



**UNIVERSIDADE PRESBITERIANA MACKENZIE**  
**Faculdade de Computação e Informática**



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

## Relatório

### Previsão de Preço de Imóveis de São Caetano do Sul



**UNIVERSIDADE PRESBITERIANA MACKENZIE**  
**Faculdade de Computação e Informática**



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

São Paulo  
1º/2024

Atividade em Grupo em ordem alfabética (Máximo 3 alunos)

Nome	RA
Lucca Eiki Amarante Millian	10390794



## Sumário

1	INTRODUÇÃO.....	4
2	INTEGRAÇÃO E DATA UNDERSTANDING.....	7
2.1	Base de dados.....	7
2.2	Atributos da Base de Dados Final.....	7
2.3	Tabela de Correlação de Pearson.....	8
2.4	Tabela de Frequência.....	8
2.5	Skewness.....	9
2.6	Kurtosis.....	9
2.7	Heatmap Correlação.....	10
2.8	Scatter plot.....	10
2.8.1	area total x preco.....	11
2.8.2	quartos x preco.....	11
2.9	Histograma Geral.....	12
2.10	Histograma com Curva: area_total.....	12
2.11	Histograma com Curva: preco.....	13
2.12	Histograma e Scatter Geral.....	14
2.13	Boxplot preco.....	15
2.14	Boxplot área_total.....	15
2.15	Boxplot quartos.....	16
2.16	Problemas Identificados.....	17
2.17	Questões – Compreendendo os Dados.....	17
3	DATA PREPARATION.....	18
3.1	Eliminação Manual de Atributos e Limpeza.....	18
3.2	Atributos Dataset: Antes da Preparação.....	20
3.3	Normalização MaxMin/Escore-Z, One-Hot Encoding e Ordenação do dataset	21
3.4	Atributos Dataset: Depois da Preparação e Ordenação das Colunas.....	21



**UNIVERSIDADE PRESBITERIANA MACKENZIE**  
**Faculdade de Computação e Informática**



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira



## 1 INTRODUÇÃO

Considere que fazemos parte de uma imobiliária que tem imóveis localizados UNICAMENTE em São Caetano do sul. Ela quer disponibilizar um preditor aos seus clientes para sugerir o valor de venda ou compra de um imóvel.

Algumas perguntas de negócio poderiam ser:

- Quero comprar um imóvel seguindo algumas características, é possível prever qual o preço ou valor do imóvel?
- Quais dados devem ser capturados para avaliar o valor a pagar?
- Posso montar um conjunto de dados para gerar um modelo preditivo que permite um vendedor avaliar o preço que deve ser ofertado do imóvel? Com esse mesmo conjunto de dados, caso eu queira comprar, é possível utilizá-lo para prever o valor de compra?

A coleta de dados foi realizada por alunos de um curso de Ciência da Computação e Sistemas de Informação em março de 2021 de uma instituição acadêmica fazendo uso de coleta de dados manuais e *webscraping* nos sites:

- [www.vivareal.com.br/](http://www.vivareal.com.br/)
- [www.creditas.com](http://www.creditas.com)
- [www.imovelweb.com.br](http://www.imovelweb.com.br)

Nesse cenário, foram coletados e montados 23 datasets por diferentes elementos, dos quais 20 deles constam em uma pasta de nome bd\_SCS.

Os atributos considerados durante essa coleta de dados foram:

- Tipo: Casa (1), Apto (2).
- Tempo do imóvel em anos: (0 = novo)
- Localização: Rua Luís Cavana, 70 - Centro, São Caetano do Sul – SP.
- Área total (m2):
- Área útil (m2):
- Número de quartos:
- Suíte: Sim (1) ou Não (0)
- Número de Banheiros
- Número de Vagas para Carros



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

- Academia: Sim (1) ou Não (0)
- Churrasqueira: Sim (1) ou Não (0)
- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do IPTU: em R\$
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.

Em um primeiro momento, o professor da disciplina fez uma integração de três *datasets*, sendo dois arquivos texto com extensão “csv”: bdSCS\_1\_csv.csv e bdSCS\_2\_csv.csv, e um do Excel com extensão xlsx: bdSCS\_3\_excel.xlsx, gerando ao final único dataset com nome bdSCS\_final\_csv.csv.

No processo de integração, foram realizadas manipulações nos conteúdos internos de cada um dos arquivos fazendo uso de algumas bibliotecas do Python, dentre elas: pandas, numpy, matplotlib, seaborn e re (expressões regulares), chegando-se a um resultado final de 13 atributos para bdSCS\_final\_csv.csv, sendo 12 deles candidatos a atributos de entrada e 1 o atributo alvo (Valor do Imóvel). Após esse processo, os atributos resultantes foram:

- Tipo: Casa (1), Apto (2).
- Bairro.
- Área total (m2)
- Número de quartos
- Número de Banheiros
- Número de Vagas para Carros
- Academia: Sim (1) ou Não (0)
- Churrasqueira: Sim (1) ou Não (0)
- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

A partir do exposto e dos resultados obtidos na etapa Data Understanding do processo CRISP-DM, realizar a etapa Data Preparation com o dataset obtido, preencher o solicitado no relatório.



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

## 2 INTEGRAÇÃO E DATA UNDERSTANDING

### 2.1 Base de dados

#### Professor

- Nome do arquivo csv utilizado: bdSCS\_1\_csv e bdSCS\_2\_csv
- Nome do arquivo Excel utilizado: bdSCS\_3\_excel
- Nome da base de dados final: bdSCS\_final\_csv.csv
- Total de Registros (Instâncias ou Observações) do arquivo de imóveis de São Caetano do Sul Final: 6381.
- Total de Atributos Final: 13.

#### Aluno

- Nome do arquivo csv utilizado: **bdSCS\_20\_csv e bdSCS\_21\_csv**
- Nome do arquivo Excel utilizado: **bdSCS\_final\_csv.csv**
- Nome da base de dados final: **bdSCS\_final\_novo\_csv.csv**
- Total de Registros (Instâncias ou Observações) do arquivo de imóveis de São Caetano do Sul Final: . **6480**
- Atributos da Base de Dados Final: **13**

#### Professor

Nesta etapa, foram considerados 13 atributos de cada imóvel, sendo eles:

- Tipo: Casa (1), Apto (2).
- Bairro.
- Área total (m2):
- Número de quartos:
- Número de Banheiros:
- Número de Vagas para Carros:
- Academia: Sim (1) ou Não (0)
- Churrasqueira: Sim (1) ou Não (0)





Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.

Aluno

Nesta etapa, foram considerados **13** atributos de cada imóvel, sendo eles:

- Tipo: Casa (1), Apto (2).
- Bairro.
- Área total (m2):
- Número de quartos:
- Número de Banheiros:
- Número de Vagas para Carros:
- Academia: Sim (1), Não (0)
- Churrasqueira: Sim (1) ou Não (0)
- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.

