) 15 imóveis com área maior que $500m^2$ (Figura 25).

Dentre os imóveis do primeiro grupo, nota-se a predominância dos que possuem áreas entre 50 e $150m^2$ que correspondem à área típica de apartamentos e pequenas casas, que são os imóveis mais procurados.

Dentre os imóveis do segundo grupo, nota-se a predominância dos que possuem áreas entre 500 a $1000m^2$ que correspondem à área típica dos terrenos dos novos loteamentos.

Se ordenarmos todos os terrenos pela área iremos obter o gráfico da Figura 26 no qual nota-se a distribuição praticamente uniforme dos terrenos com área menor que $400m^2$. A partir deste valor, surgem alguns *outliers* com áreas de até $2500m^2$.

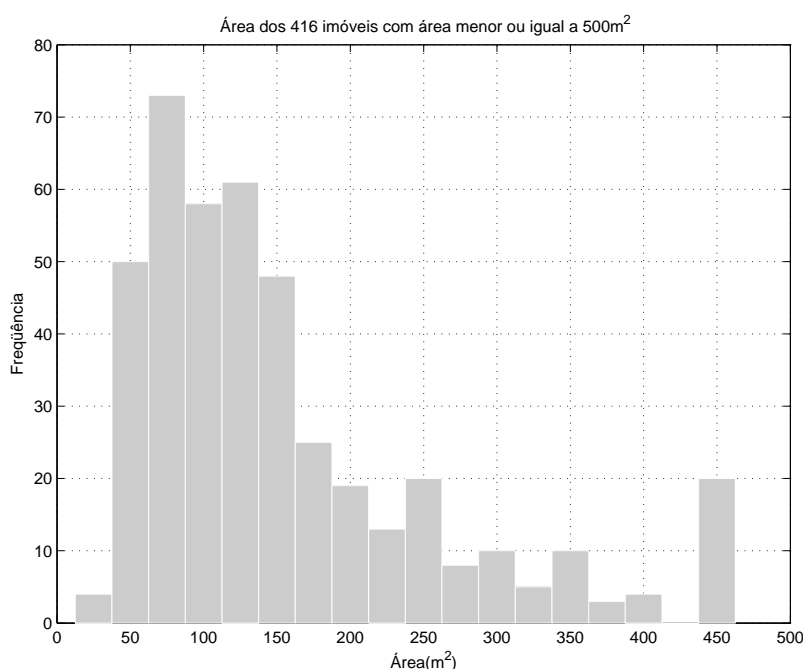


Figura 24: Histograma da Área de Todos os 416 Imóveis com Área menor ou igual a $500m^2$.

4.3.1 Área dos 164 Apartamentos à Venda

Nos apartamentos à venda (Figura 27), as áreas mais comuns estão entre 50 e $90m^2$ que correspondem aos imóveis de baixo custo.

Imóveis de alto padrão, com área superior a $200m^2$ são bem mais raros.

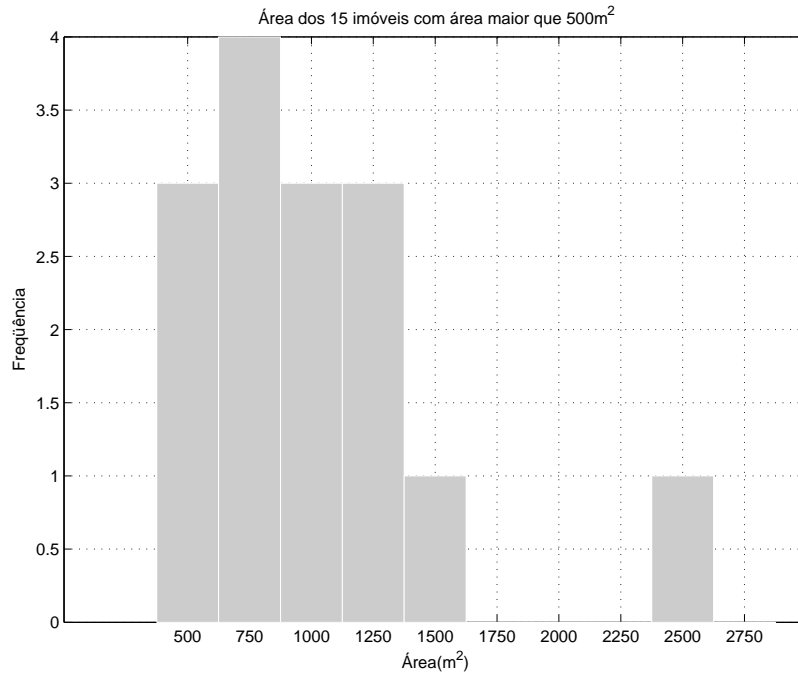


Figura 25: Histograma da Área de Todos os 15 Imóveis com Área maior que 500m².

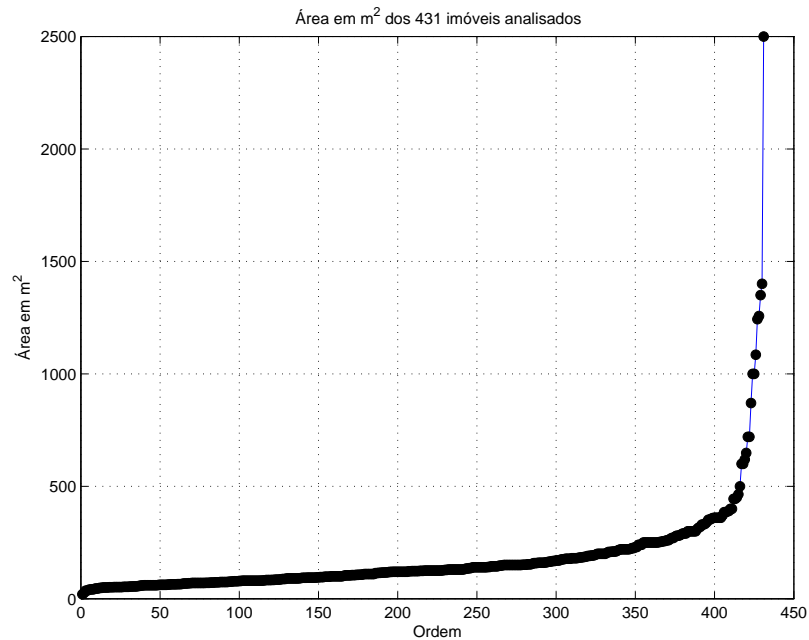


Figura 26: Área dos 431 Imóveis.

Ordenando os apartamentos à venda pela área, obtemos o gráfico da Figura 28 no qual nota-se que a maioria dos apartamentos tem área entre 50 e 140m². A partir deste valor, surgem alguns *outliers* com áreas de até 300m².

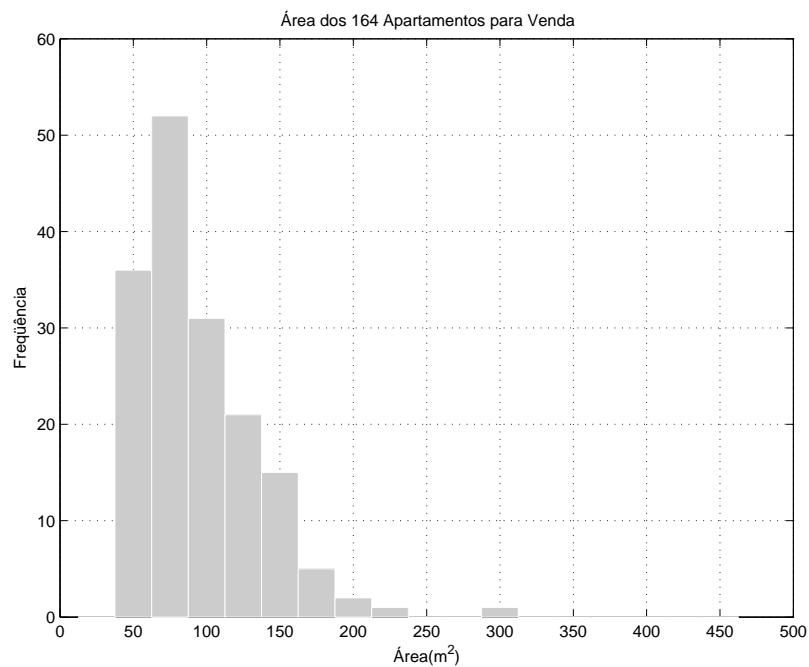


Figura 27: Histograma da Área dos 164 Apartamentos à Venda.

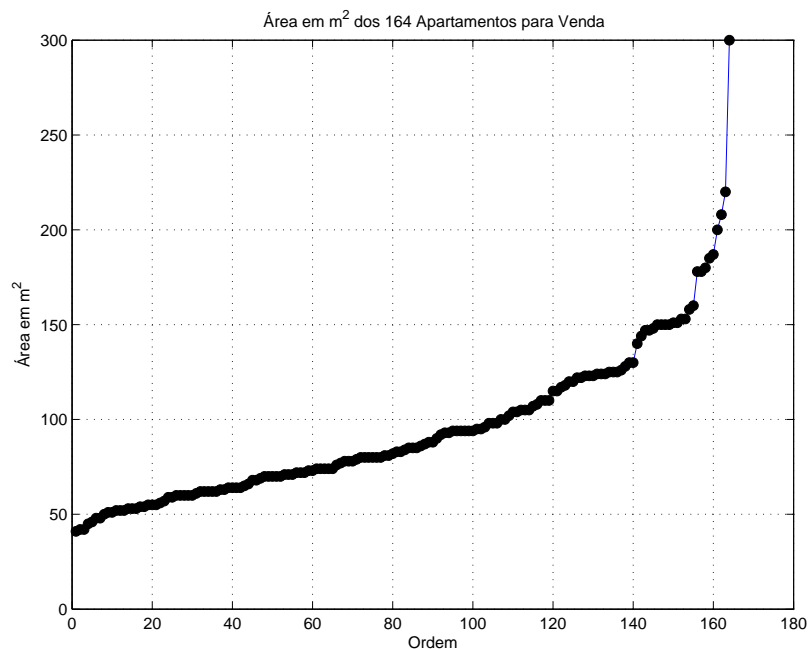


Figura 28: Área dos 164 Apartamentos à Venda.

4.3.2 Área dos 25 Apartamentos para Locação

Os apartamentos para locação (Figura 29), além de estarem em menor número que os apartamentos à venda (Figura 27), também possuem áreas mais modestas que estes, variando de 50 a $150m^2$, havendo apenas 4 imóveis com área superior a $200m^2$ e nenhum com área maior

que $300m^2$.

Ordenando os apartamentos para locação pela área, obtemos o gráfico da Figura 30 no qual nota-se a existência de três grupos: os pequenos, com área entre 40 e $50m^2$; os médios, com área entre 50 e $100m^2$ e os de alto padrão, com área entre 100 e $130m^2$.

Nota-se que não há, para locação, os apartamentos de luxo, com área acima de $200m^2$ que estão disponíveis para venda.

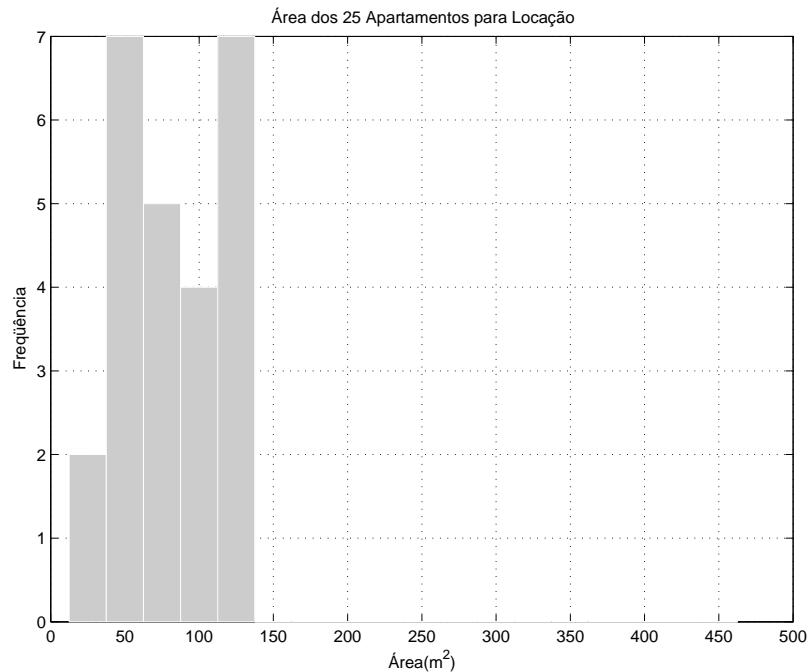


Figura 29: Histograma da Área dos 25 Apartamentos para Locação.

4.3.3 Área das 197 Casas à Venda

Embora as áreas mais comuns das casas à venda (Figura 31) varie entre 100 e $150m^2$, nota-se que há uma boa oferta de casas nos mais diversos tamanhos, desde as mais modestas com área menor que $75m^2$ até as mais luxuosas, com área maior que $300m^2$.

Chega-se à mesma conclusão analisando-se o gráfico da Figura 32.

4.3.4 Área das 24 Casas para Locação

De modo similar ao que ocorre no mercado de apartamentos, as casas para locação (Figura 33) são em menor número que as casas à venda (Figura 31).

No entanto, ao contrário do que ocorre no mercado de apartamentos, há casas disponíveis para praticamente todos os valores de áreas, notando-se quatro agrupamentos distintos em tornos das áreas 100, 250, 350 e $450m^2$.

