

SISTEMA AGREGADOR PARA ANÁLISE DE CARACTERÍSTICAS DE IMÓVEIS COM RASTREAMENTO WEB E APRENDIZADO DE MÁQUINA

Felipe Israel Corrêa

Trabalho de Conclusão do Curso de Engenharia da Computação

Orientador: **PROF. Dr. Renato Dourado Maia**

Organização

- 1 - Introdução
- 2 - Revisão Literária
- 3 - Materiais e Métodos
- 4 - Resultados e Discussão
- 5 - Considerações Finais
- 6 - Referências

1 - Introdução

O constante crescimento dos centros urbanos tem ocasionado, proporcionalmente, o aumento por uma moradia, fato importante para empresas do ramo imobiliário.

Atualmente grande parte das ofertas de imóveis é feita por meio de *sites*, gerando um imenso volume de dados digitais.

Problema

- Inúmeras opções de sites;
- Informações disponibilizadas;

Objetivo

- Utilização de Rastreador *Web* e Aprendizagem de Máquina, para criar um sistema capaz de agregar os dados fornecidos pelas imobiliárias e adicionar uma gama de informações mais detalhadas do que a atualmente obtida pelos usuários.
- Possibilitar ao usuário informações não somente sobre o imóvel, mas também sobre o bairro e a cidade;

Justificativa

- Previsto um retorno de crescimento em 2018, para este ramo, ocasionado pela baixa da taxa SELIC e novas regras do programa Minha Casa Minha Vida (DINO, 2017);
- Criado por meio de ferramentas de análise e programação gratuitas;

2 - Revisão Literária

- Conceitos importantes:
 - Rastreador *Web*;
 - Aprendizado de Máquina;

Rastreamento *Web*

- Consiste em coletar automaticamente os dados de páginas web, extrair as informações específicas e armazená-las para uso posterior (BERNARD, 2017);
- *Tags* disponíveis nos códigos-fonte das páginas HTML e XML (HEMENWAY; CALISHAIN, 2003);
- Fluxo de busca:

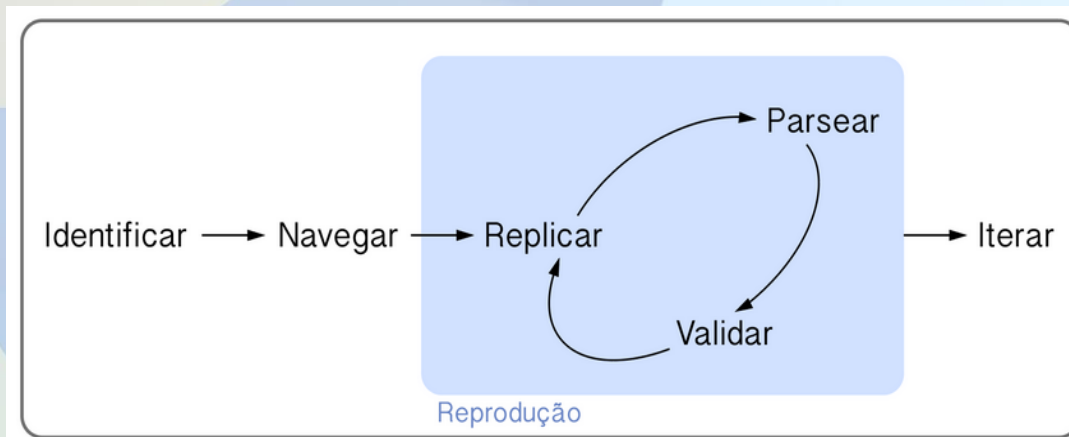


FIGURA 1: Diagrama de fluxo de um rastreador web.
Fonte: Lente (2018)

Aprendizado de Máquina

- Melhorar a predição de modelos e tomar a decisão baseada nos dados, por meio do conhecimento adquirido (RASCHKA; MIRJALILI, 2017);
- Aplicações;
- Tarefas;
- Tipo de aprendizagem;

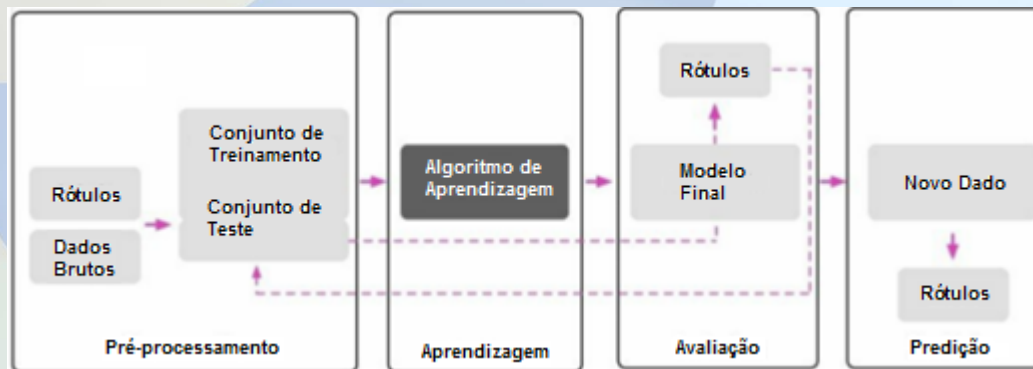


FIGURA 2: Fluxo de operações de um sistema de AM.

Fonte: Raschka e Mirjalili (2017, p. 53 – Adaptada)