## 2.2 Tabela de Correlação de Pearson

Professor



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

	tipo	area_total	quartos	academia	churrasqueira	elevador	salao_festa	piscina	condominio	preco
tipo	1.000000	-0.302339	-0.195422	0.026421	-0.013209	0.103412	0.052864	0.003737	-0.055293	-0.036151
area_total	-0.302339	1.000000	0.532947	-0.009867	0.023002	-0.085138	-0.032768	0.016376	0.049051	0.270127
quartos	-0.195422	0.532947	1.000000	-0.067305	0.009903	-0.027897	-0.032116	0.038520	0.055467	0.066994
academia	0.026421	-0.009867	-0.067305	1.000000	-0.117725	-0.397707	0.176674	-0.364900	-0.009944	-0.027186
churrasqueira	-0.013209	0.023002	0.009903	-0.117725	1.000000	0.296261	0.269835	0.436321	-0.007544	0.001135
elevador	0.103412	-0.085138	-0.027897	-0.397707	0.296261	1.000000	0.285759	0.044278	-0.026394	-0.018076
salao_festa	0.052864	-0.032768	-0.032116	0.176674	0.269835	0.285759	1.000000	0.319325	-0.010579	0.003303
piscina	0.003737	0.016376	0.038520	-0.364900	0.436321	0.044278	0.319325	1.000000	-0.000539	0.025697
condominio	-0.055293	0.049051	0.055467	-0.009944	-0.007544	-0.026394	-0.010579	-0.000539	1.000000	0.012667
preco	-0.036151	0.270127	0.066994	-0.027186	0.001135	-0.018076	0.003303	0.025697	0.012667	1.000000

## Aluno

Unnamed: 0	1.000000	-0.067480	0.042539	0.010821	0.009078	0.012037	-0.015046	-0.005275	0.003427	-0.289575	0.009318
tipo	-0.067480	1.000000	-0.297700	-0.192891	0.029297	-0.009103	0.104005	0.058751	0.006464	-0.047396	-0.034264
area_total	0.042539	-0.297700	1.000000	0.535531	-0.007559	0.022753	-0.082016	-0.033618	0.017120	0.058425	0.271137
quartos	0.010821	-0.192891	0.535531	1.000000	-0.063413	0.009336	-0.027335	-0.033835	0.038630	0.066690	0.068472
academia	0.009078	0.029297	-0.007559	-0.063413	1.000000	-0.096058	-0.362613	0.189222	-0.335010	-0.001367	-0.027258
churrasqueira	0.012037	-0.009103	0.022753	0.009336	-0.096058	1.000000	0.310488	0.288377	0.447799	-0.008916	0.002984
elevador	-0.015046	0.104005	-0.082016	-0.027335	-0.362613	0.310488	1.000000	0.308932	0.069034	-0.032022	-0.013297
salao_festa	-0.005275	0.058751	-0.033618	-0.033835	0.189222	0.288377	0.308932	1.000000	0.333113	-0.016107	0.005984
piscina	0.003427	0.006464	0.017120	0.038630	-0.335010	0.447799	0.069034	0.333113	1.000000	-0.000715	0.026976
condominio	-0.289575	-0.047396	0.058425	0.066690	-0.001367	-0.008916	-0.032022	-0.016107	-0.000715	1.000000	0.012762
preco	0.009318	-0.034264	0.271137	0.068472	-0.027258	0.002984	-0.013297	0.005984	0.026976	0.012762	1.000000

## 2.3 Tabela de Frequência

### Professor

	fabs	frel
tipo		
1	109	1.708196
2	6272	98.291804



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

	fabs	frel
quartos		
1	72	1.128350
2	3741	58.627174
3	2348	36.796740
4	214	3.353706
5	4	0.062686
6	2	0.031343

Aluno

	fabs	frel
tipo		
1	113	1.743558
2	6368	98.256442

	fabs	frel
quartos		
1	73	1.126369
2	3799	58.617497
3	2384	36.784447
4	219	3.379108
5	4	0.061719
6	2	0.030859



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

## 2.4 Skewness

Professor

tipo	-7.455521
area_total	6.659797
quartos	0.732790
academia	0.909522
churrasqueira	-0.761851
elevador	-1.689357
salao_festa	-1.408184
piscina	-0.374116
condominio	0.155014
preco	2.833031

Aluno

```
Unnamed: 0      0.000000
tipo            -7.375426
area_total      6.636522
quartos         0.732885
academia        0.867834
churrasqueira  -0.782557
elevador        -1.662110
salao_festa     -1.435816
piscina         -0.394376
condominio      0.248000
preco           2.858902
dtype: float64
```

## 2.5 Kurtosis

Professor

tipo	53.601595
area_total	110.777605
quartos	0.338157
academia	-1.173137
churrasqueira	-1.420028
elevador	0.854195
salao_festa	-0.017025
piscina	-1.860621
condominio	1.240078
preco	7.515773



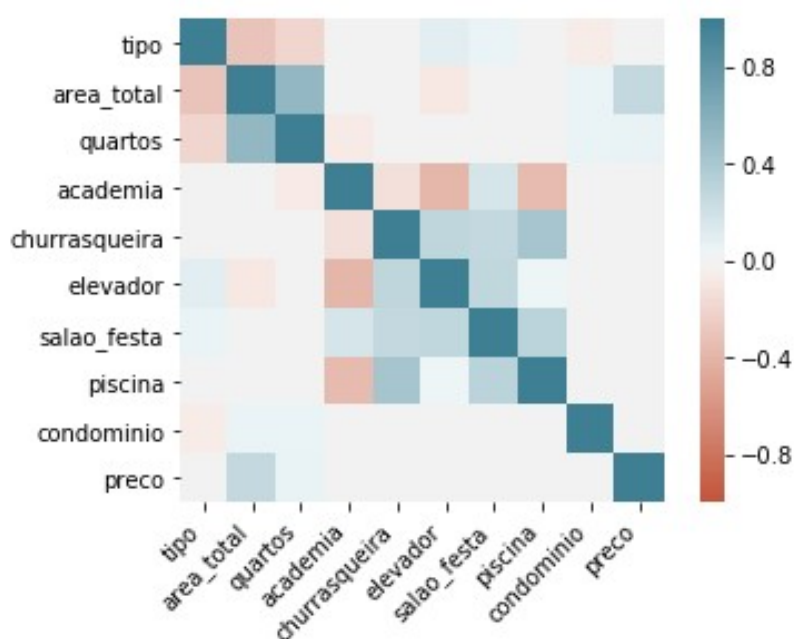
Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

Aluno

```
Unnamed: 0      -1.200000  
tipo            52.413080  
area_total     110.532471  
quartos        0.330553  
academia       -1.163931  
churrasqueira  -1.301012  
elevador        1.188557  
salao_festa     0.225419  
piscina        -1.773306  
condominio      1.837394  
preco           7.684721  
dtype: float64
```