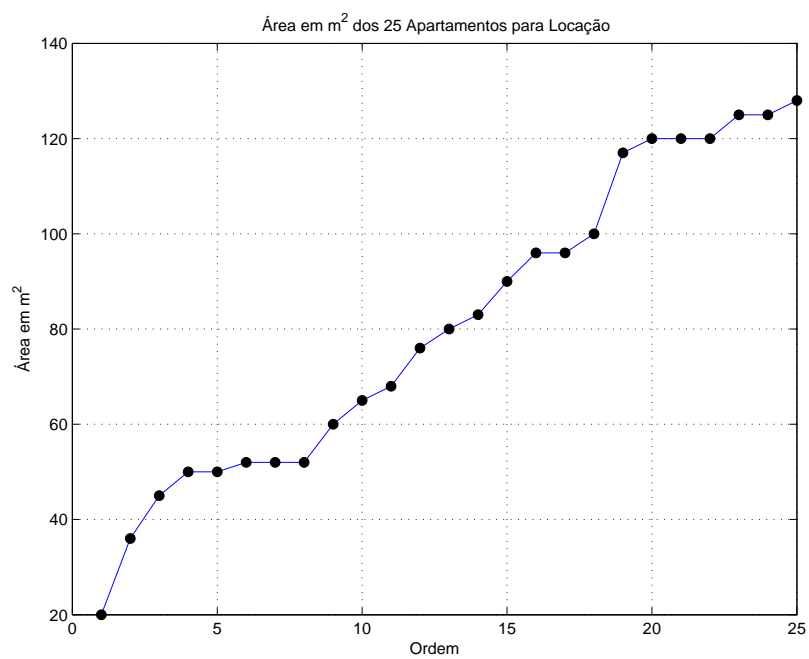


Figura 30: Área dos 25 Apartamentos para Locação.

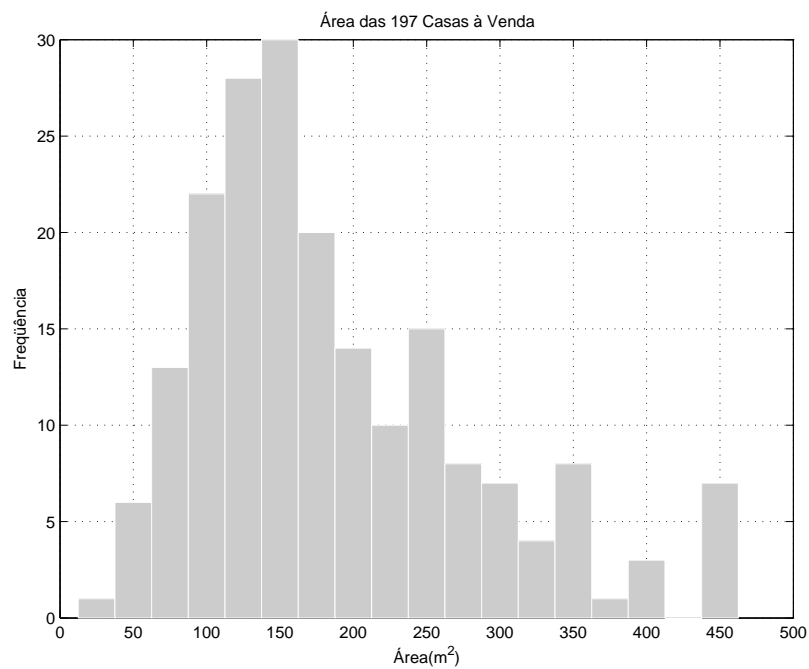


Figura 31: Histograma da Área das 197 Casas à Venda.

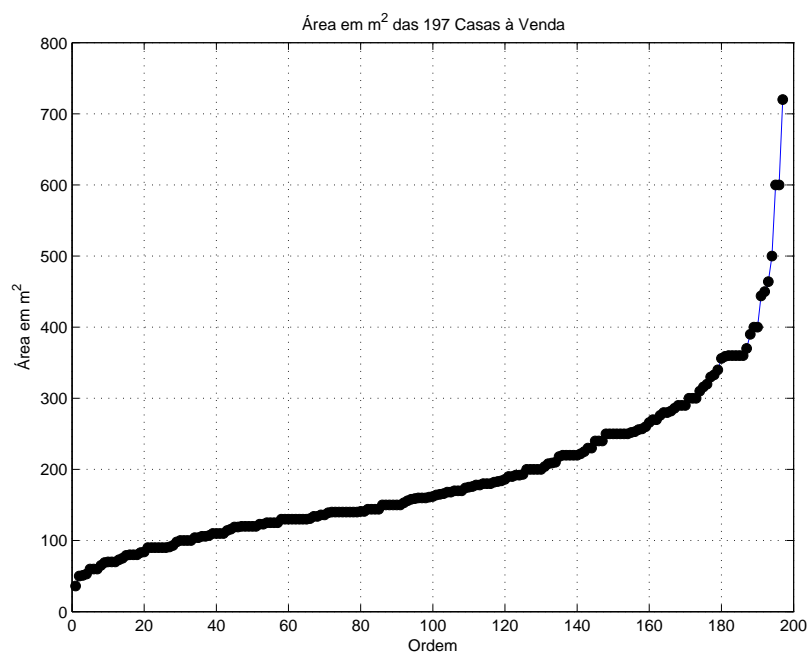


Figura 32: Área das 197 Casas à Venda.

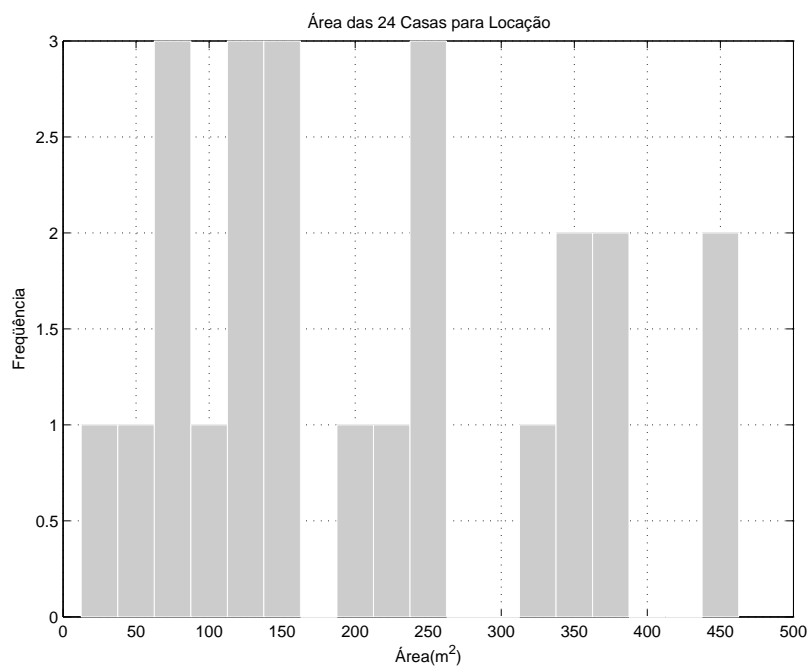


Figura 33: Histograma da Área das 24 Casas para Locação.

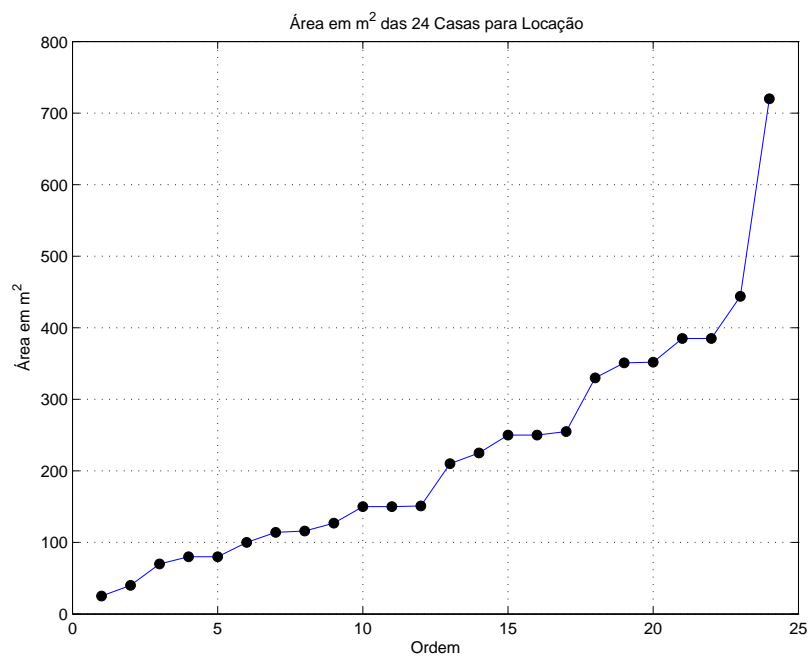


Figura 34: Área das 24 Casas para Locação.

4.3.5 Área dos 21 Terrenos à Venda

Apesar de haver relativamente poucos terrenos à venda (Figura 35), nota-se a existência de três grupos quanto à área: um grupo com áreas entre 100 e 300 m^2 , outro com áreas em torno dos 100 m^2 e o último, do qual consta apenas um imóvel, com um terreno de 2500 m^2 de área.

No gráfico da Figura 35 nota-se como se destaca dentre os demais o terreno de 2500 m^2 .

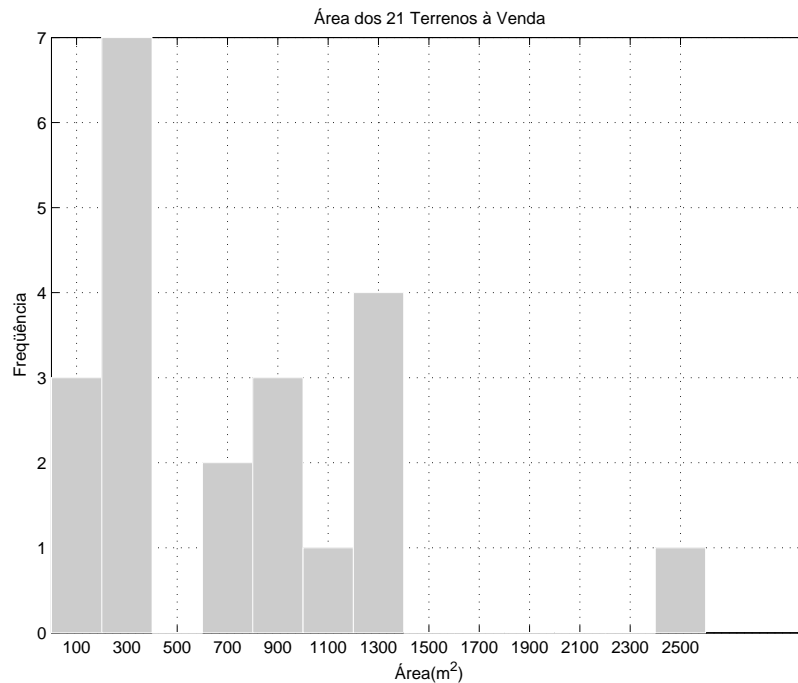


Figura 35: Histograma da Área dos 21 Terrenos à Venda.

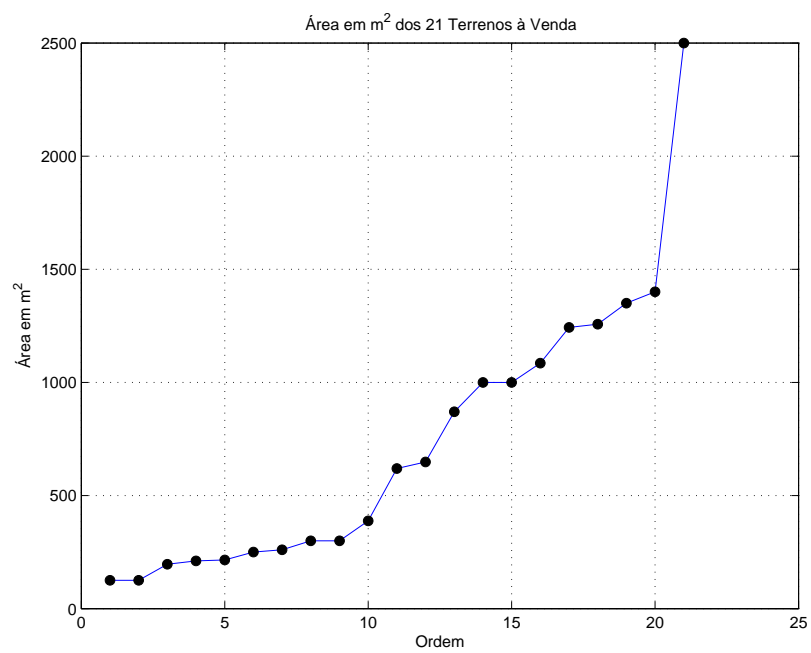


Figura 36: Área dos 21 Terrenos à Venda.

4.4 Preço em Reais

Analisando o preço em reais dos 431 imóveis (Figura 37), constatamos que a grande maioria tem preço menor que R\$800.000,00.

Pelo gráfico da (Figura 38) notamos que os dados podem ser agrupados em três grupos:

- i) Há cerca de 50 imóveis com preços menores que R\$10.000,00. Trata-se dos imóveis destinados a locação cujo preço refere-se ao aluguel mensal que é da ordem de 1% do valor do imóvel.
- ii) Em seguida, há cerca de 375 imóveis com preços entre R\$10.000,00 e R\$800.000,00 reais.
- iii) Finalmente, há um grupo de 6 *outliers* com preço acima de R\$800.000,00 reais, que formam os imóveis de alto padrão.

Como há uma variação muito grande nos preços, é interessante usar o logaritmo decimal do preço conforme mostrado nas Figuras 39 e 40.

Usando o logaritmo decimal, fica ainda mais claro a diferenciação em três grupos pois os 50 imóveis mais baratos ficam mais destacados.

Com os gráficos das Figuras 39 e 42 nota-se um acúmulo de imóveis de alto padrão em torno dos bairros Aquários, mais ao norte, e Bosque dos Eucaliptos, mais ao sul. É interessante notar que, a região central não é tão valorizada como estes bairros que, além de bem mais jovens, oferecem mais conforto e segurança a seus moradores.