3 - Materiais e Métodos

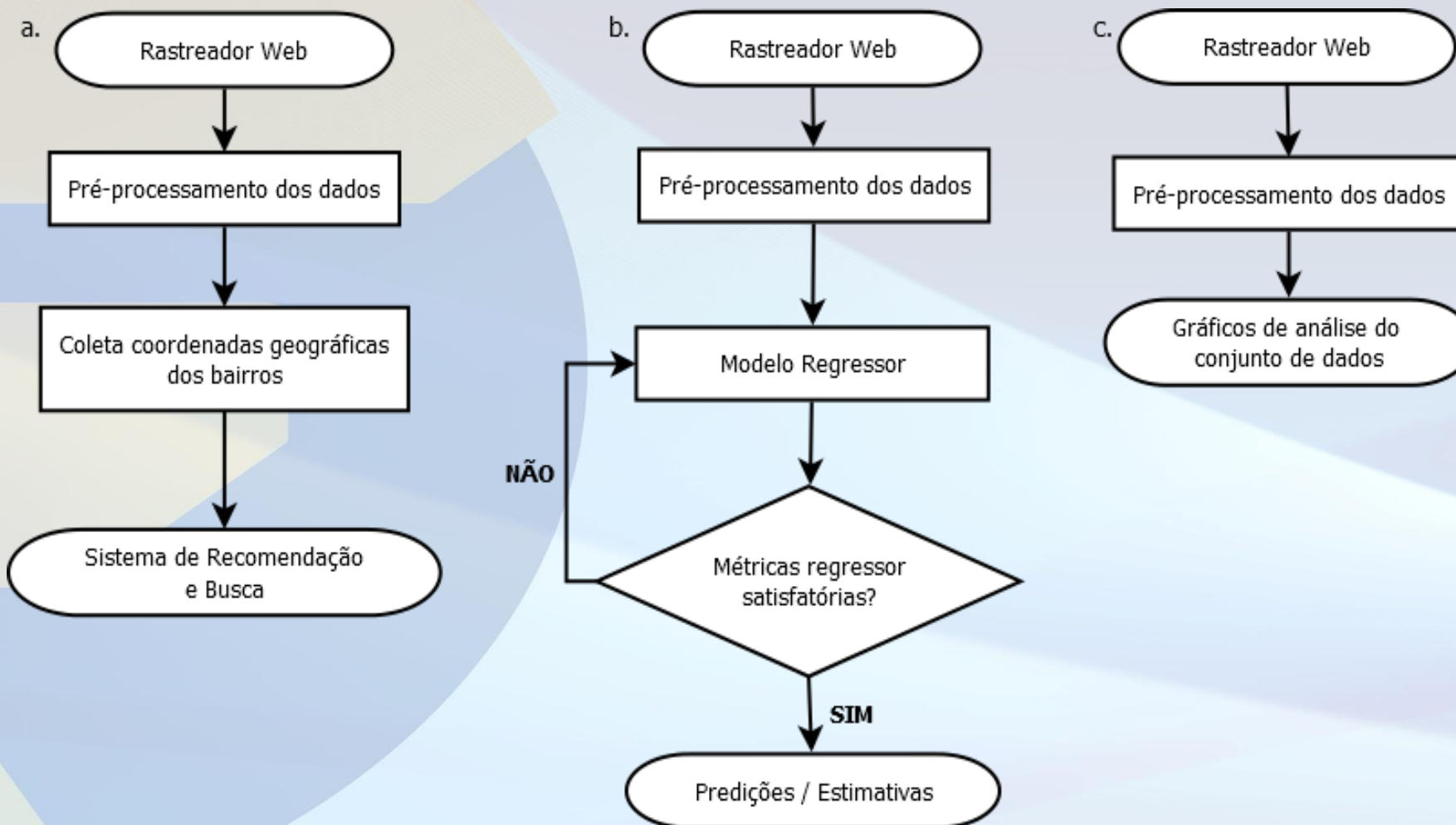
Métodos

- Etapas:
 - Linguagem de programação e bibliotecas;
 - Rastreador *Web* e obtenção dos dados;
 - Armazenamento dos dados obtidos;
 - Criação dos modelos de aprendizado de máquina;
 - Análise exploratória dos dados;

Materiais

- Ambiente de desenvolvimento:
 - *Visual Studio Code;*
- Linguagem:
 - *Python;*
- Bibliotecas:
 - *Requests*
 - *BeautifulSoup;*
 - *Mysql.connector;*
- *Googlemaps Geocoding;*
- *Scikit-learn;*
- *Numpy;*
- *Pandas;*
- *Matplotlib;*
- *Bokeh;*
- Questionário:
 - *Google Forms;*

4 - Resultados e Discussão



Rastreador Web

TABELA 1

Características dos imóveis

ID	DIA	BAIRRO	QUARTOS	BANHEIROS	GARAGENS	AREA	VALOR
1	2018-04-02	Augusta Mota	3	1	2	90	320000
2	2018-04-02	Augusta Mota	2	1	1	85	300000
3	2018-04-02	Ibituruna	3	2	2	0	287500
4	2018-04-02	Canelas	2	1	1	62.5	198000
5	2018-04-02	São José	4	3	1	0	285000

Fonte: O Autor

Regressor

- Pré-processamento;

TABELA 2

Conjunto de dados após pré-processamento

ID	DIA	BAIRRO	QUARTOS	BANHEIROS	GARAGENS	AREA	VALOR
1	2018-02-04	Planalto	2	1	1	48	333075.79
2	2018-02-04	Todos Os Santos	2	2	0	51.215	333075.79
3	2018-02-04	Cidade Santa Maria	3	2	2	120	480000
4	2018-02-04	Sao Jose	4	3	1	90.752	285000
5	2018-02-04	Centro	4	5	2	149.838	750000
6	2018-02-04	Morada Do Sol	3	1	2	155.432	370000