## 2.6 Heatmap Correlação

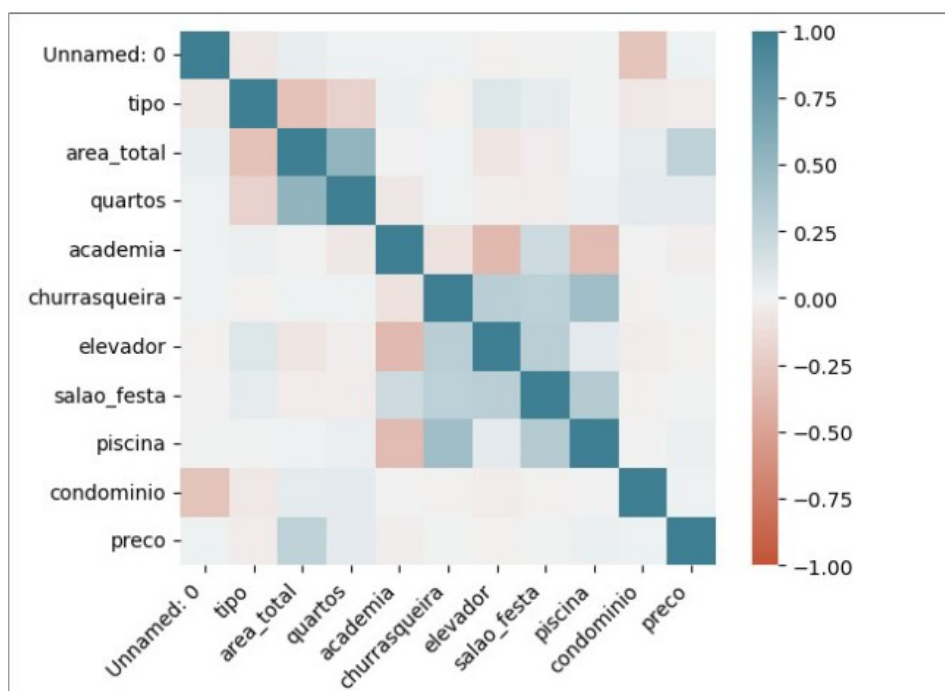
Professor





Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

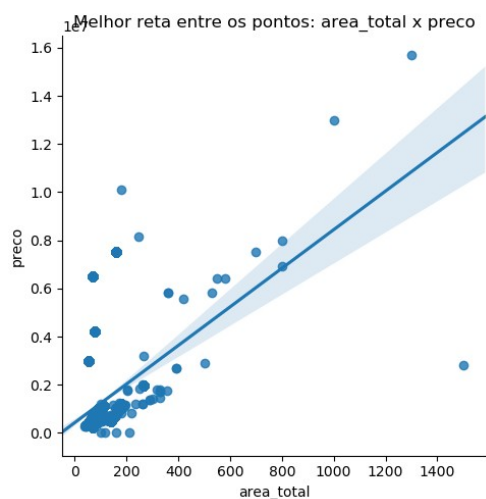
Aluno



## 2.7 Scatter plot

### 2.7.1 area total x preco

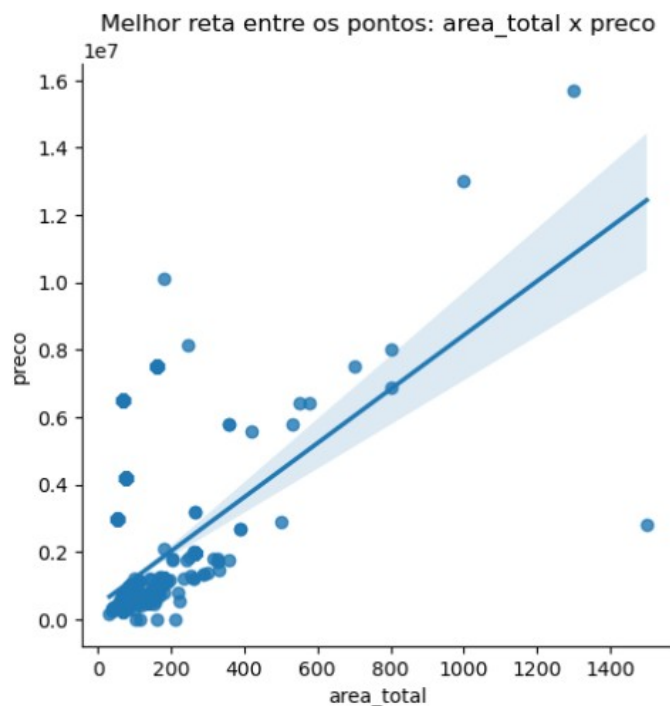
Professor





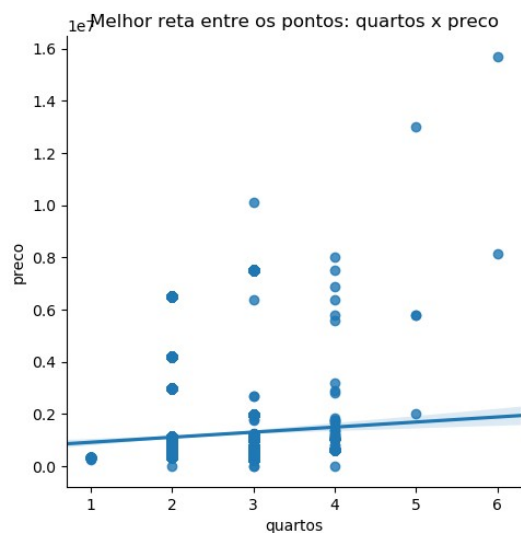
Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

Aluno



## 2.7.2 quartos x preco

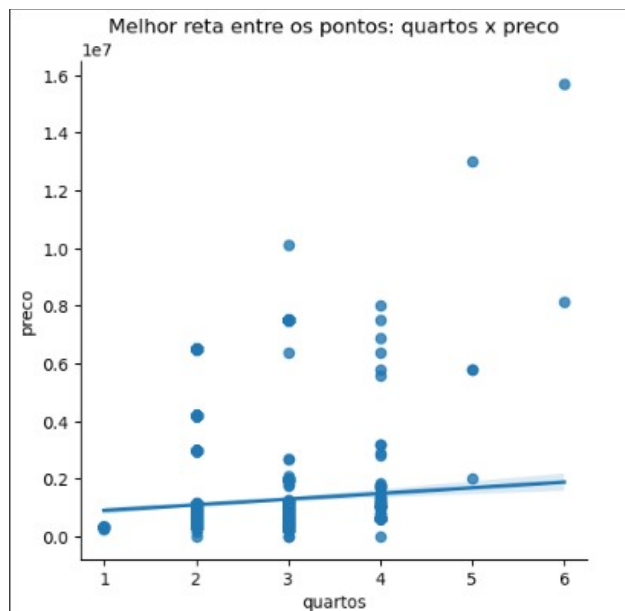
Professor





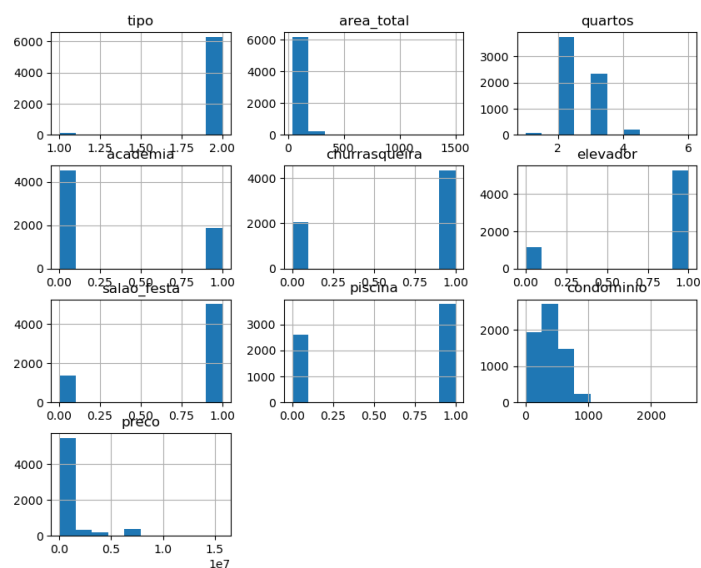
Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