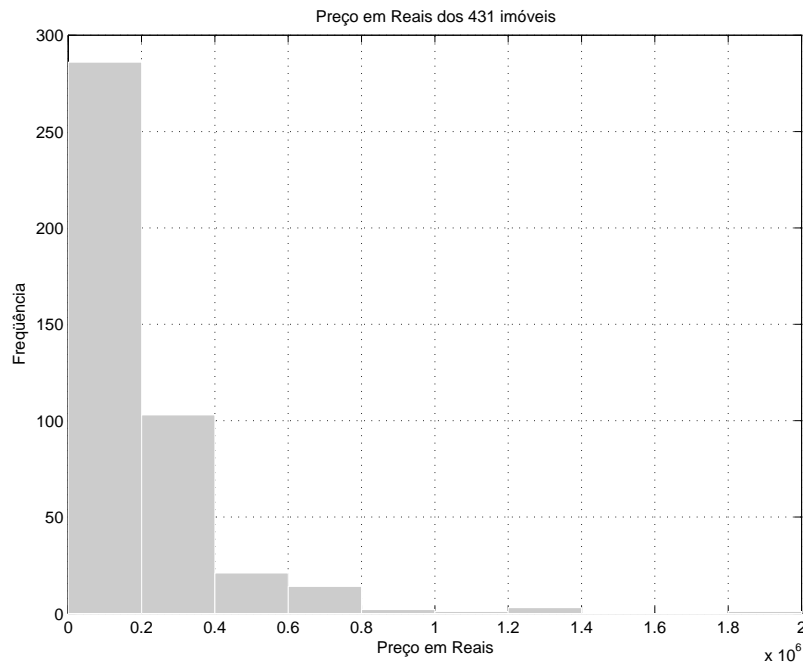


Figura 37: Histograma do Preço em Reais de Todos os 431 Imóveis.

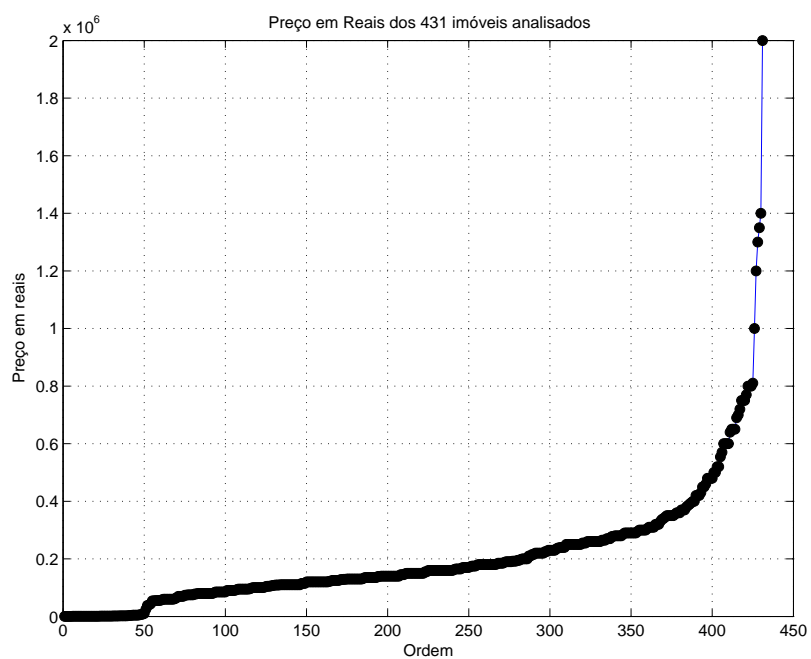


Figura 38: Preço dos 431 Imóveis.

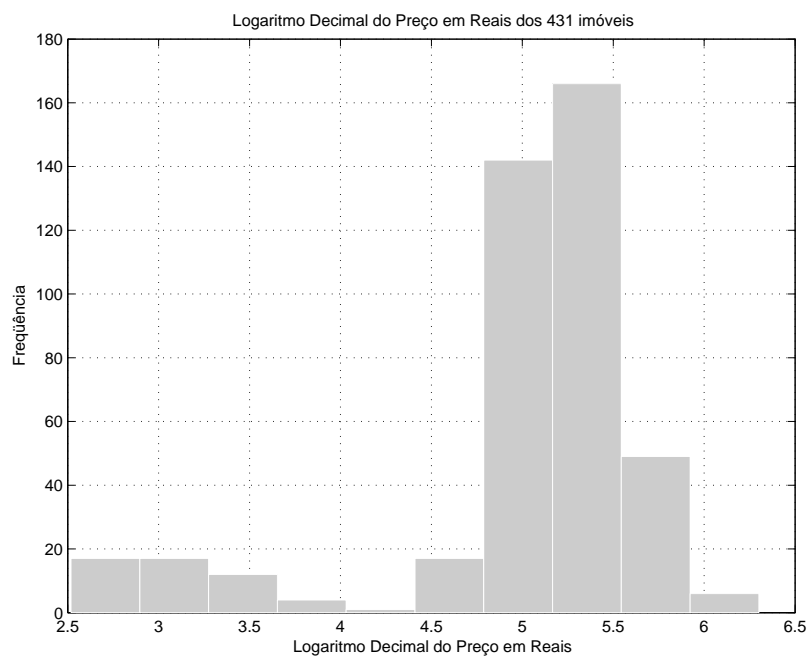


Figura 39: Histograma do Logaritmo Decimal do Preço em Reais de Todos os 431 Imóveis.

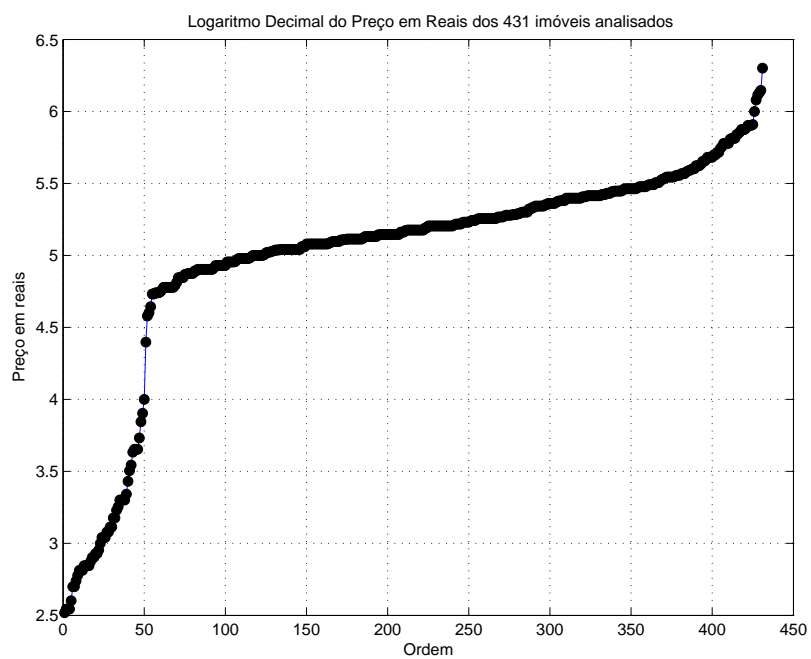


Figura 40: Logaritmo Decimal do Preço dos 431 Imóveis.

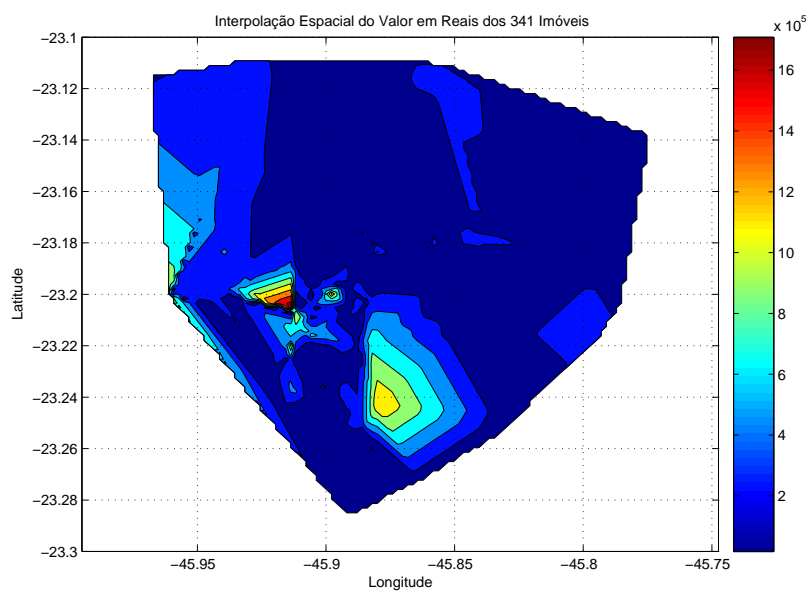


Figura 41: Valor do Imóvel em função da Latitude e da Longitude, para todos os 431 imóveis.

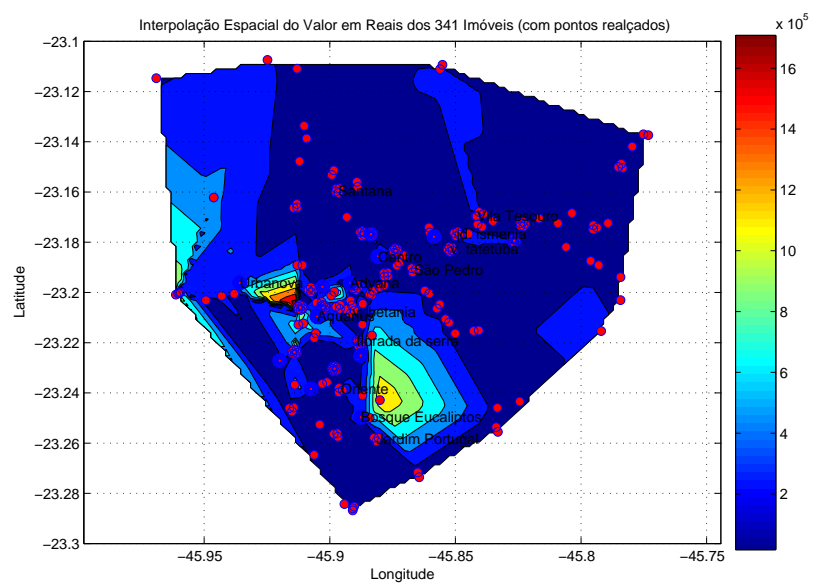


Figura 42: Valor do Imóvel em função da Latitude e da Longitude, para todos os 431 imóveis, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

4.4.1 Preço em Reais dos 164 Apartamentos à Venda

Os gráficos das Figuras 43 a 46 demonstram que a maioria dos apartamentos à venda, têm preços entre $R\$50.000,00$ e $R\$400.000,00$, sendo bem mais comuns os apartamentos mais baratos, com preços variando entre $R\$60.000,00$ e $R\$120.000,00$.

Os gráficos das Figuras 47 e 48 demonstram que os apartamentos à venda concentram-se próximos ao centro da cidade. Os imóveis de maior valor acumulam-se no entorno dos bairros Aquários e Vila Adyana, ao norte, e Bosque dos Eucaliptos, mais ao sul.

É interessante notar que a região nordeste da cidade, em torno do Jardim Ismênia, obteve uma relativa valorização apesar da grande distância ao centro e às outras regiões mais valorizadas. Uma possível explicação para este fenômeno deriva da proximidade com a Petrobrás que está ampliando suas instalações na cidade e contrantando muitos novos funcionários os quais desejam morar próximo ao seu local de trabalho.

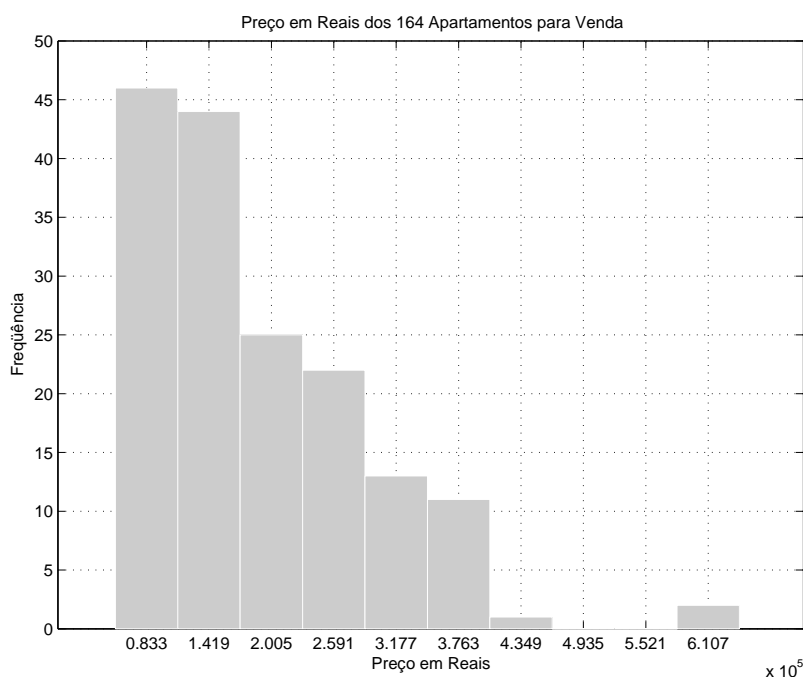


Figura 43: Histograma do Preço em Reais dos 164 Apartamentos à Venda.