Fonte: O Autor

Regressor

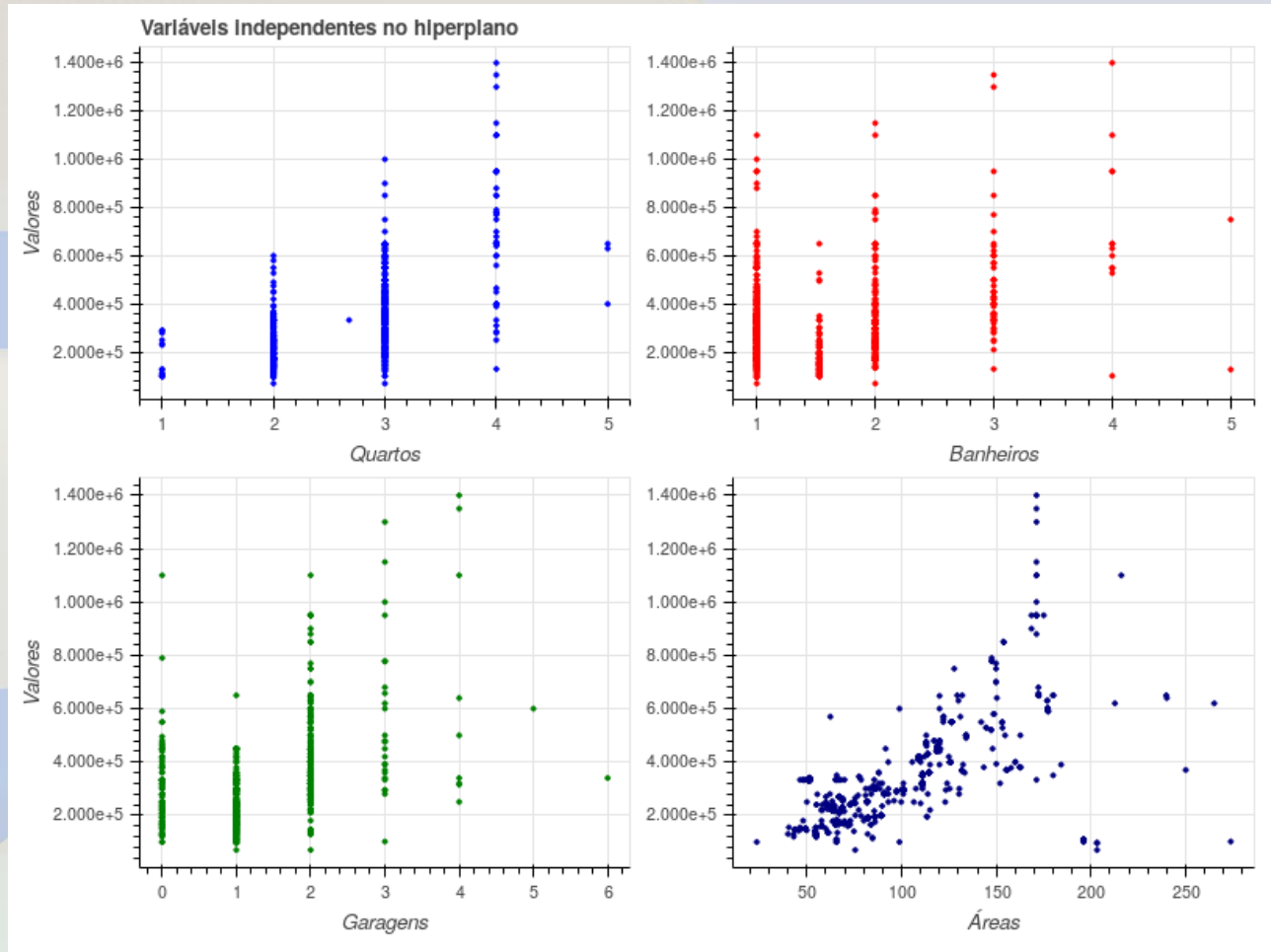


FIGURA 4: Variáveis independentes no hiperplano.
Fonte: O Autor

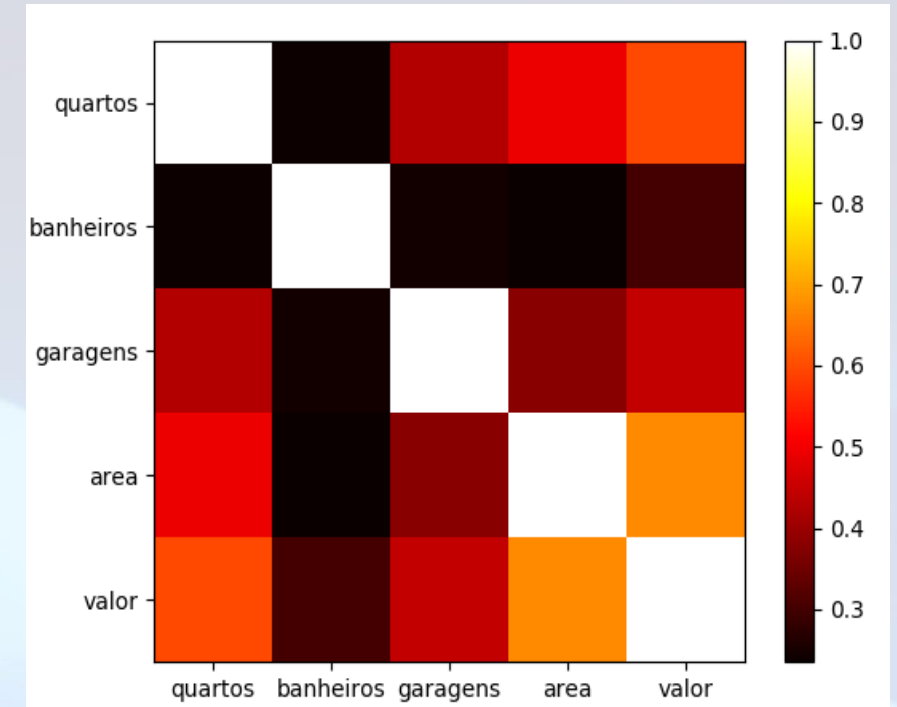


FIGURA 5: Matriz de correlação.
Fonte: O Autor

Regressor

TABELA 3

Resultado do cálculo das métricas do modelo linear

Média Validação Cruzada	Desvio-Padrão da Validação Cruzada	Média do Erro Absoluto (R\$)	R ²
0,6389	0,0916	67.268,09	0,4527

Fonte: O Autor

TABELA 4

Resultado das métricas da árvore de decisão de acordo com sua profundidade

Profundidade	Média Validação Cruzada	Desvio-Padrão da Validação Cruzada	Média do Erro Absoluto (R\$)	R ²
5	0,7683	0,1230	51.808,86	0,5145
10	0,7623	0,1556	37.086,16	0,5738
15	0,7769	0,1398	35.937,07	0,5805
20	0,7797	0,1308	32.791,86	0,6265
25	0,7714	0,1506	36.885,14	0,5408

Fonte: O Autor

TABELA 5

Resultado das métricas da floresta aleatória de acordo com a quantidade de árvores e profundidade

Árvores / Profundidade	Média Validação Cruzada	Desvio-Padrão da Validação Cruzada	Média do Erro Absoluto (R\$)	R ²
5 / 5	0,8048	0,1071	51.561,28	0,6585
5 / 10	0,8358	0,8300	34.714,23	0,7047
5 / 15	0,8491	0,0747	36.565,53	0,6918
5 / 20	0,8405	0,0813	34.840,55	0,6929
5 / 25	0,8442	0,0797	35.548,32	0,7051
20 / 20	0,8545	0,0766	30.885,83	0,7714
20 / 25	0,8560	0,0734	31.539,07	0,7711
25 / 5	0,8253	0,0859	43.852,90	0,7406
25 / 10	0,8583	0,0673	31.711,47	0,7657
25 / 15	0,8610	0,0681	30.933,56	0,7945
25 / 20	0,8577	0,0749	30.708,28	0,7711
25 / 25	0,8568	0,0737	30.828,96	0,7860