Aluno



## 2.8 Histograma Geral

Professor



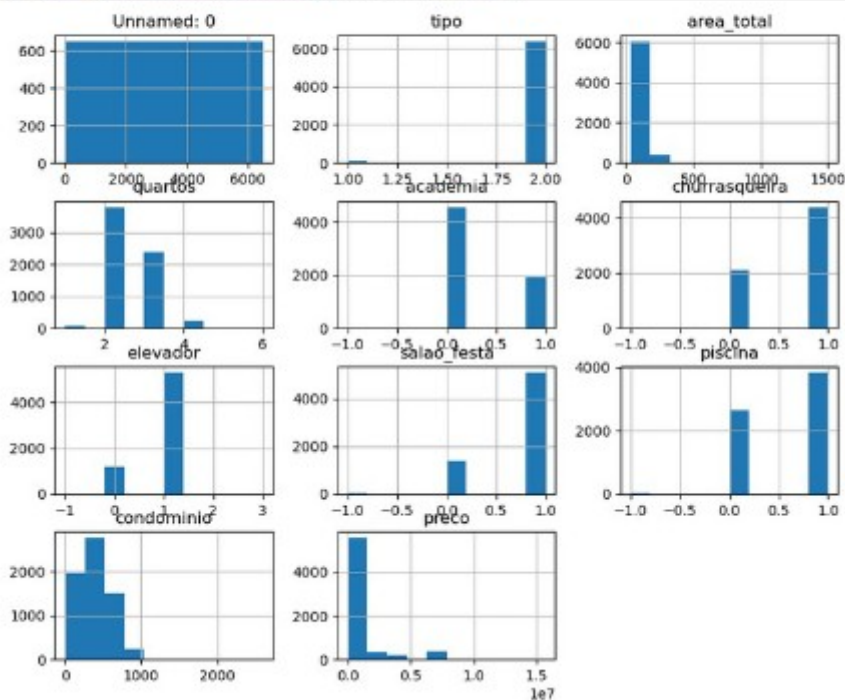




Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

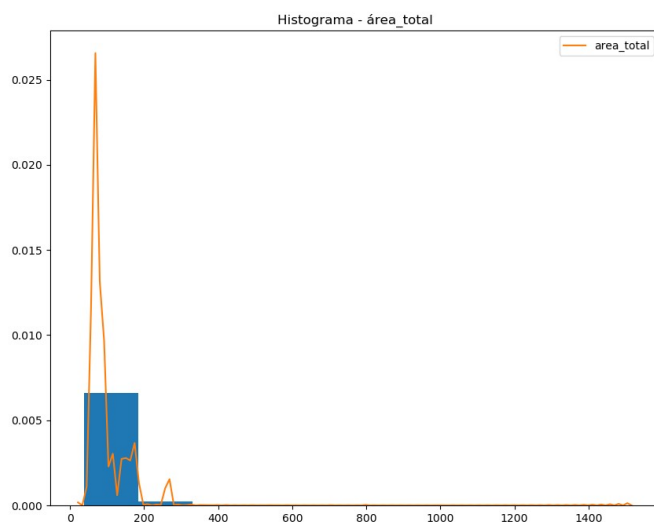
Aluno

```
[24]: # Apresenta o histograma de todos os atributos
dfImoveis_SF5_final.hist().all()
# Salva a Figura
plt.savefig('imagens/histogramaimoveis.png', bbox_inches='tight')
```



## 2.9 Histograma com Curva: area\_total

Professor





Curso: Bacharelado em Ciência da Computação

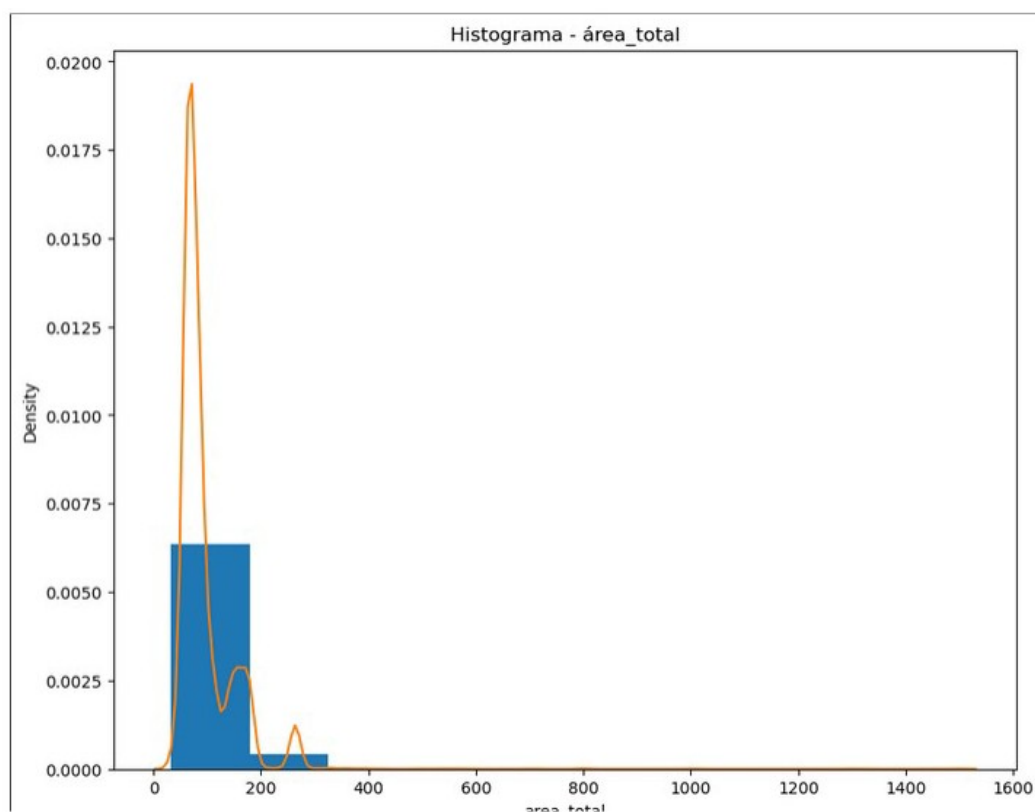
Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