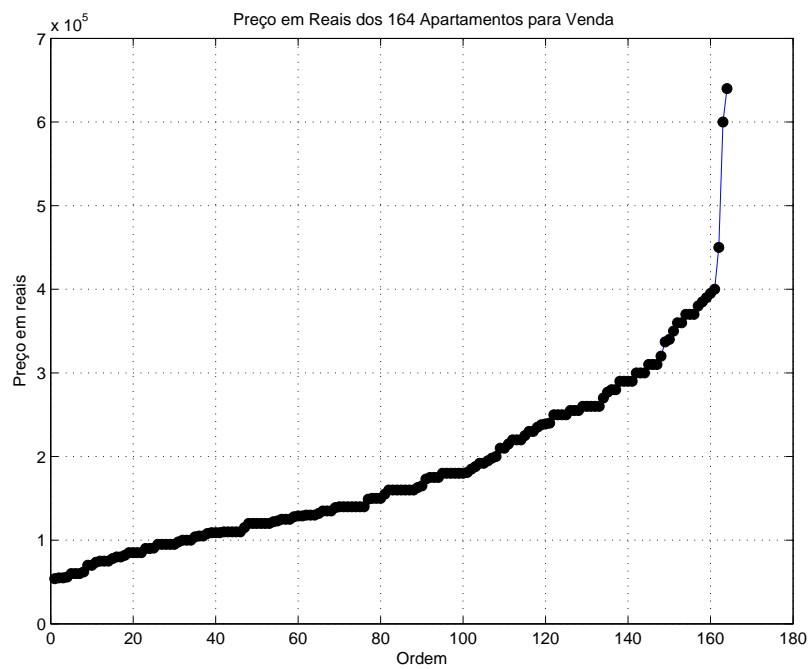


Figura 44: Preço dos 164 Apartamentos à Venda.

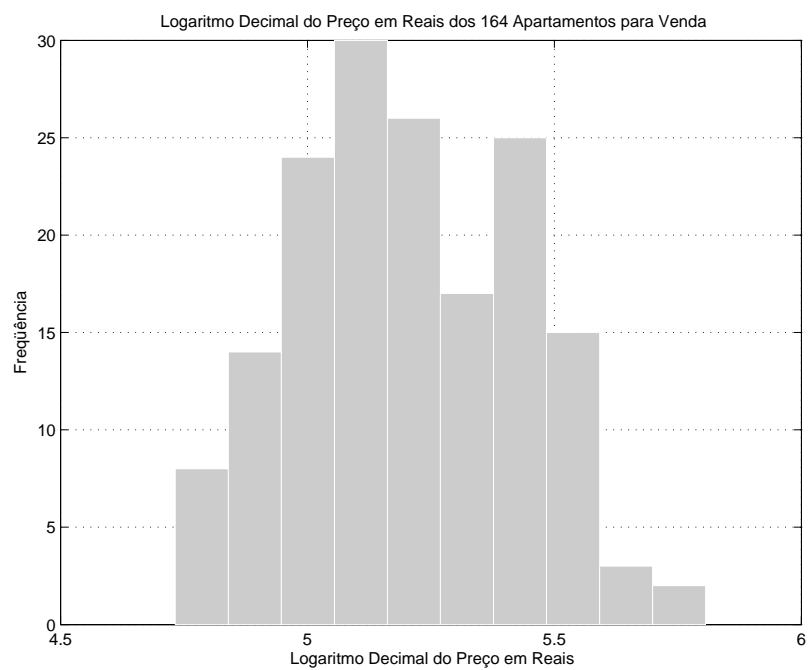


Figura 45: Histograma do Logaritmo Decimal do Preço em Reais dos 164 Apartamentos à Venda.

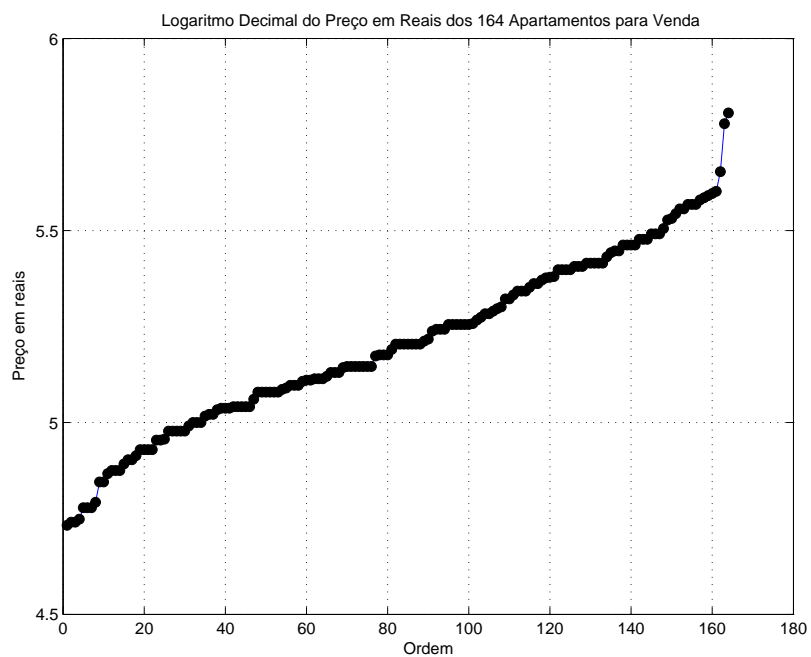


Figura 46: Logaritmo Decimal do Preço dos 164 Apartamentos à Venda.

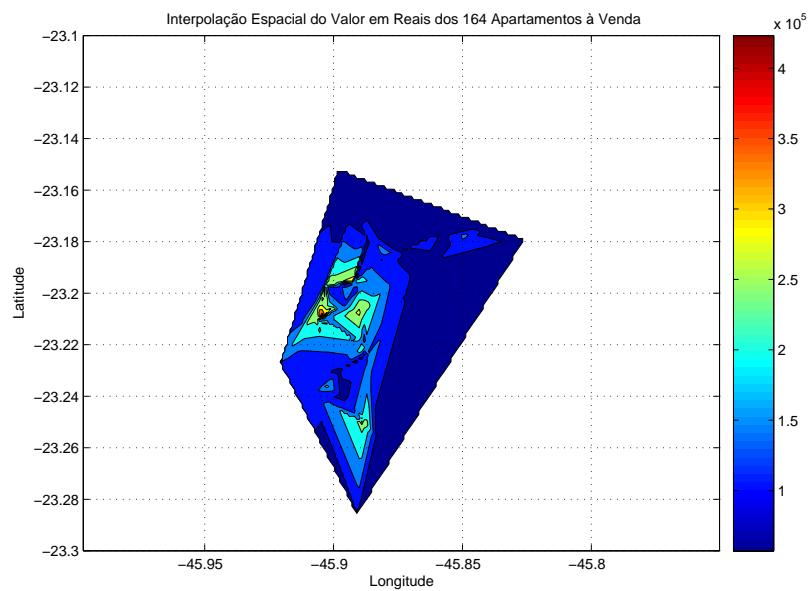


Figura 47: Valor do Imóvel em função da Latitude e da Longitude, para todos os 164 Apartamentos à Venda.

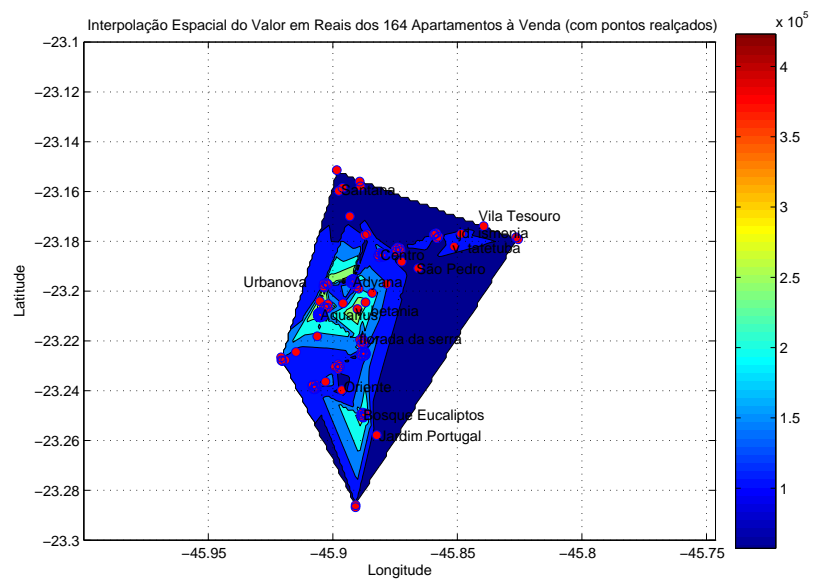


Figura 48: Valor do Imóvel em função da Latitude e da Longitude, para todos os 164 Apartamentos à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

4.4.2 Preço em Reais dos 25 Apartamentos para Locação

Os gráficos das Figuras 49 e 50 demonstram que os apartamentos para locação podem ser agrupados em três grupos:

- i) Na faixa de preços entre $R\$400,00$ e $R\$800,00$, estão os apartamentos mais baratos, sendo os preços mais comuns os da faixa entre $R\$700,00$ e $R\$800,00$.
- ii) Em seguida, há os imóveis de preço médio, com preços variando entre $R\$1.000,00$ e $R\$1.400,00$.
- iii) Finalmente, há uma elite de 2 imóveis com aluguéis em torno de $R\$1.800,00$.

Os gráficos das Figuras 51 e 52 demonstram, com mais vigor, a mesma subdivisão em três grupos acima citada.

Os gráficos das Figuras 53 e 54 demonstram, para os apartamentos para locação, um comportamento semelhante ao observado para os apartamentos à venda:

- i) Os imóveis estão em áreas próximas ao centro.
- ii) Os imóveis mais caros estão entre os bairros Jd. Aquários e Vila Adyana.

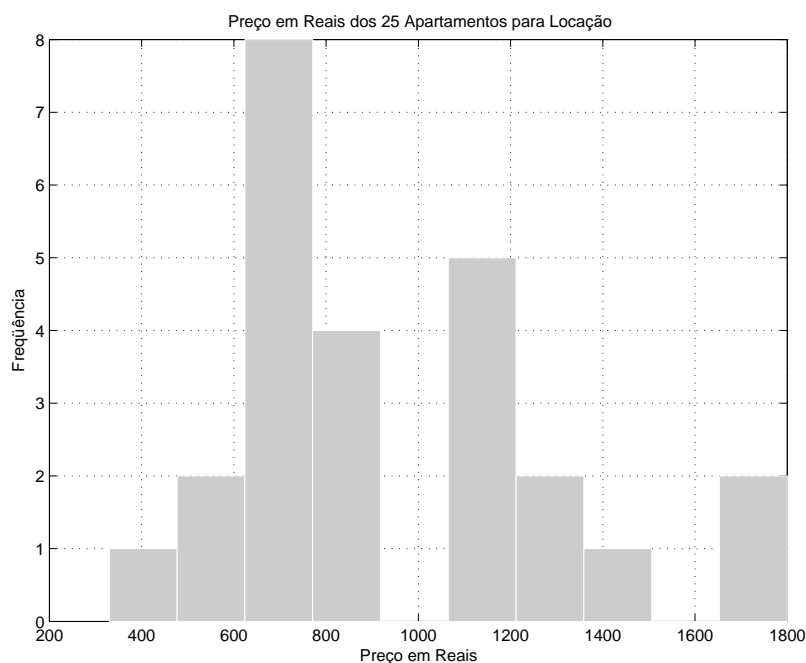


Figura 49: Histograma do Preço em Reais dos 25 Apartamentos para Locação.

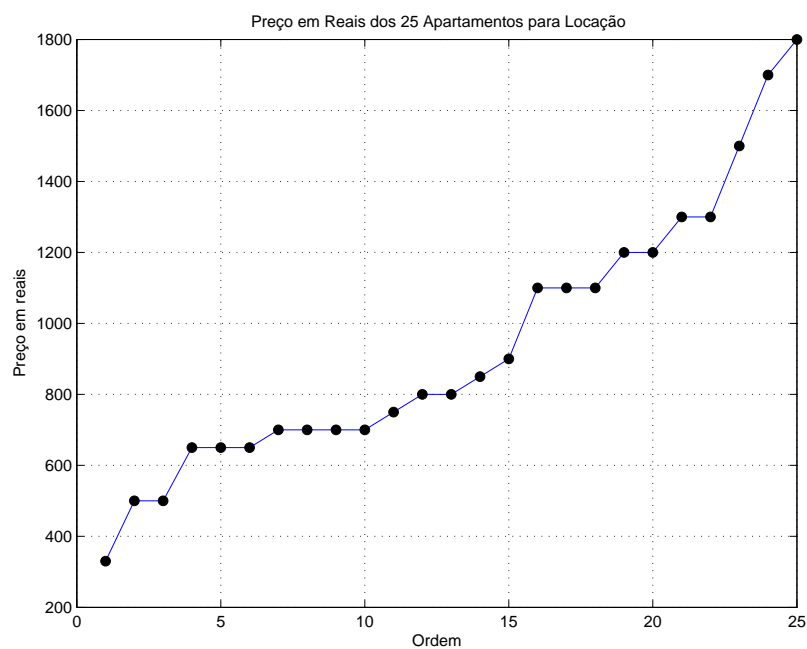


Figura 50: Preço dos 25 Apartamentos para Locação.