Fonte: O Autor

Regressor

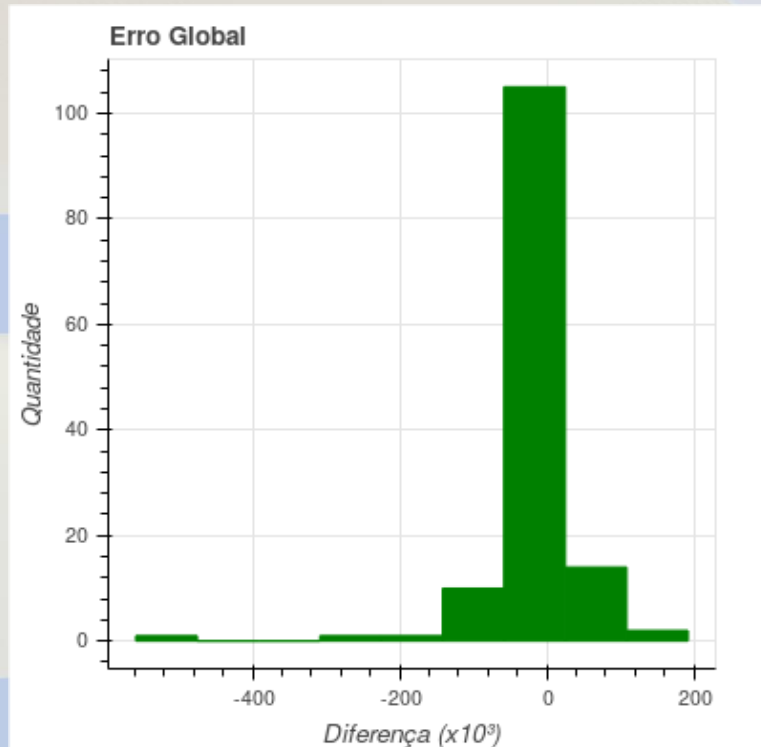


FIGURA 6: Erro global.
Fonte: O Autor

TABELA 6
Média das características dos apartamentos com base na quantidade de quartos

Quartos	Média Banheiros	Média Garagens	Média Área (m²)
1	1,28	0,53	70,43
2	1,39	1,02	73,40
3	1,53	1,54	104,85
4	2,20	2,34	150,68
5	2,66	2	128,37

Fonte: O Autor

Regressor

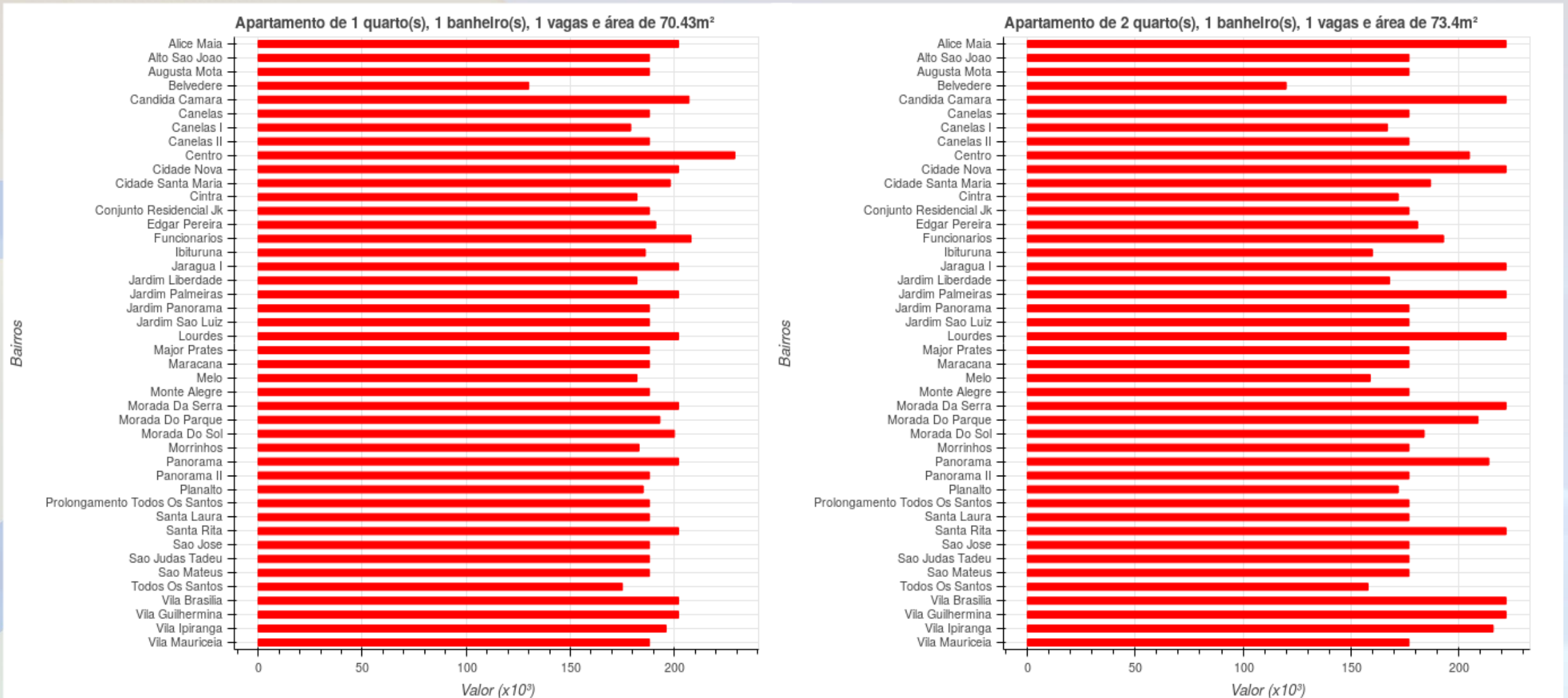


FIGURA 7: Estimativas de preços para apartamentos de 1 e 2 quartos.
Fonte: O Autor