sk = 6.659796686013957, Ck = 110.77760469964831

Média=96.0025074439743, Mediana=76.0, Moda= 90

Aluno



sk = 6.636521900984331, Ck = 110.53247069853705

Média=95.9267088412282, Mediana=76.0, Moda=0 90

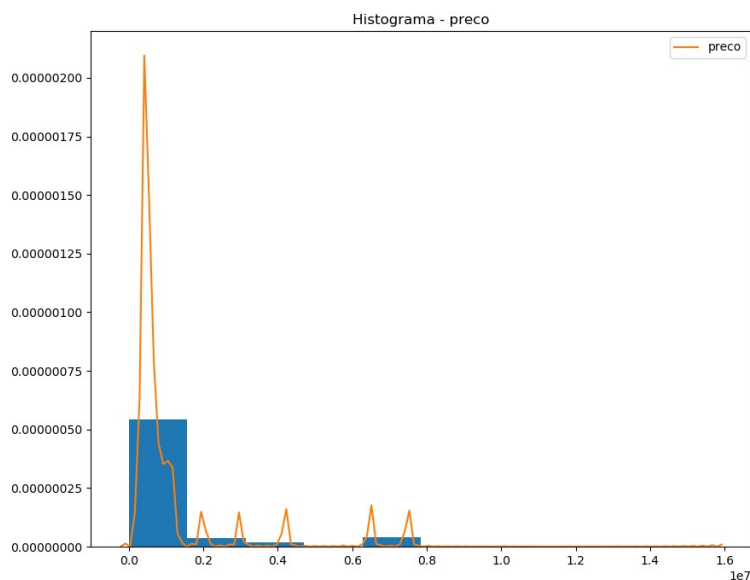
Name: area\_total, dtype: int64

## 2.10 Histograma com Curva: preco

Professor



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

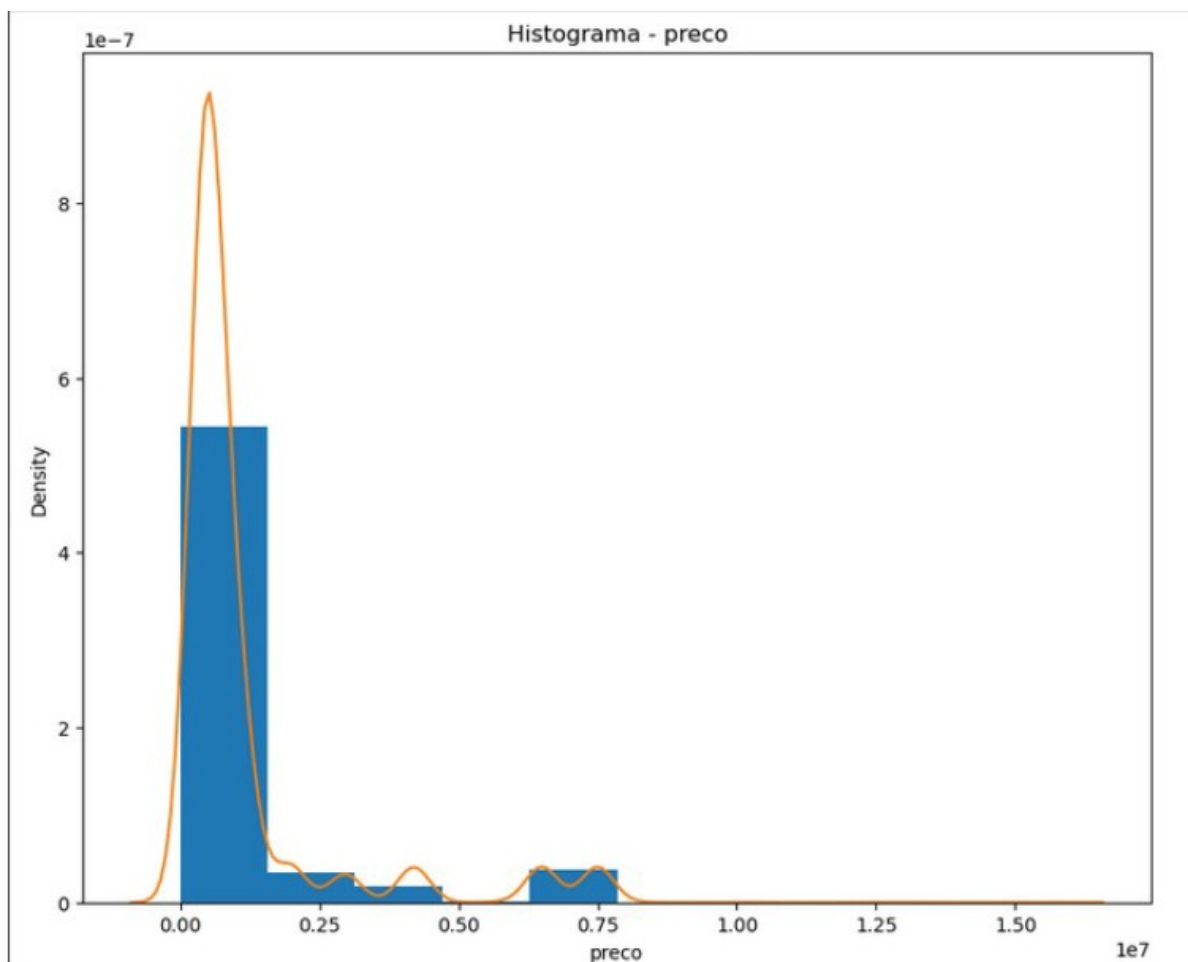


sk = 2.8330314405741173, Ck = 7.515772619574491  
Média=1193589.0520294625, Mediana=510000.0, Moda= 990000