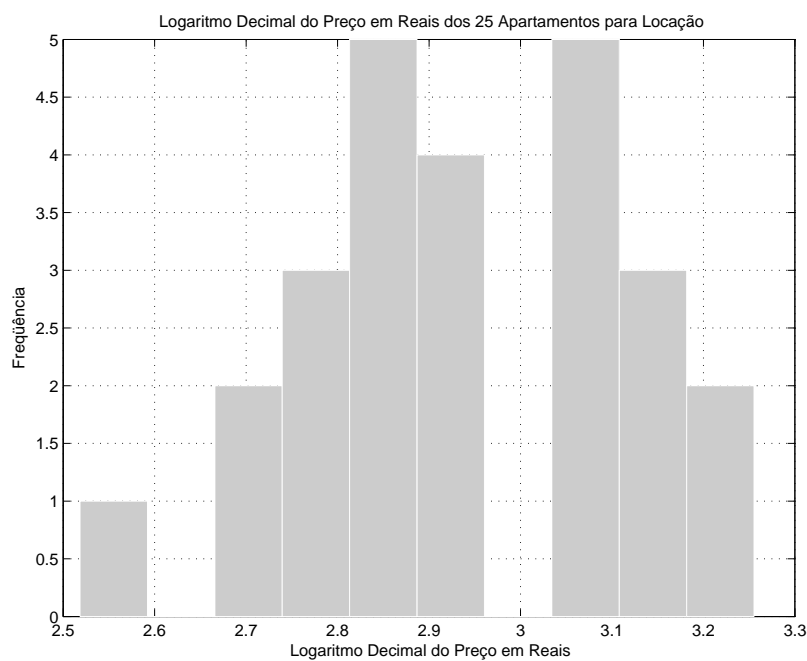


Figura 51: Histograma do Logaritmo Decimal do Preço em Reais dos 25 Apartamentos para Locação.

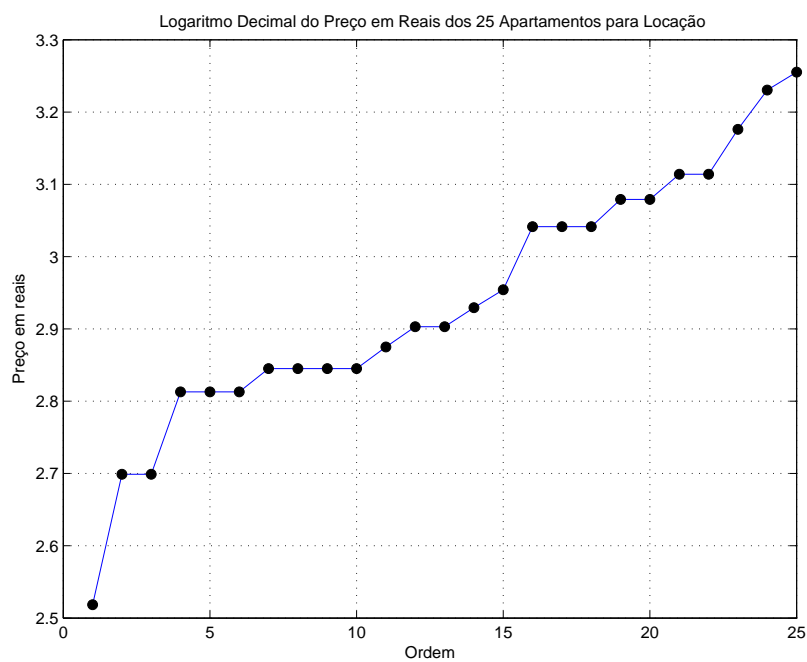


Figura 52: Logaritmo Decimal do Preço dos 25 Apartamentos para Locação.

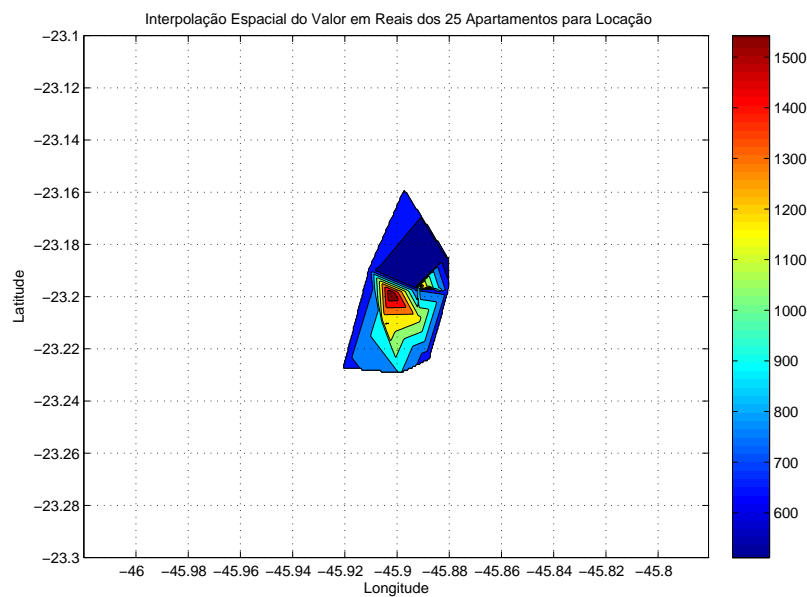


Figura 53: Valor do Imóvel em função da Latitude e da Longitude, para todos os 25 Apartamentos para Locação.

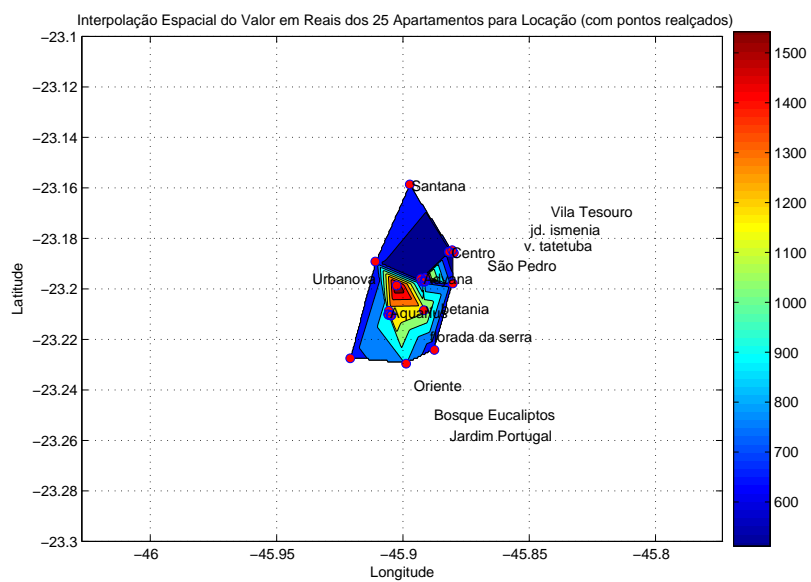


Figura 54: Valor do Imóvel em função da Latitude e da Longitude, para todos os 25 Apartamentos para Locação, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

4.4.3 Preço em Reais das 197 Casas à Venda

Os gráficos das Figuras 55 a 58 demonstram que a maioria das casas à venda têm preço inferior a R\$800.000,00, sendo mais comuns as com preços abaixo de R\$40.000,00. Há uma elite de 6 imóveis com preços acima de R\$800.000,00.

Os gráficos das Figuras 59 e 60 mostram que os imóveis de maior preço agrupam-se em torno do Jardim Esplanada e do Jardim Oriente (próximo ao Bosque dos Eucaliptos). Nota-se que há casas de alto padrão no lado oeste da cidade, próximas à Reserva do Parateí, considerado um dos pontos de expansão para imóveis terrenos atualmente em São José dos Campos.

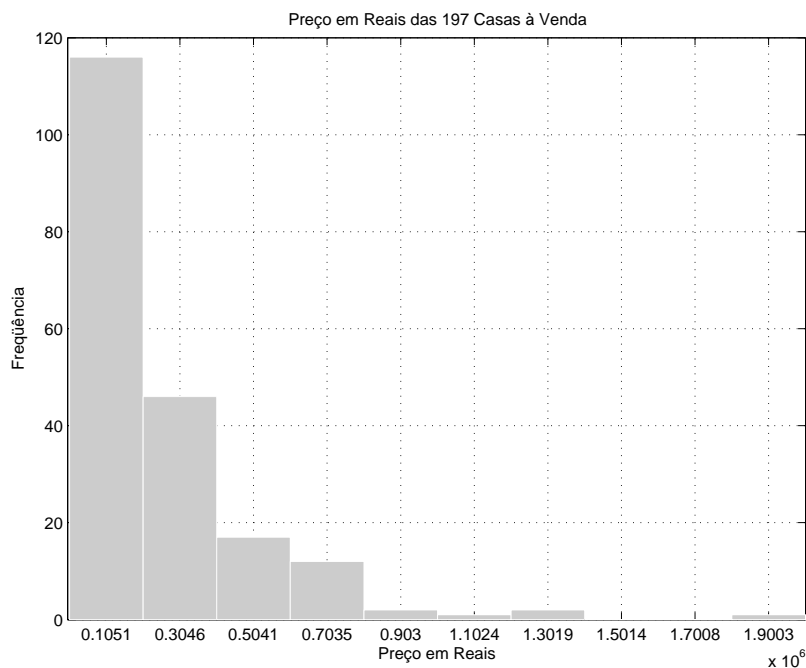


Figura 55: Histograma do Preço em Reais das 197 Casas à Venda.

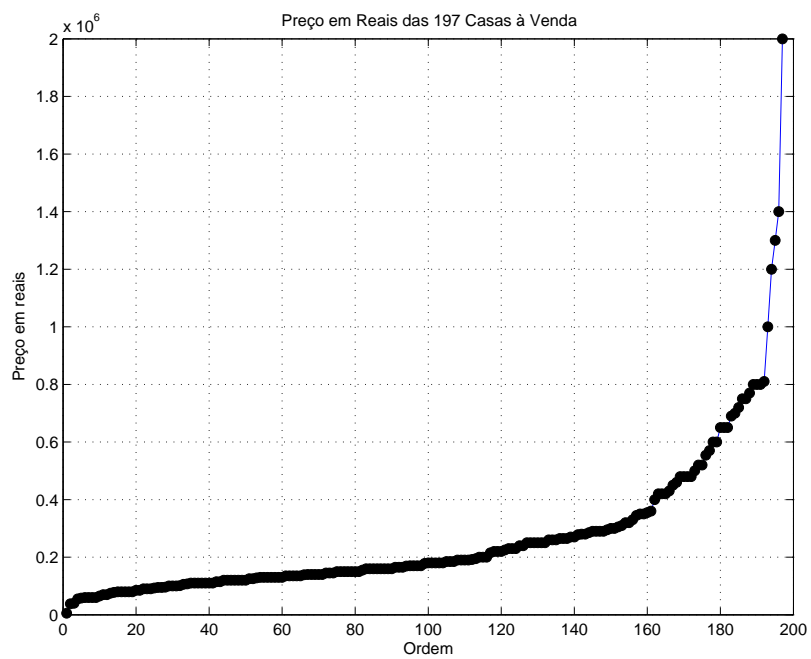


Figura 56: Preço das 197 Casas à Venda.

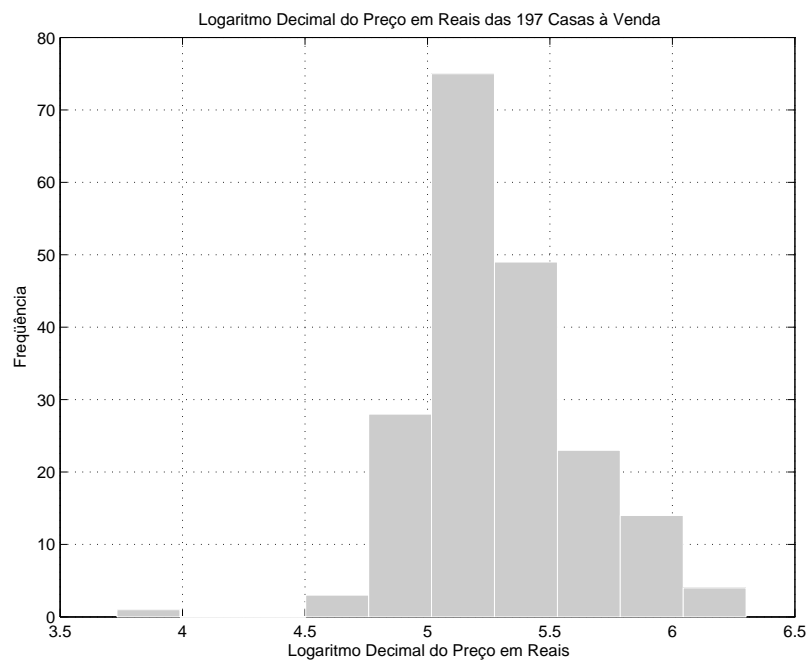


Figura 57: Histograma do Logaritmo Decimal do Preço em Reais das 197 Casas à Venda.

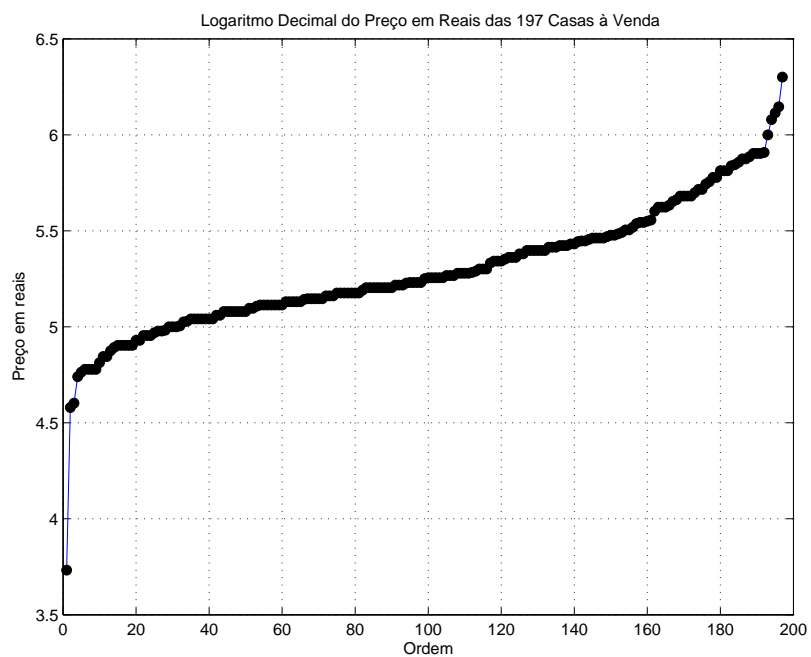


Figura 58: Logaritmo Decimal do Preço das 197 Casas à Venda.