Regressor

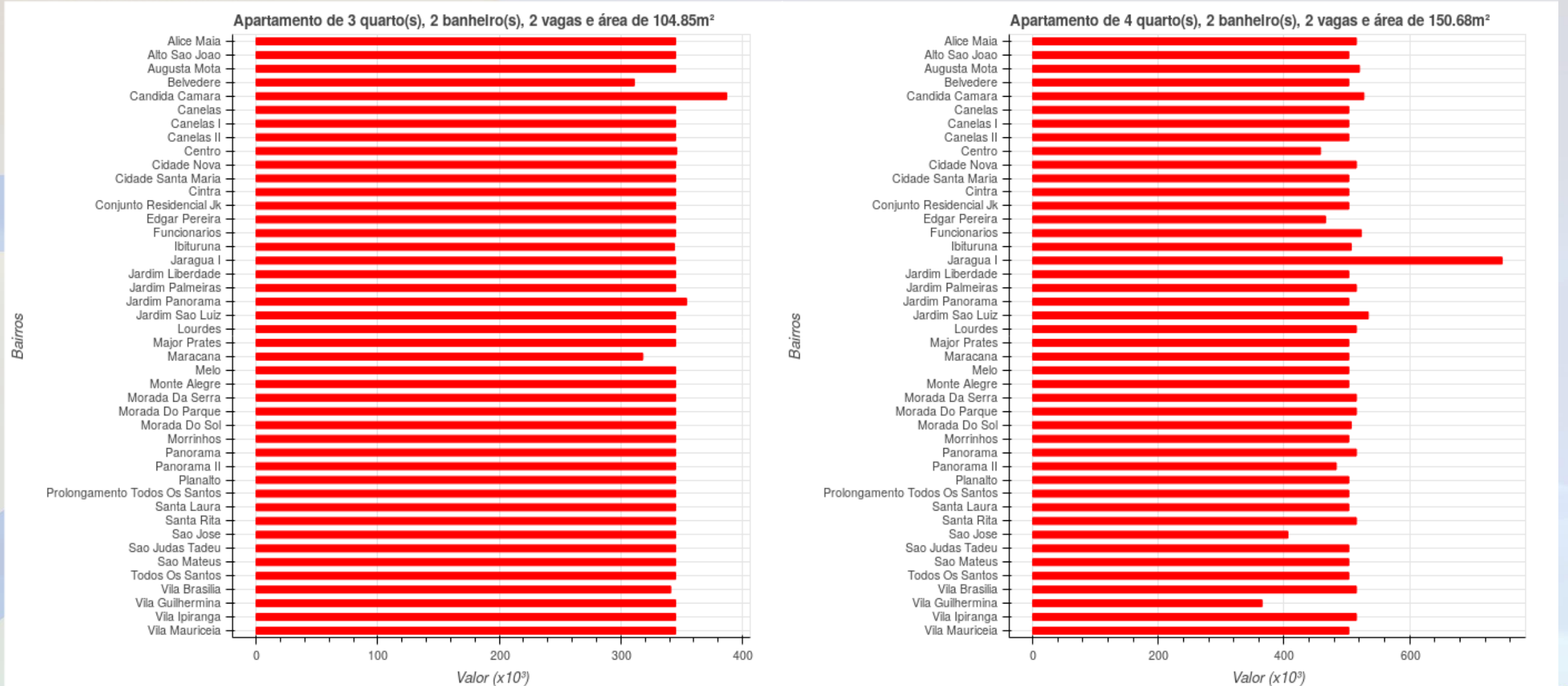


FIGURA 8: Estimativas de preços para apartamentos de 3 e 4 quartos.
Fonte: O Autor

Regressor

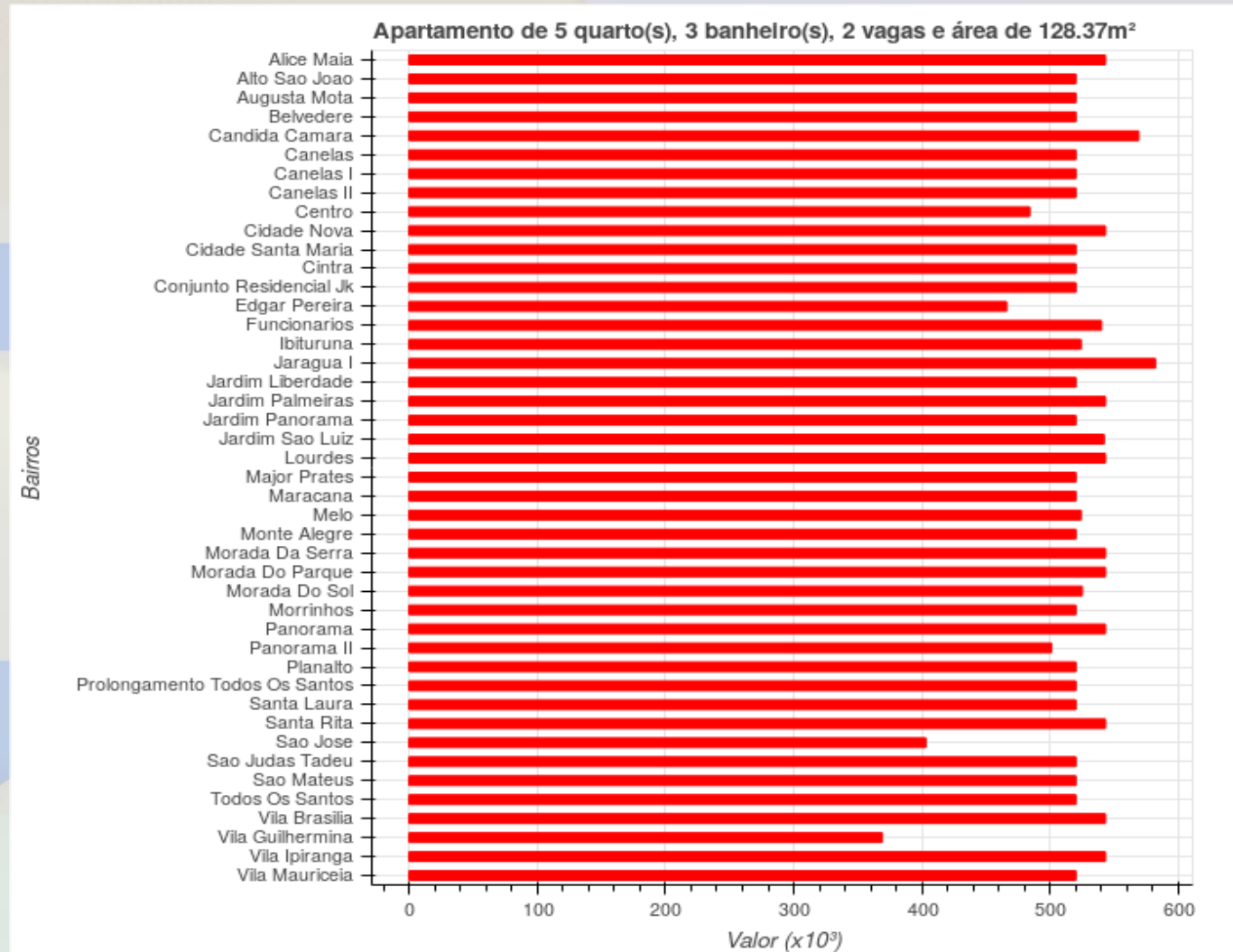


FIGURA 9: Estimativas de preços para apartamentos de 5 quartos.