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

Aluno



$sk = 2.858901664757256$ ,  $Ck = 7.684720663544072$

Média=1184245.0276191945, Mediana=510000.0, Moda=0 450000

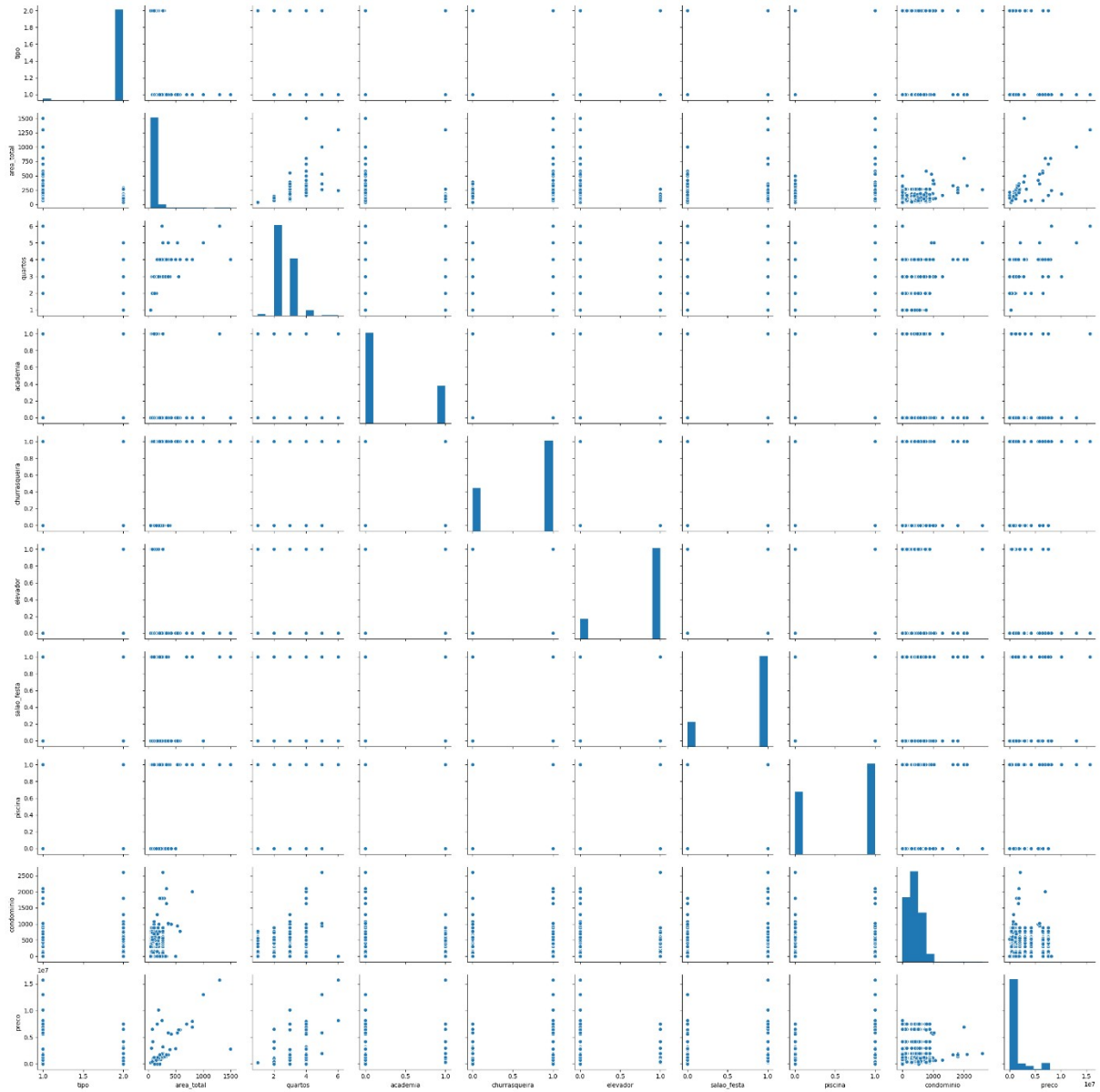
Name: preco, dtype: int64

## 2.11 Histograma e Scatter Geral

Professor



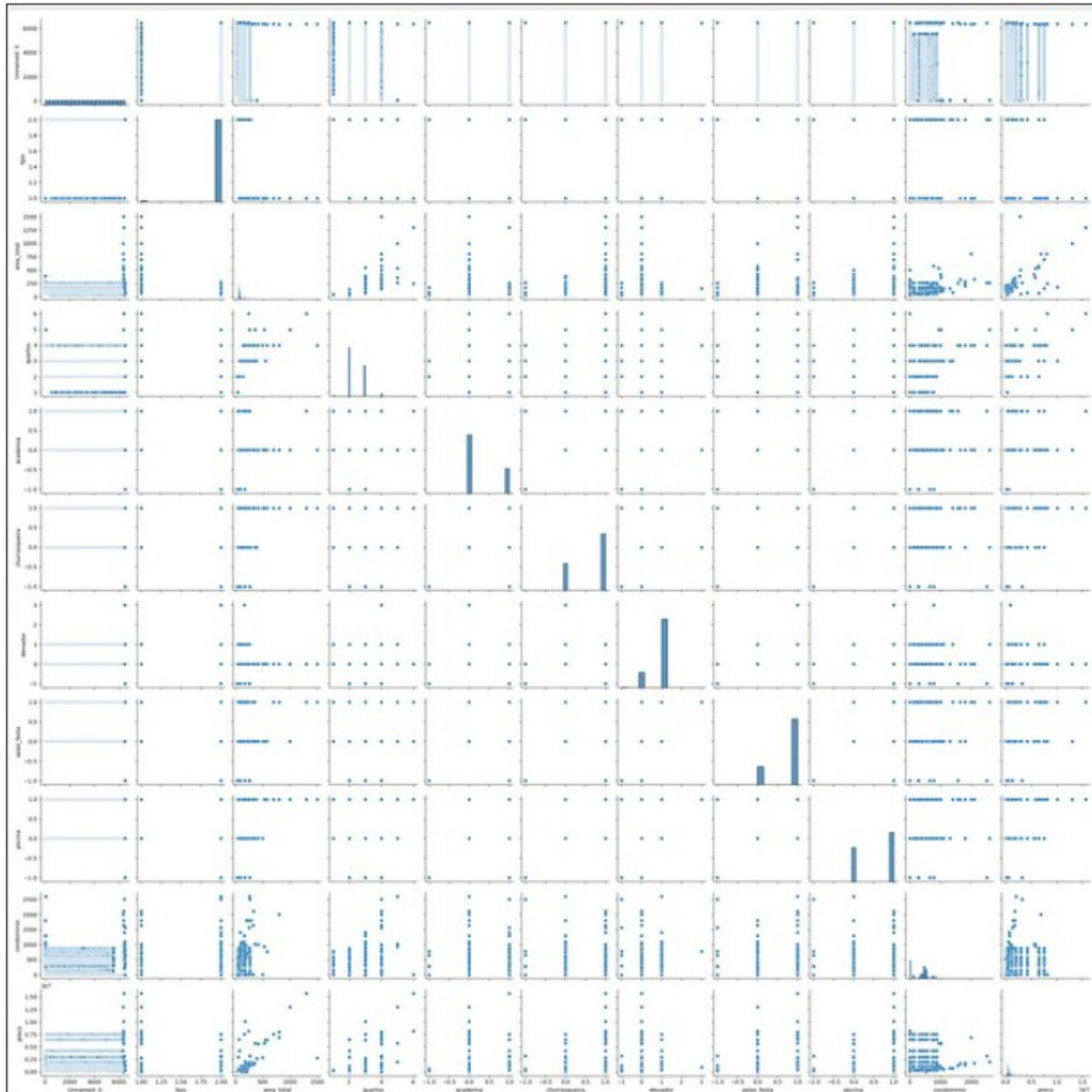
Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira



Aluno



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira



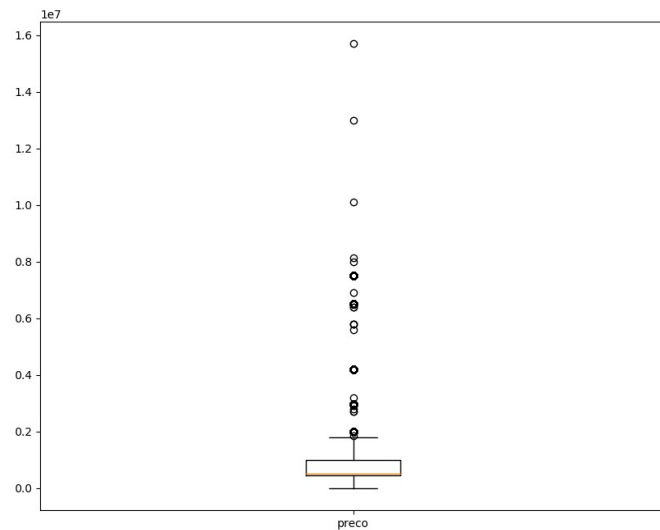
## 2.12 Boxplot preco

Professor

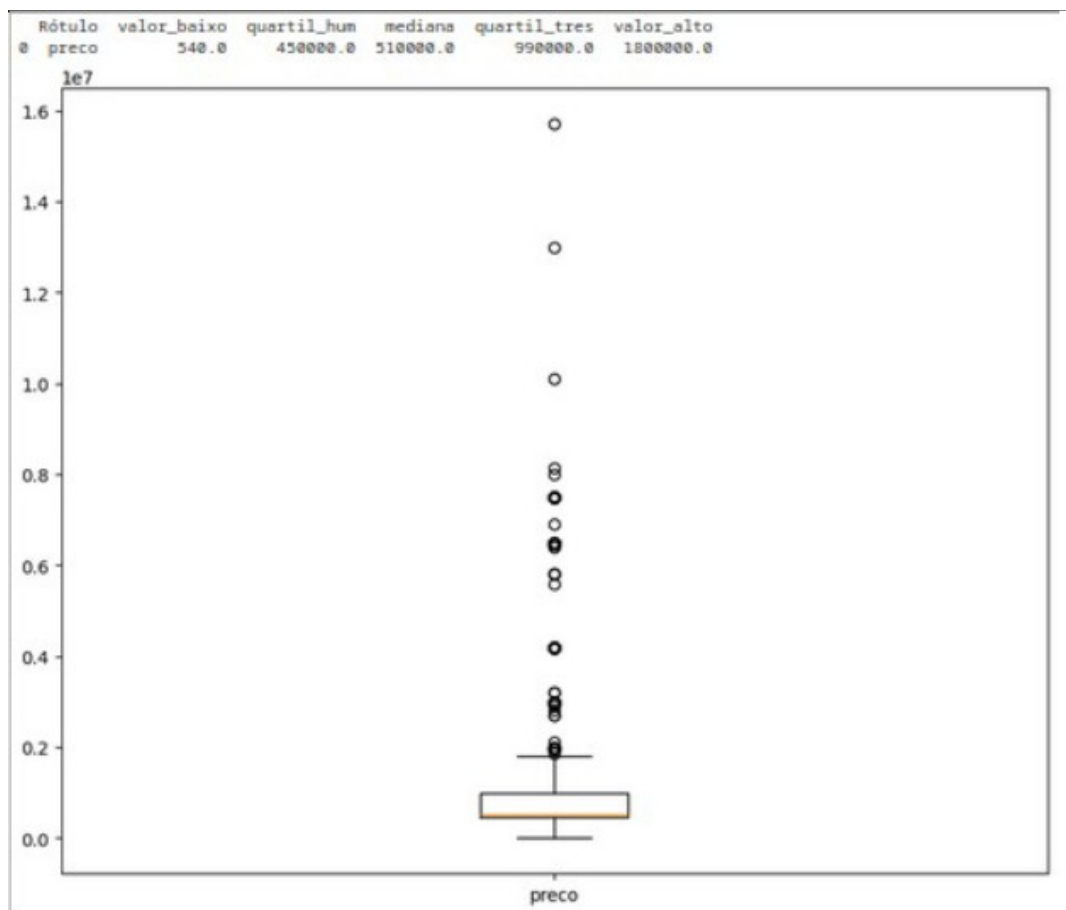
Rótulo	valor_baixo	quartil_hum	mediana	quartil_tres	valor_alto
preco	540.0	450000.0	510000.0	990000.0	1800000.0



Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira



Aluno





Curso: Bacharelado em Ciência da Computação

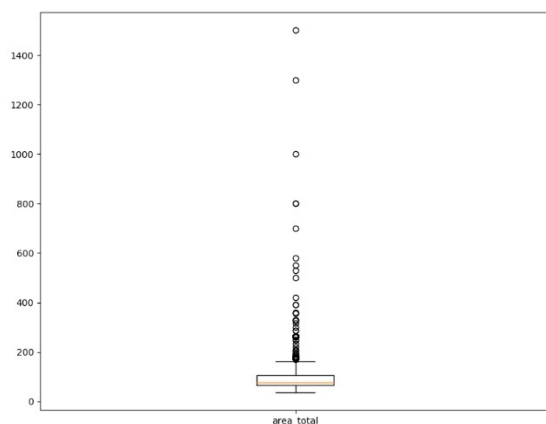
Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