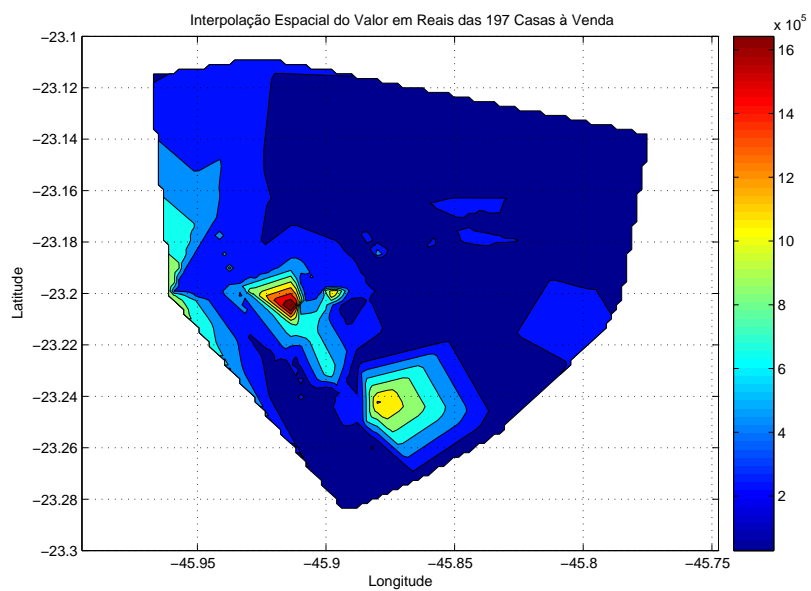


Figura 59: Valor do Imóvel em função da Latitude e da Longitude, para todas as 197 Casas à Venda.

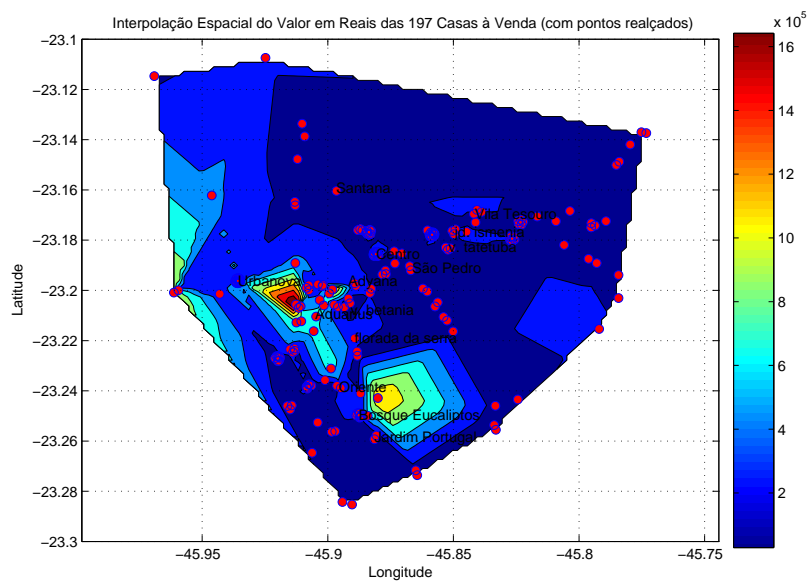


Figura 60: Valor do Imóvel em função da Latitude e da Longitude, para todas as 197 Casas à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

4.4.4 Preço em Reais das 24 Casas para Locação

Os gráficos das Figuras 61 a 64 demonstram uma distribuição homogênea dos preços dos imóveis, embora haja predominância das casas mais baratas, com preço em torno de R\$1500,00.

Os gráficos das Figuras 65 e 66 mostram que os imóveis de maior preço agrupam-se em torno do Jardim Colinas e do Jardim Esplanada.

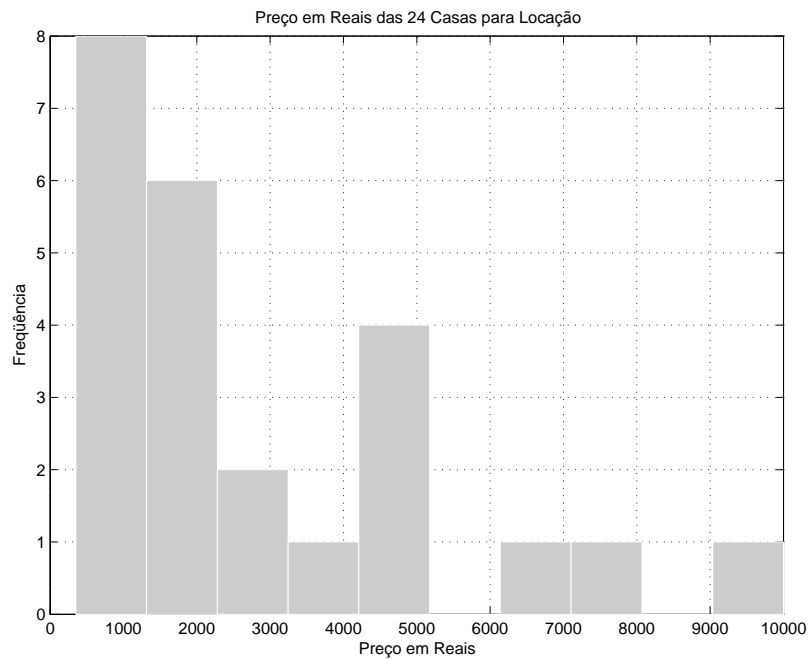


Figura 61: Histograma do Preço em Reais das 24 Casas para Locação.

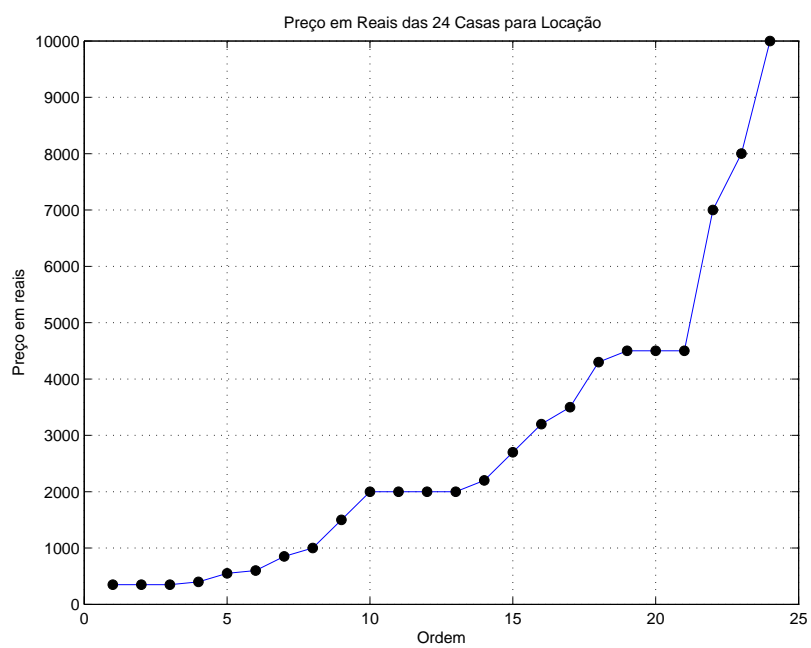


Figura 62: Preço das 24 Casas para Locação.

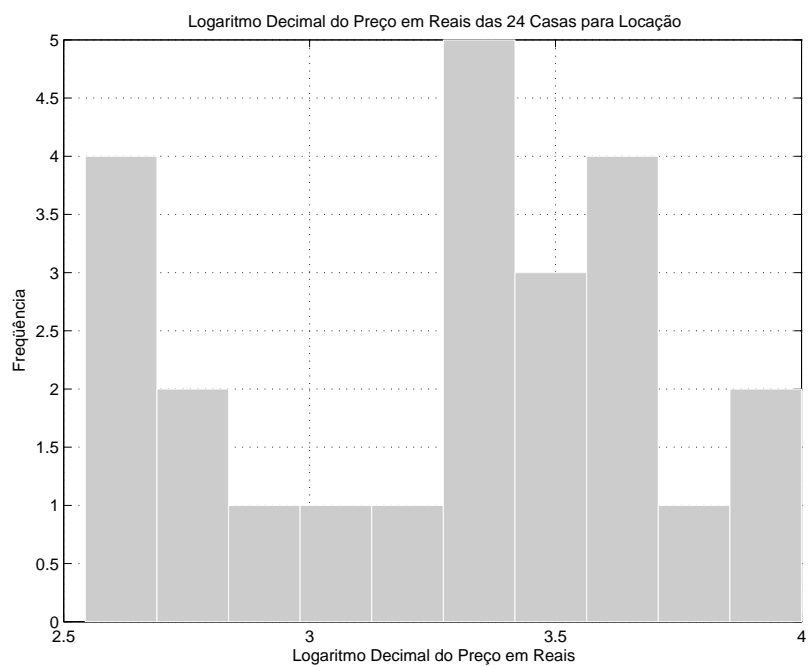


Figura 63: Histograma do Logaritmo Decimal do Preço em Reais das 24 Casas para Locação.

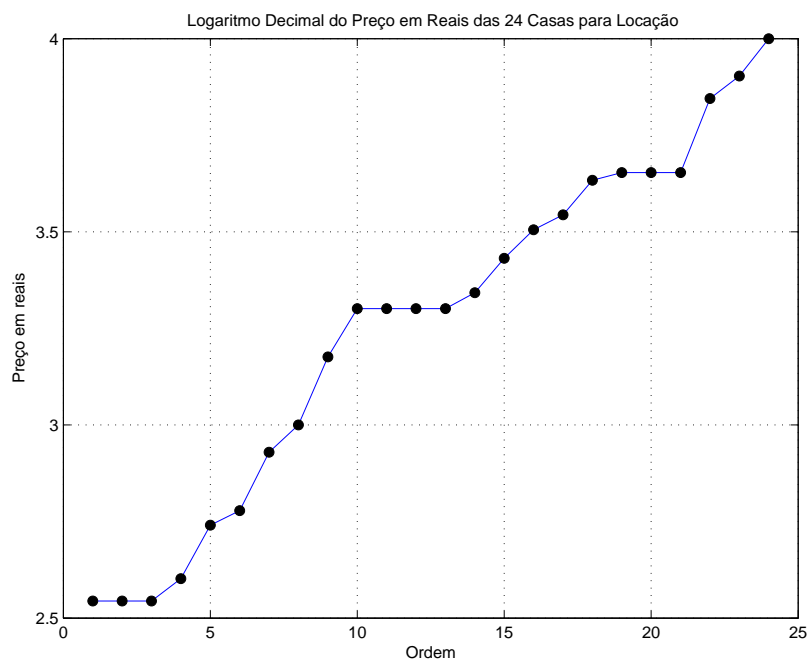


Figura 64: Logaritmo Decimal do Preço das 24 Casas para Locação.

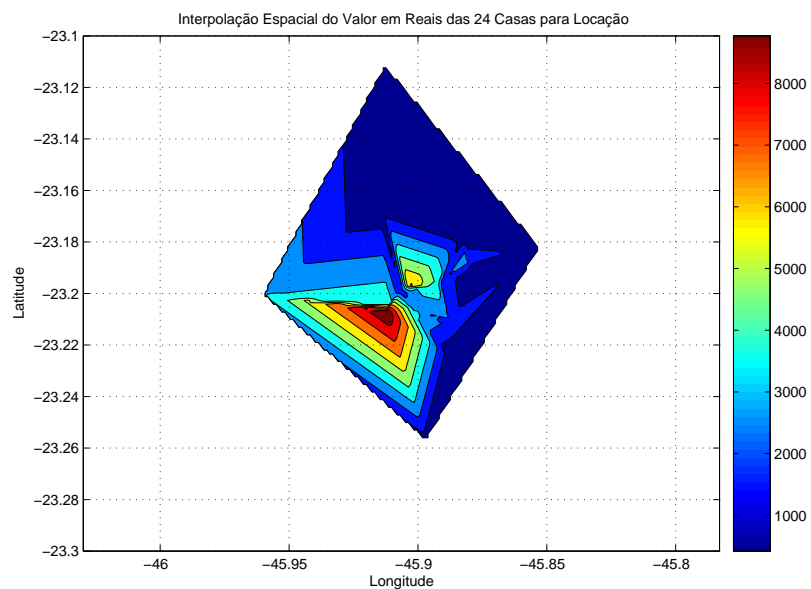


Figura 65: Valor do Imóvel em função da Latitude e da Longitude, para todas as 24 Casas para Locação.

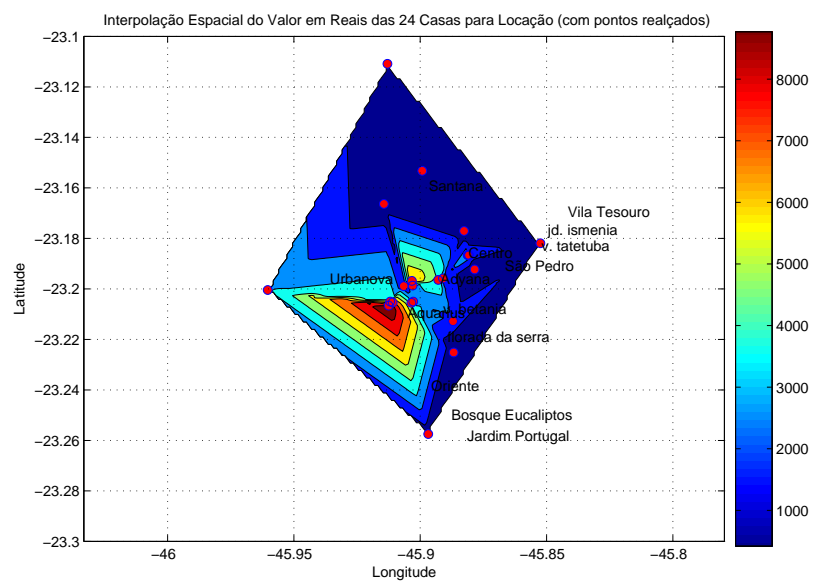


Figura 66: Valor do Imóvel em função da Latitude e da Longitude, para todas as 24 Casas para Locação, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

4.4.5 Preço em Reais dos 21 Terrenos à Venda

Os gráficos das Figuras 67 a 70 demonstram uma predominância de imóveis mais baratos, com preços abaixo de R\$100.000,00 embora haja um terreno *outlier* com preço acima de R\$1.300.000,00.