Fonte: O Autor

Primeiro Imóvel

Apartamentos encontrados:

Bairro Centro com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 200000.0.
Bairro Centro com 2 quarto(s), 1 banheiro(s), 0 vaga(s) de garagem, área de 0.0m² e valor de R\$ 225000.0.
Bairro Centro com 2 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 64.4m² e valor de R\$ 230000.0.
Bairro Centro com 2 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 60.65m² e valor de R\$ 230000.0.
Bairro Centro com 3 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 230000.0.

Imóveis recomendados:

Bairro Jardim Panorama com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 75.53m² e valor de R\$ 200000.0.
Bairro Todos Os Santos com 3 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 200000.0.
Bairro Canelas I com 3 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 200000.0.
Bairro Morada Do Parque com 3 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 200000.0.
Bairro Canelas com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 0.0m² e valor de R\$ 200000.0.

Primeiro Imóvel

Apartamentos encontrados:

Bairro Centro com 2 quarto(s), 2 banheiro(s), 0 vaga(s) de garagem, área de 89.74m² e valor de R\$ 235000.0.
Bairro Centro com 3 quarto(s), 1 banheiro(s), 0 vaga(s) de garagem, área de 99.75m² e valor de R\$ 258000.0.
Bairro Centro com 2 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 64.4m² e valor de R\$ 230000.0.
Bairro Centro com 3 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 180.8m² e valor de R\$ 298000.0.
Bairro Centro com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 64.4m² e valor de R\$ 235000.0.

Imóveis recomendados:

Bairro Candida Canara com 3 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 77.35m² e valor de R\$ 220000.0.
Bairro Jardim Panorama com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 75.53m² e valor de R\$ 288000.0.
Bairro Prolongamento Todos Os Santos com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 69.75m² e valor de R\$ 215000.0.
Bairro São José com 2 quarto(s), 1 banheiro(s), 1 vaga(s) de garagem, área de 57.72m² e valor de R\$ 180000.0.
Bairro Jardim Liberdade com 3 quarto(s), 2 banheiro(s), 1 vaga(s) de garagem, área de 78.59m² e valor de R\$ 198100.0.

FIGURA 10: Busca e recomendação métricas euclidiana e do cosseno..

Fonte: O Autor

<input type="checkbox"/> Ibituruna, 3 quarto(s), 2 banheiro(s), 1 vaga(s) e valor de R\$ 230000.	<input type="checkbox"/> Cândida Câmara, 3 quarto(s), 2 banheiro(s), 1 vaga(s) e valor de R\$ 220000.
<input type="checkbox"/> Edgar Pereira, 3 quarto(s), 2 banheiro(s), 1 vaga(s) e valor de R\$ 230000.	<input type="checkbox"/> Prolongamento Todos Os Santos, 2 quarto(s), 1 banheiro(s), 1 vaga(s) e valor de R\$ 215000.
<input type="checkbox"/> Morada Do Sol, 2 quarto(s), 1 banheiro(s), 0 vaga(s) e valor de R\$ 230000.	<input type="checkbox"/> Ibituruna, 2 quarto(s), 2 banheiro(s), 2 vaga(s) e valor de R\$ 250000.
<input type="checkbox"/> Cidade Nova, 2 quarto(s), 1 banheiro(s), 1 vaga(s) e valor de R\$ 230000.	<input type="checkbox"/> Cidade Nova, 3 quarto(s), 1 banheiro(s), 2 vaga(s) e valor de R\$ 255000.
<input type="checkbox"/> Funcionários, 2 quarto(s), 1 banheiro(s), 1 vaga(s) e valor de R\$ 230000.	<input type="checkbox"/> Ibituruna, 3 quarto(s), 2 banheiro(s), 2 vaga(s) e valor de R\$ 280000.

FIGURA 11: Imóveis apresentados aos usuários calculados por métricas diferentes.
Fonte: O Autor

TABELA 7

Resultado cálculos de eficiência das recomendações

	Média F1-score	Média Precisão	Média Recall
Dissimilaridade Euclidiana	0,2322	0,2090	0,2613
Similaridade do Cosseno	0,1959	0,1566	0,2613