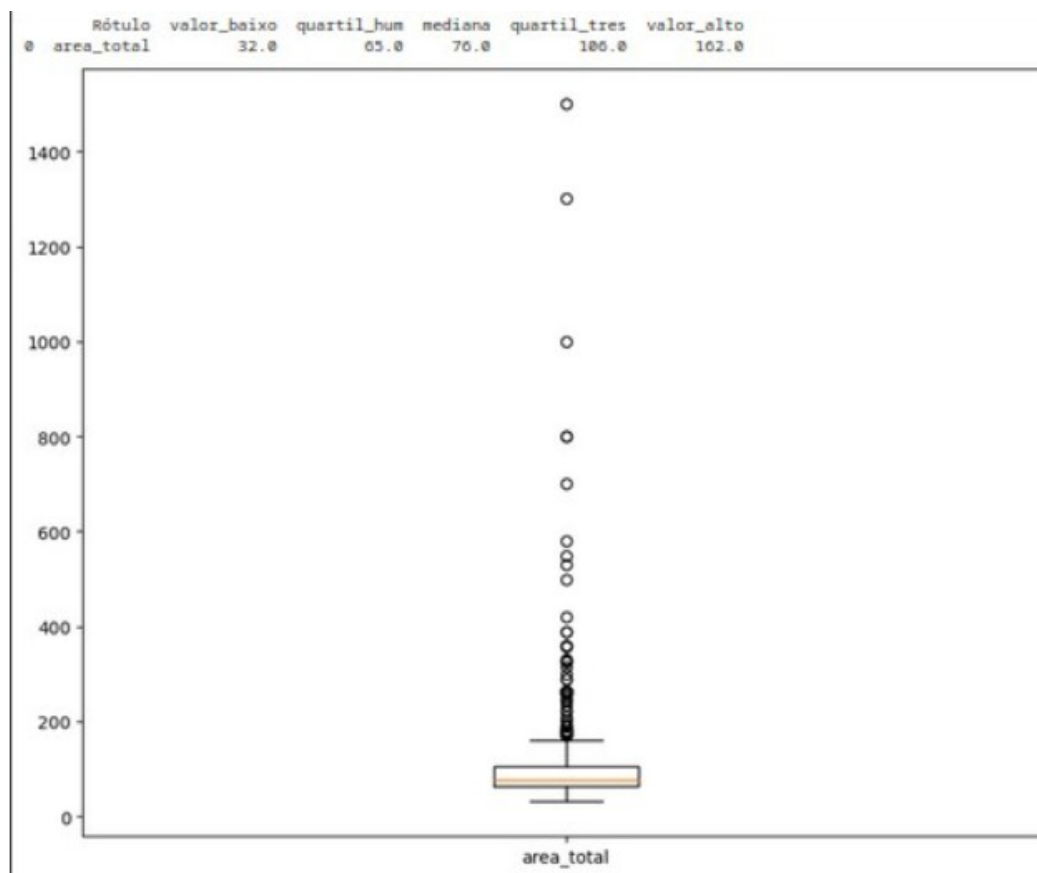
## 2.13 Boxplot área\_total

Professor

Rótulo	valor_baixo	quartil_hum	mediana	quartil_tres	valor_alto
area_total	38.0	65.0	76.0	106.0	162.0



Aluno





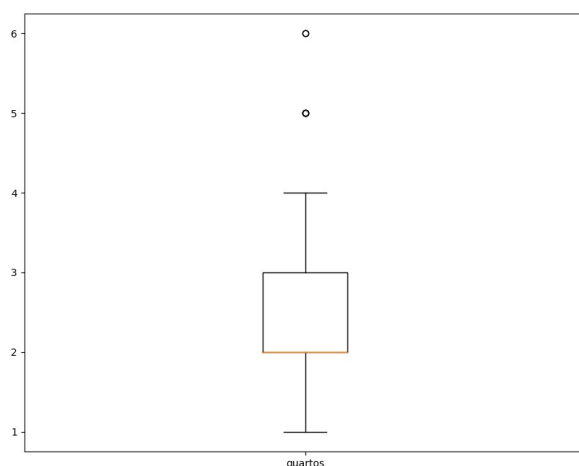


Curso: Bacharelado em Ciência da Computação  
Disciplina: Inteligência Artificial – 7º N  
Prof. Dr. Ivan Carlos Alcântara de Oliveira

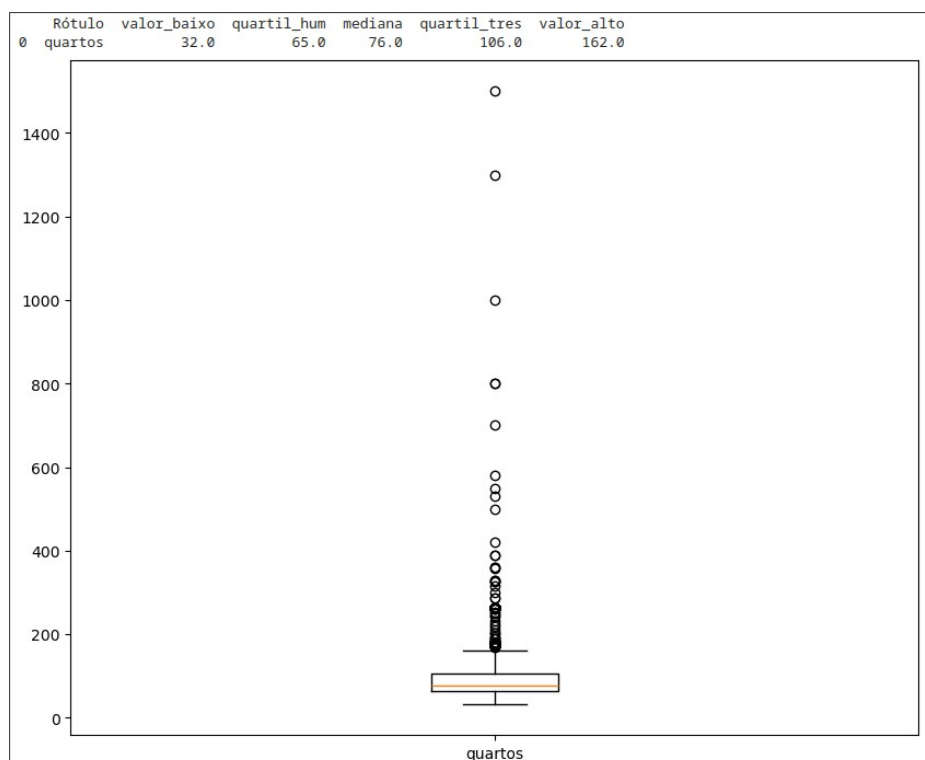
## 2.14 Boxplot quartos

Professor

Rótulo	valor_baixo	quartil_hum	mediana	quartil_tres	valor_alto
quartos	1.0	2.0	2.0	3.0	4.0



Aluno





Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

### **2.15 Problemas Identificados**

Dos 17 atributos inicialmente selecionados, o tempo do imóvel, suíte, área útil e valor do IPTU foram removidos por não apresentar valores significativos na maior parte das observações.

O atributo localização continha o endereço completo, mas foi considerado somente o valor do bairro. Por este fato, o atributo foi renomeado para bairro.

O atributo condomínio não apresenta dados em alguns registros. Logo, algo deve ser feito na etapa de preparação dos dados.

O atributo banheiros e vagas apresentam valores inconsistentes internamente. Logo, algo deve ser feito na etapa de preparação dos dados.

### **2.16 Questões – Compreendendo os Dados**

A resposta a essas questões deve constar no Notebook (IA\_EAD\_Atividade\_DataUnderstanding.ipynb) em local apresentado no seu interior. Os resultados obtidos do código elaborado deve ser apresentado como respostas às questões elencadas abaixo.

a) Qual a quantidade de imóveis por bairro do tipo 2?

Resposta:

bairro

Barcelona            728

Boa Vista            396

Centro                171

Cerâmica            916

Fundação            251

Jardim Planalto      1

Jardim Sao Caetano 152

Mauá                  191

Nova Gerti           8

Nova Gerty          260



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

Olímpico	726
Osvaldo Cruz	153
Santa Maria	987
Santa Paula	992
Santo Antônio	435
São José	1

b) Qual o valor médio dos preços dos imóveis do tipo 1 do bairro Santa Paula?

Resposta: nan

c) Quantos imóveis do tipo 1 que tem menos do que 2 banheiros com área total maior do que 100 m<sup>2</sup>

Resposta: 72

d) Qual o valor médio do condomínio dos imóveis do tipo 1 que tem academia, elevador e piscina?