Os gráficos das Figuras 65 e 66 mostram que os imóveis de maior preço agrupam-se em torno do Jardim Alvorada e o vizinho Jardim das Indústrias.

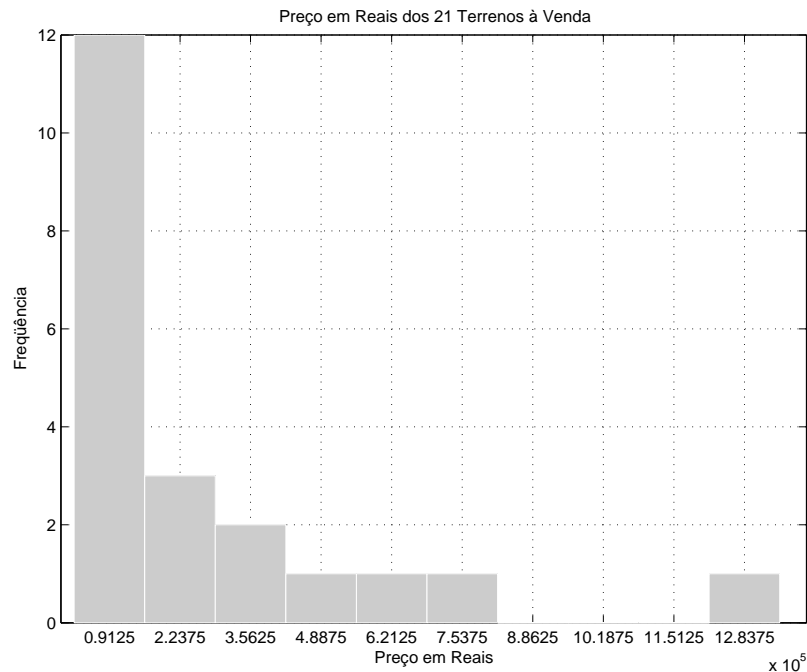


Figura 67: Histograma do Preço em Reais dos 21 Terrenos à Venda.

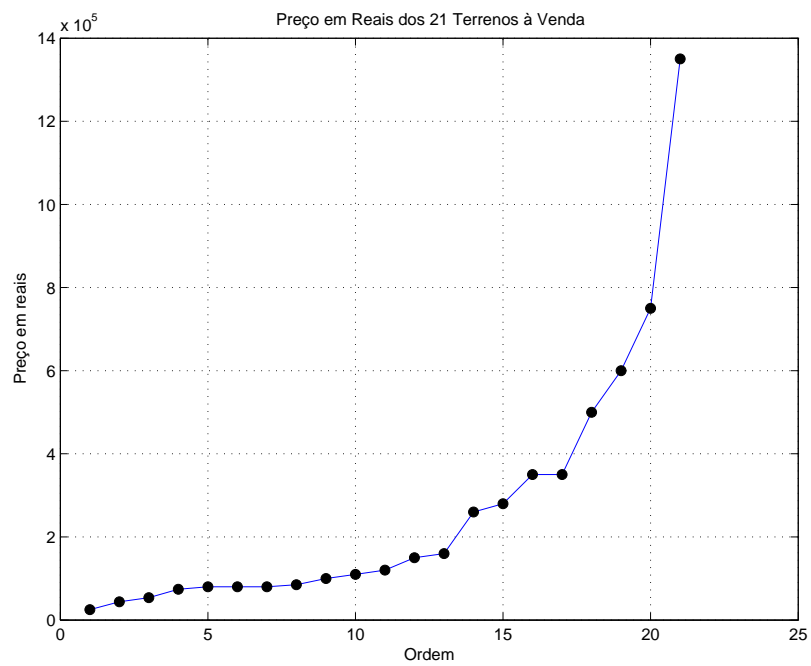


Figura 68: Preço dos 21 Terrenos à Venda.

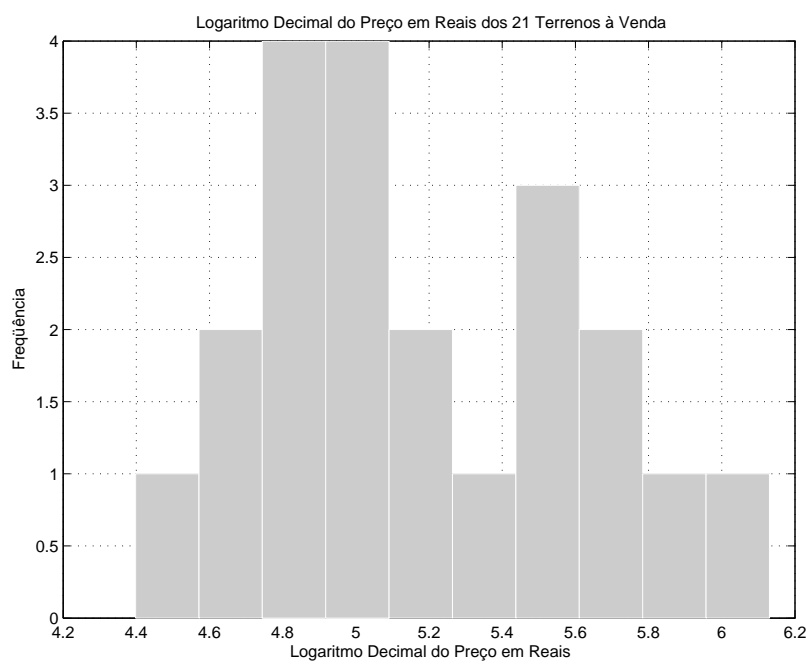


Figura 69: Histograma do Logaritmo Decimal do Preço em Reais dos 21 Terrenos à Venda.

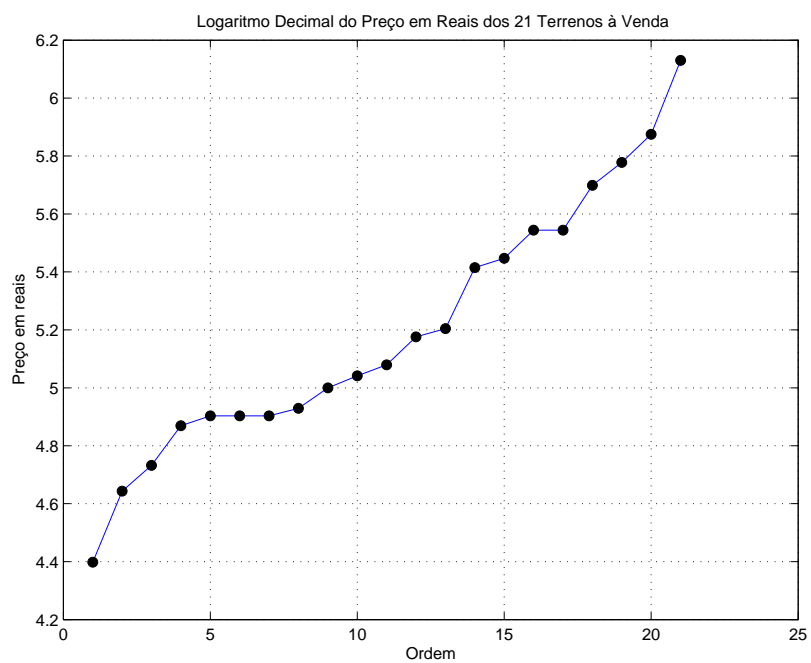


Figura 70: Logaritmo Decimal do Preço dos 21 Terrenos à Venda.

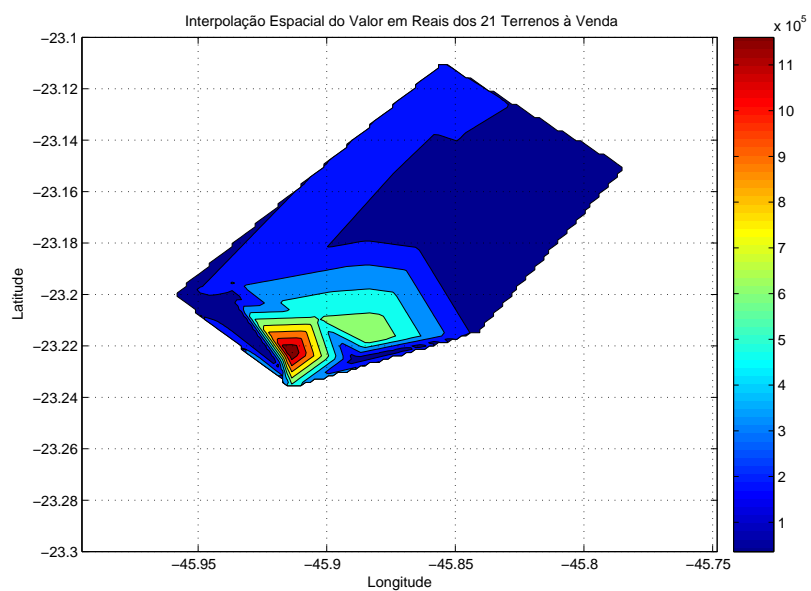


Figura 71: Valor do Imóvel em função da Latitude e da Longitude, para todos os 21 Terrenos à Venda.

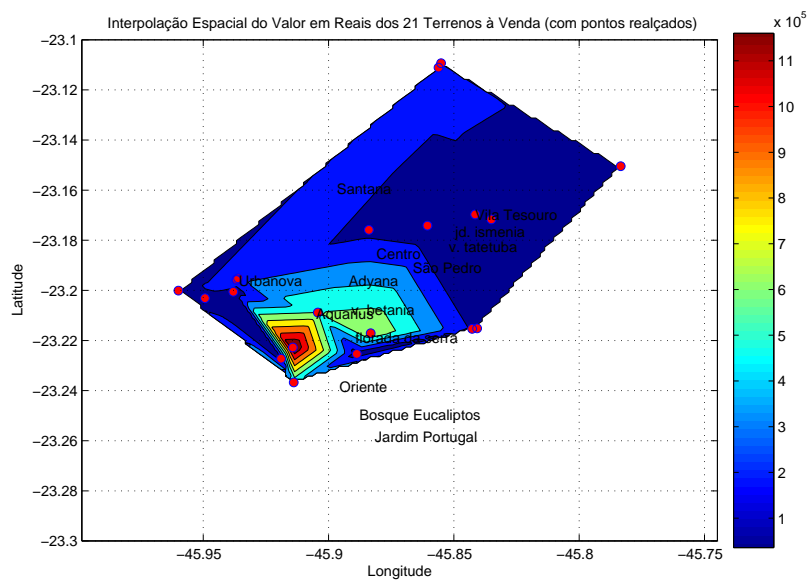


Figura 72: Valor do Imóvel em função da Latitude e da Longitude, para todos os 21 Terrenos à Venda, com realce no posicionamento destes imóveis e alguns bairros de São José dos Campos.

5 Regressão Múltipla

A partir dos gráficos da análise exploratória, definimos os seguintes modelos de regressão múltipla:

- i) $\text{Log Preço} = \beta_0 + \beta_1 * \text{Área}$
- ii) $\text{Log Preço} = \beta_0 + \beta_1 * \text{Área} + \beta_2 * \text{Longitude} + \beta_3 * \text{Latitude}$
- iii) $\text{Log Preço} = \beta_0 + \beta_1 * \text{Área} + \beta_2 * \text{Longitude} + \beta_3 * \text{Latitude} + \beta_4 * \text{Dormitórios} + \beta_5 * \text{Suítes} + \beta_6 * \text{Vagas}$

Para cada modelo, foi feita a regressão linear com cálculo do r^2 ajustado (que simbolizamos como r_{aj}^2) e validação cruzada com $k=3$.

5.1 Modelos de Estimativa de Preço para os 164 Apartamentos à Venda

Analisando os resultados obtidos (Tabela 5) notamos que:

- i) À medida em que adicionamos mais variáveis ao modelo, aumenta o valor do r_{aj}^2 e reduz-se o erro RMS.
- ii) O coeficiente da variável **Área**, apesar de positivo (conforme esperado) é praticamente zero indicando que a localização do imóvel (expressa pelas variáveis **Longitude** e **Latitude**) explica mais o preço do que a área.
- iii) Conforme esperado, os coeficientes das variáveis **Dormitórios**, **Suítes** e **Vagas** são positivos.

Tabela 5: Modelos de Estimativa de Preço para os 164 Apartamentos à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6667 | 0.0547 | 4.7636 | 0.0047 | - | - | - | - | - |
| 2 | 0.7255 | 0.0503 | -152.7722 | 0.0044 | -2.9975 | -0.8617 | - | - | - |
| 3 | 0.7701 | 0.0456 | -131.5545 | 0.0027 | -2.7476 | -0.4367 | 0.0494 | 0.0178 | 0.0812 |

Para comparar a importância de cada coeficiente do modelo 3 para a estimativa de preço, calculamos algumas estatísticas (Tabela 6).

Quando multiplicamos cada coeficiente pelo respectivo desvio padrão e, em seguida, operamos 10 elevado ao resultado, obtemos a coluna R\$ a qual mostra quantos reais a variação de um desvio padrão acrescenta ao preço final, mantidos constantes as demais variáveis.