Fonte: O Autor

Análise Exploratória dos Dados

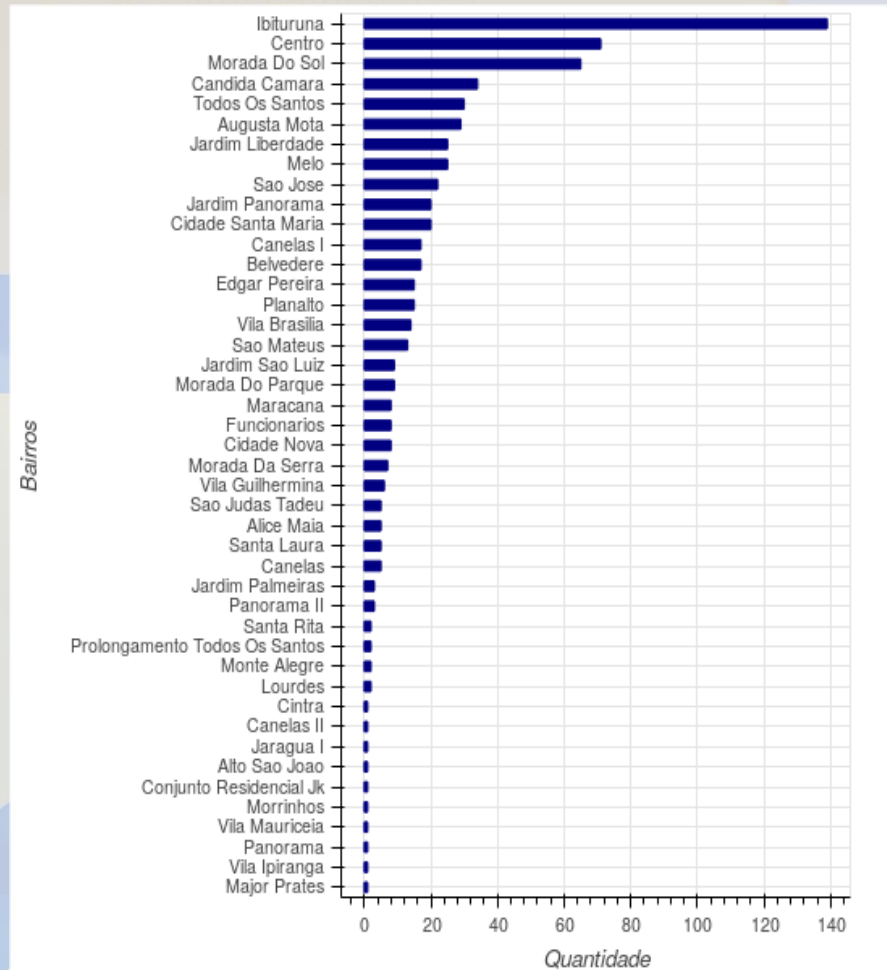


FIGURA 12: Imóveis por bairro.
Fonte: O Autor

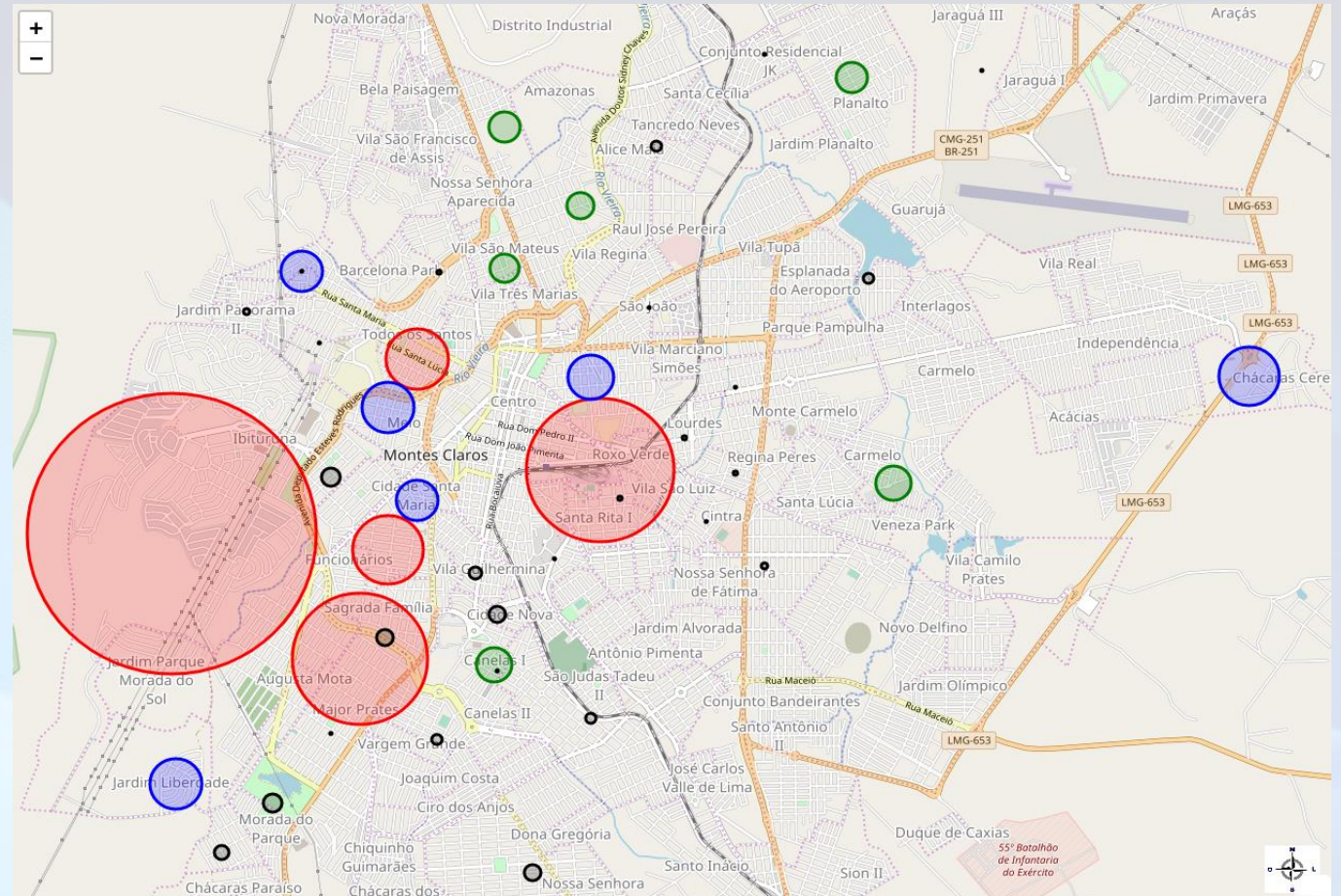


FIGURA 13: Distribuição de imóveis na cidade.
Fonte: O Autor

Análise Exploratória dos Dados

- Estimativa para o menor apartamento no bairro Centro;

```
7 dataset['area'].min()
8 previsao_preco('Centro', 1, 1, 0, 23.5)
```

Imóvel no Centro com 1 quarto(s), 1 banheiro(s), 0 vaga(s) de garagem e área de 23.5m². Valor previsto R\$ 153197.37

FIGURA 14: Estimativa de valor.

Fonte: O Autor

- Estimativa para o maior apartamento no bairro Centro;

```
7 dataset['quartos'].max()
8 dataset['banheiros'].max()
9 dataset['garagens'].max()
10 dataset['area'].max()
11 previsao_preco('Centro', 5, 5, 6, 273.74)
```

Imóvel no Centro com 5 quarto(s), 5 banheiro(s), 6 vaga(s) de garagem e área de 273.74m². Valor previsto R\$ 688400.0

FIGURA 15: Estimativa de valor.

Fonte: O Autor

Análise Exploratória dos Dados

- Estimativa para valor de apartamento ao aumentar-se a área;

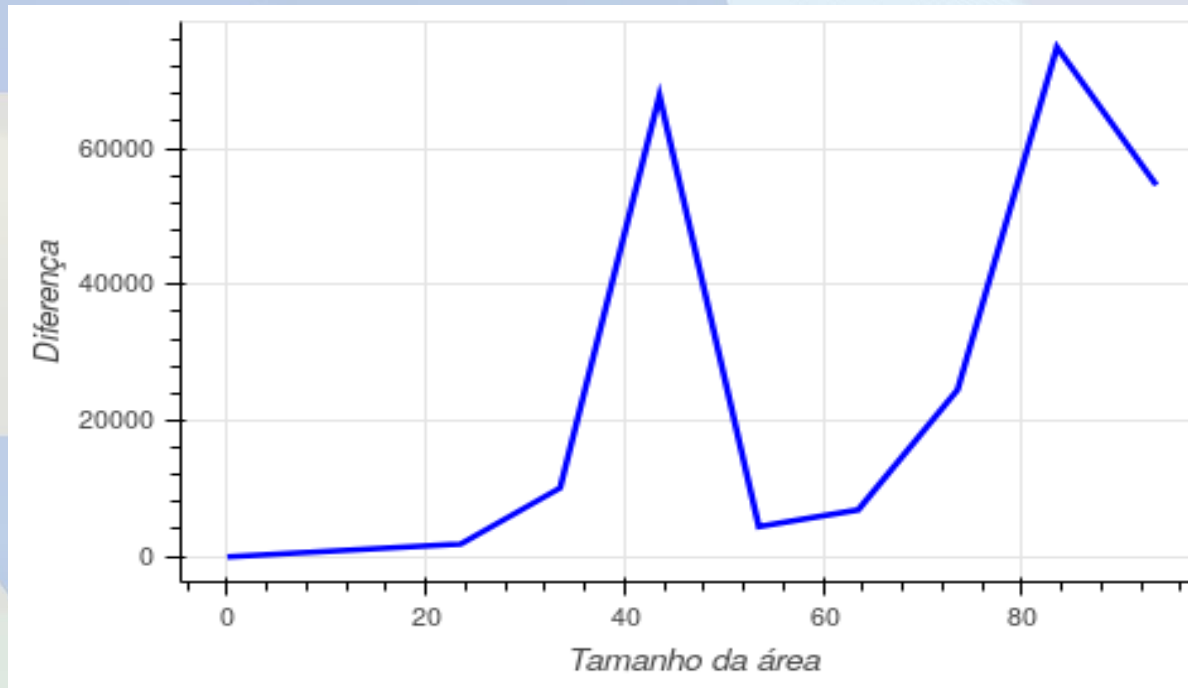


FIGURA 16: Diferença nos valor a cada aumento da área.
Fonte: O Autor

Análise Exploratória dos Dados

- Estimativa de preços para aumentos de quartos, banheiros e vagas de garagem;

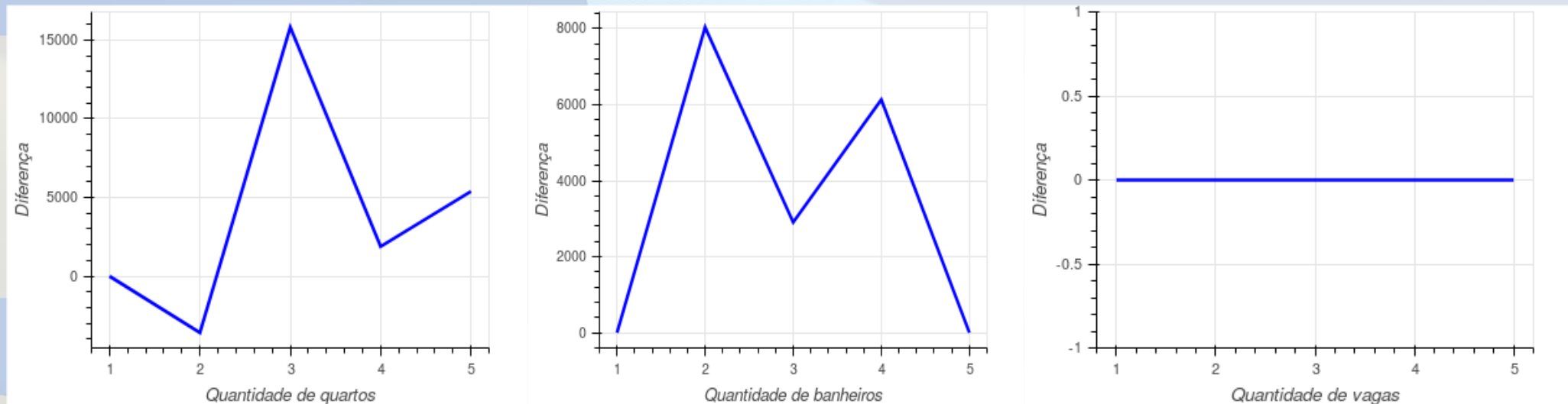


FIGURA 17: Diferenças de preços baseadas no aumento da quantidade de quartos, banheiros e vagas de garagem.

Fonte: O Autor

5 - Considerações Finais

- Este trabalho propôs um sistema que forneça ao usuário auxílio quando da decisão de efetivar a compra ou busca pelo apartamento ideal;
- Através dos testes realizados constatou-se o funcionamento correto de todos os procedimentos propostos;
- Melhorias futuras:
 - Abrangência de outros tipos de imóveis;
 - Diferentes condições de transações financeiras;
 - Criação de um sistema *web*;

6 - Referências

- BERNARD, Benoit. *Web Scraping and Crawling Are Perfectly Legal, Right?* 2017. Disponível em: <<https://benbernardblog.com/web-scraping-and-crawling-are-perfectly-legal-right/>>. Acesso em: 23 mar. 2018.
- DINO. Perspectivas do mercado imobiliário em 2018. *Revista Exame*, 19 de outubro de 2017. Disponível em: <<https://exame.abril.com.br/negocios/dino/perspectivas-do-mercado-imobiliario-em-2018/>>. Acesso em: 19 jan. 2018.
- HEMENWAY, Kevin; CALISHAIN, Tara. *Spidering Hacks: 100 Industrial-Strength Tips & Tools*. Sebastopol: O'Reilly, 2003.

- LENTE, Caio. *O Fluxo do Web Scraping*. 18/02/2018. Disponível em: <<http://curso-r.com/blog/2018/02/18/2018-02-18-fluxo-scraping/>>. Acesso em: 26 mar. 2018.
- RASCHKA, Sebastian; MIRJALILI, Vahid. *Python Machine Learning*. 2.ed. Birmingham: Packt Publishing, 2017.

SISTEMA AGREGADOR PARA ANÁLISE DE CARACTERÍSTICAS DE IMÓVEIS COM RASTREAMENTO WEB E APRENDIZADO DE MÁQUINA

OBRIGADO!

Felipe Israel Corrêa

Trabalho de Conclusão do Curso de Engenharia da Computação

Orientador: **PROF. Dr. Renato Dourado Maia**

Montes Claros - MG
Junho de 2018