Da análise da Tabela 5 notamos que:

- i) Embora haja uma grande variabilidade no preço, que varia R\$ 54.000 a R\$ 640.000 de modo exponencial, a utilização do LOG reduziu esta variação para o intervalo de 4.732 a 5.806, linearizando a variável dependente.
- ii) A variável que mais explica o preço final é a área do apartamento, seguida de perto pelo número de vagas na garagem. É interessante notar que, neste modelo, o número de vagas é mais importante que o número de suítes!

Tabela 6: Estatísticas para os 164 Apartamentos à Venda

| Variável | Mínimo | Média | Máximo | Desvio Padrão | Coefficiente | R\$ |
|-------------|---------|---------|---------|---------------|--------------|-------|
| Preço | 54000 | 184829 | 640000 | 102901 | - | - |
| Log Preço | 4.732 | 5.206 | 5.806 | 0.231 | - | - |
| Área | 41.000 | 94.878 | 300.000 | 40.567 | 0.003 | 1.287 |
| Longitude | -45.921 | -45.893 | -45.826 | 0.016 | -2.748 | 0.904 |
| Latitude | -23.287 | -23.208 | -23.151 | 0.023 | -0.437 | 0.977 |
| Dormitórios | 1.000 | 2.768 | 5.000 | 0.764 | 0.049 | 1.091 |
| Suítes | 0.000 | 0.445 | 2.000 | 0.534 | 0.018 | 1.022 |
| Vagas | 0.000 | 1.518 | 4.000 | 0.687 | 0.081 | 1.137 |

A Figura 73 mostra o gráfico obtido com a função `rcoplot` do Matlab. Neste gráfico, há 8 pontos outliers destacados em vermelho.

Após a remoção dos 8 outliers, repetimos o procedimento de regressão múltipla, obtendo os dados da Tabela 7.

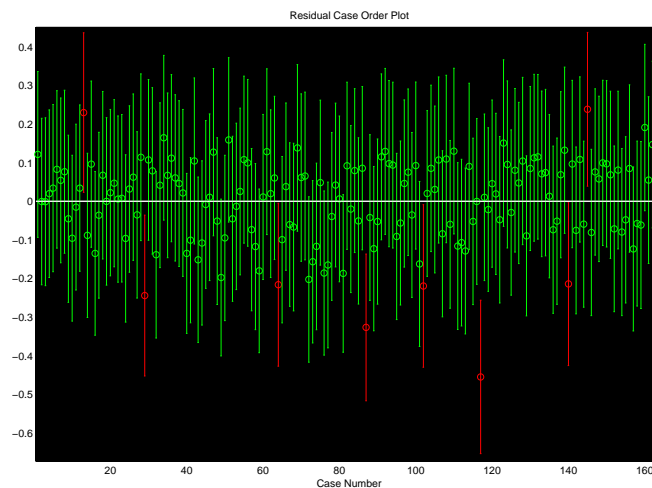


Figura 73: Resíduos calculados para o Modelo 3. Estão destacados em vermelho os *outliers*.

Tabela 7: Modelos de Estimativa de Preço para 156 Apartamentos à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6788 | 0.0518 | 4.7515 | 0.0049 | - | - | - | - | - |
| 2 | 0.7474 | 0.0468 | -159.1996 | 0.0045 | -3.2200 | -0.6984 | - | - | - |
| 3 | 0.8354 | 0.0388 | -120.5686 | 0.0024 | -2.5901 | -0.2743 | 0.0417 | 0.0000 | 0.1342 |

Analisando os resultados obtidos (Tabela 7) notamos que:

- Conforme esperado, os resultados são bem melhores que os obtidos com os *outliers* mostrados na Tabela 7.
- O coeficiente da variável **Suítes** é estatisticamente zero. Logo, podemos remover esta variável do modelo 3.

É interessante notar que, mesmo após retirarmos 8 *outliers*, o gráfico de resíduos ainda mostra outros 5 *outliers* (Figura 74).

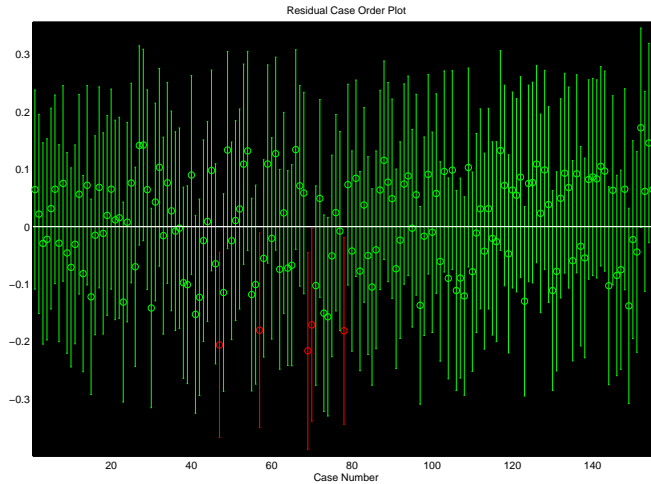


Figura 74: Resíduos calculados para o Modelo 3 com 156 Apartamentos à Venda. Estão destacados em vermelho os *outliers*.

5.2 Modelos de Estimativa de Preço para os 25 Apartamentos para Locação

Analisando os resultados obtidos (Tabela 8) notamos que:

- À medida em que adicionamos mais variáveis ao modelo, aumenta o valor do r_{aj}^2 e reduz-se o erro RMS. No entanto, os valores de r_{aj}^2 e RMS são sempre piores que os obtidos para o modelo de Apartamentos à Venda. Provavelmente, isto ocorre porque temos poucos apartamentos para locação com preços que variam muito.

- ii) De forma análoga ao que ocorreu no modelo para Apartamentos à Venda, o coeficiente da variável **Área**, apesar de positivo (conforme esperado) é praticamente zero.
- iii) Ao contrário do esperado, os coeficientes das variáveis **Dormitórios**, **Suítes** e **Vagas** são bem próximos de zero e o coeficiente de **Suítes** é negativo!

Tabela 8: Modelos de Estimativa de Preço para os 25 Apartamentos para Locação

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.5991 | 0.1273 | 2.5909 | 0.0043 | - | - | - | - | - |
| 2 | 0.6424 | 0.1180 | -160.8097 | 0.0041 | -2.6727 | -1.7565 | - | - | - |
| 3 | 0.6443 | 0.1711 | -46.5552 | 0.0033 | -0.0667 | -1.9841 | 0.0021 | -0.0183 | 0.1007 |

Após a remoção de 2 outliers, obtemos os dados da Tabela 9

Tabela 9: Modelos de Estimativa de Preço para os 23 Apartamentos para Locação

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.7208 | 0.1143 | 2.5998 | 0.0047 | - | - | - | - | - |
| 2 | 0.7169 | 0.1214 | -74.9158 | 0.0046 | -1.0290 | -1.3045 | - | - | - |
| 3 | 0.7345 | 0.1160 | -49.7036 | 0.0060 | -0.6385 | -0.9905 | -0.0865 | -0.0084 | 0.0598 |

5.3 Modelos de Estimativa de Preço para as 197 Casas à Venda

Analisando os resultados obtidos (Tabela 10) notamos que:

- i) À medida em que adicionamos mais variáveis ao modelo, aumenta o valor do r_{aj}^2 e reduz-se o erro RMS. Porém, os resultados foram ainda piores que os obtidos para o modelo de Apartamentos para Locação.
- ii) Repetiu-se o fato de o coeficiente da variável **Área** ser praticamente zero e o coeficiente da variável **Suíte** ser negativo.

Tabela 10: Modelos de Estimativa de Preço para as 197 Casas à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.5283 | 0.0829 | 4.8609 | 0.0023 | - | - | - | - | - |
| 2 | 0.6064 | 0.0759 | -93.0893 | 0.0020 | -2.5091 | 0.7374 | - | - | - |
| 3 | 0.6355 | 0.0737 | -82.7670 | 0.0018 | -2.3878 | 0.9508 | 0.0701 | -0.0199 | 0.0105 |

Após a remoção de 7 outliers, obtemos os dados da Tabela 11

Tabela 11: Modelos de Estimativa de Preço para as 190 Casas à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6519 | 0.0718 | 4.8706 | 0.0023 | - | - | - | - | - |
| 2 | 0.7593 | 0.0577 | -106.3714 | 0.0020 | -2.7363 | 0.6135 | - | - | - |
| 3 | 0.8012 | 0.0531 | -98.9198 | 0.0017 | -2.7318 | 0.8986 | -0.0704 | -0.0343 | 0.0142 |

5.4 Modelos de Estimativa de Preço para as 24 Casas para Locação

Analisando os resultados obtidos (Tabela 12) notamos que:

- Ao contrário do esperado, o modelo mais simples, que correlaciona preço com área tem melhores valores de r_{aj}^2 e RMS que os modelos mais sofisticados.
- Confirmou-se a tendência, já vista nos modelos anteriores, de o coeficiente da variável **Área**, apesar de positivo, ser bem próximo de zero.
- Conforme esperado, os coeficientes das variáveis **Dormitórios**, **Suítes** e **Vagas** são positivos.

Tabela 12: Modelos de Estimativa de Preço para as 24 Casas para Locação

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6024 | 0.3336 | 2.7633 | 0.0022 | - | - | - | - | - |
| 2 | 0.5899 | 0.3736 | -111.3102 | 0.0021 | -1.1208 | -2.7020 | - | - | - |
| 3 | 0.6001 | 0.3799 | -80.0515 | 0.0018 | -0.6887 | -2.2065 | 0.0103 | 0.2445 | 0.0260 |

Após a remoção de 7 outliers, obtemos os dados da Tabela 13

Tabela 13: Modelos de Estimativa de Preço para 23 Casas para Locação

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6621 | 0.3101 | 2.7227 | 0.0023 | - | - | - | - | - |
| 2 | 0.6505 | 0.4136 | -108.7978 | 0.0021 | -1.1892 | -2.4566 | - | - | - |
| 3 | 0.6810 | 0.3099 | -78.2315 | 0.0022 | -0.7793 | -1.9503 | -0.0336 | 0.03350 | 0.0102 |

5.5 Modelos de Estimativa de Preço para os 21 Terrenos à Venda

Analisando os resultados obtidos (Tabela 14) notamos que:

- Conforme esperado, a adição das variáveis **Dormitórios**, **Suítes** e **Vagas** não melhorou os valores de r_{aj}^2 e RMS.

- ii) O coeficiente da variável **Área** foi o mais baixo dos modelos fazendo crer que, com os dados analisados, a localização explica bem mais o preço do que a área do terreno.

Tabela 14: Modelos de Estimativa de Preço para os 21 Terrenos à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.5282 | 0.3493 | 4.8018 | 0.0005 | - | - | - | - | - |
| 2 | 0.6415 | 0.3568 | -164.1786 | 0.0005 | -2.9700 | -1.4101 | - | - | - |
| 3 | 0.5647 | 0.3543 | -164.1786 | 0.0005 | -2.9700 | -1.4101 | 0 | 0 | 0 |

Após a remoção de 2 outliers, obtemos os dados da Tabela 15

Tabela 15: Modelos de Estimativa de Preço para 19 Terrenos à Venda

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|--------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.6310 | 0.3042 | 4.7706 | 0.0005 | - | - | - | - | - |
| 2 | 0.7515 | 0.2913 | -131.9598 | 0.0005 | -2.8272 | -0.3018 | - | - | - |
| 3 | 0.6893 | 0.2579 | -131.9598 | 0.0005 | -2.8272 | -0.3018 | 0 | 0 | 0 |

6 Conclusão

Considerando o critério do maior r_{aj}^2 e do menor RMS os melhores modelos analisados são descritos na Tabela 16.

Tabela 16: Melhores Equações para Cada Modelo

| Modelo | r_{aj}^2 | RMS | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 |
|---------------|------------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Apto. Venda | 0.8351 | 0.0388 | -120.5686 | 0.0024 | -2.5901 | -0.2743 | 0.0417 | - | 0.1342 |
| Apto. Locação | 0.7345 | 0.1160 | -49.7036 | 0.0060 | -0.6385 | -0.9905 | -0.0865 | -0.0084 | 0.0598 |
| Casa Venda | 0.8012 | 0.0531 | -98.9198 | 0.0017 | -2.7318 | 0.8986 | 0.0704 | -0.0343 | 0.0142 |
| Casa Locação | 0.6810 | 0.3099 | -78.2315 | 0.0022 | -0.7793 | -1.9503 | -0.0336 | 0.03350 | 0.0102 |
| Terrenos | 0.7515 | 0.2913 | -131.9598 | 0.0005 | -2.8272 | -0.3018 | - | - | - |

Acreditamos ter obtidos bons resultados com valores de r_{aj}^2 entre 0.6810 e 0.8351, com RMS variando entre 0.0388 e 0.3099.

7 Referências Bibliográficas

7.1 Sites

- [1] www.redesaojose.com.br, acessado em 09 de abril de 2008 e 04 de maio de 2008.
- [2] maps.google.com, acessado em 20 de maio de 2008.

7.2 Papers

- [3] CASSUNDÉ Jr., Nilton. *Um estudo sobre as principais variáveis influenciadoras da composição de preço do mercado imobiliário: uma análise sobre o bairro de Boa Viagem - Recife/PE*. VII Seminários em Administração FEA-USP. 2004.
- [4] CHOPRA, Sumit et. alli. *Discovering the hidden structure of house prices with a non-parametric latent manifold model*. KDD 2007: 173-182. 2007.
- [5] GRESS, Bernard. *Using semi-parametric spatial autocorrelation models to improve hedonic housing price prediction*. Economics, University of California, Riverside, CA. 2004.
- [6] GUEDES, Jackson Carvalho. *Aplicação de redes neurais na avaliação de bens*. Petrobrás. 2008.
- [7] HARRISON, Jr., David, RUBINFELD, Daniel L. *Hedonic housing prices and the demand for clean air*. Journal of Environmental Economics and Management 5(1): 81-102. 1978.