Resposta: 1035318.0908265214

e) Qual a área média dos imóveis (tipo 1 e 2) que tem 2 quartos ou mais e valor do imóvel maior do que R\$ 600000,00?

Resposta: 154.14350064350063

f) Qual a quantidade de imóveis do tipo 1 que pelo menos 2 quartos, ao menos 2 vagas de garagem, com salão de festas e churrasqueira?

Resposta: 15

g) Qual a quantidade de imóveis do tipo 1 possui seu preço maior do que a média dos preços dos imóveis (também do tipo 1)?

Resposta: 25

h) Quantos imóveis do tipo 2 são do bairro Barcelona, Santa Maria e Nova Gerty e tem pelo menos 3 quartos?

Resposta: 556

i) Apresentar a relação de imóveis de qualquer tipo com preço maior do que R\$ 600.000,00 com pelo menos 2 vagas de garagem?

Resposta: 1920



### 3 DATA PREPARATION

#### 3.1 Eliminação Manual de Atributos e Limpeza

##### Professor

Atividade		Valores/Ação realizada
Base de dados antes da preparação	Nome da Base	bdSCS_final_profIvan.csv
	Total de objetos	6381
	Total de Atributos	13
Eliminação Manual de Atributos	Removido o atributo	Tipo, pois havia poucos valores com Tipo = 1 (casa), podiam afetar a sua predição.
	Antes de remover o atributo Tipo (coluna)	Foram removidos os objetos com Tipo = '1' (casa)
	Total de Objetos restantes	6272
	Total de Atributos restantes	12
Limpeza de Dados – Dados Incompletos	Objetos contendo o atributo condomínio	11 deles com valores nulos. Foram todos removidos.
	Total de Objetos restantes	6261
Limpeza de Dados – Dados Inconsistentes	Atributo banheiros contendo 179 “—” e 45 “2-3”	Removidos todos os objetos com esses valores
	Total de Objetos restantes	6037
	Os atributos banheiros e vagas estavam com o tipo de dados object	Foram convertidos para inteiro
	O conteúdo do atributo bairro apresentavam dois problemas	1) acentos nos nomes dos campos 2) Bairro Jardim São Caetano tem com acento e sem acento.
	Ação realizada no atributo bairro	Os acentos foram removidos, resolvendo os dois problemas.
	Remoção dos objetos nos quais o atributo	tem o valor 0 (757 objetos)



Curso: Bacharelado em Ciência da Computação  
 Disciplina: Inteligência Artificial – 7º N  
 Prof. Dr. Ivan Carlos Alcântara de Oliveira

	condomínio	
	Total de Objetos restantes	5280
Limpeza de Dados - Dados Redundantes - Objetos	Foram identificados alguns objetos redundantes	Total 809. Todos foram removidos.
	Total de Objetos restantes	4471
Limpeza de Dados - Dados Redundantes - Atributos	Inexistente	
Limpeza de Dados - Ruídos	Com base no boxplot,	foi selecionado o atributo preço para remoção de outliers
	Técnica utilizada para remoção dos outliers do atributo preço	outliers baixos estão abaixo de $Q1 - 1,5 \cdot FIQ$ e outliers altos estão acima de $Q3 + 1,5 \cdot FIQ$ , onde $FIQ = Q3 - Q1$ .
	Total de Objetos restantes	3831

### Aluno

Atividade		Valores/Ação realizada
Base de dados antes da preparação	Nome da Base	bdSCS_final_csv.csv
	Total de objetos	6256
	Total de Atributos	13
Eliminação Manual de Atributos	Removido o atributo	Tipo, pois havia poucos valores com Tipo = 1 (casa), podiam afetar a sua predição.
	Antes de remover o atributo Tipo (coluna)	Foram removidos os objetos com Tipo = '1' (casa)
	Total de Objetos restantes	6144
	Total de Atributos restantes	12
Limpeza de Dados - Dados Incompletos	Objetos contendo o atributo condomínio	11 deles com valores nulos. Foram todos removidos.
	Total de Objetos restantes	6133



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

Limpeza de Dados – Dados Inconsistentes	Atributo banheiros contendo 179 “—” e 45 “2-3”	Removidos todos os objetos com esses valores
	Total de Objetos restantes	<b>6133</b>
	Os atributos banheiros e vagas estavam com o tipo de dados object	Foram convertidos para inteiro
	O conteúdo do atributo bairro apresentavam dois problemas	1) acentos nos nomes dos campos 2) Bairro Jardim São Caetano tem com acento e sem acento.
	Ação realizada no atributo bairro	Os acentos foram removidos, resolvendo os dois problemas.
	Remoção dos objetos nos quais o atributo condomínio	tem o valor 0 ( <b>762</b> objetos)
	Total de Objetos restantes	<b>5371</b>
Limpeza de Dados – Dados Redundantes - Objetos	Foram identificados alguns objetos redundantes	Total <b>811</b> . Todos foram removidos.
	Total de Objetos restantes	<b>4560</b>
Limpeza de Dados - Dados Redundantes - Atributos	Inexistente	
Limpeza de Dados - Ruídos	Com base no boxplot,	foi selecionado o atributo preço para remoção de outliers
	Técnica utilizada para remoção dos outliers do atributo preço	outliers baixos estão abaixo de $Q1 - 1,5 \cdot FIQ$ e outliers altos estão acima de $Q3 + 1,5 \cdot FIQ$ , onde $FIQ = Q3 - Q1$ .
	Total de Objetos restantes	<b>3918</b>

### 3.2 Atributos Dataset: Antes da Preparação

#### Professor

Nesta etapa, foram considerados 13 atributos de cada imóvel, sendo eles:



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

- Tipo: Casa (1), Apto (2).
- Bairro.
- Área total (m2):
- Número de quartos:
- Número de Banheiros:
- Número de Vagas para Carros:
- Academia: Sim (1) ou Não (0)
- Churrasqueira: Sim (1) ou Não (0)
- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.

### **Aluno**

Nesta etapa, foram considerados **13** atributos de cada imóvel, sendo eles:

- Tipo: Casa (1), Apto (2).
- Bairro.
- Área total (m2):
- Número de quartos:
- Número de Banheiros:
- Número de Vagas para Carros:
- Academia: Sim (1) ou Não (0)
- Churrasqueira: Sim (1) ou Não (0)
- Elevador: Sim (1) ou Não (0)
- Salão de Festas: Sim (1) ou Não (0)
- Piscina: Sim (1) ou Não (0)
- Valor do Condomínio: em R\$.
- Valor do Imóvel: em R\$.



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

### 3.3 Normalização MaxMin/Escore-Z, One-Hot Encoding e Ordenação do dataset

#### Professor

Atividade		Atributo
Normalização MaxMin	Convertidos dentro do intervalo: [0, 1]	Quartos, Banheiros e Vagas
Normalização (Padronização) Escore-Z	Convertidos para média 0 e desvio padrão 1	Área_total e condomínio
One-Hot Encoding	Transformação de Variáveis Categóricas para Numéricas	Bairro
	O atributo categórico apresentava	13 valores diferentes
	Logo, aumentou o número de atributos	Em mais 13
Ordenação do dataset	O atributo alvo não estava no final, então	Preço teve a ordem modificada para o final do dataset

### 3.4 Atributos Dataset: Depois da Preparação e Ordenação das Colunas

#### Professor

Ao final foram obtidos 3831 objetos contendo 24 atributos, sendo:

- Candidatos a Atributos Preditores (23):
  1. area\_total: Transformação Escore-Z.
  2. quartos: Transformação MaxMin (0-1).
  3. banheiros: Transformação MaxMin (0-1).
  4. vagas: Transformação MaxMin (0-1).
  5. academia: Sim (1) ou Não (0).
  6. churrasqueira: Sim (1) ou Não (0).
  7. elevador: Sim (1) ou Não (0).
  8. salao\_festas: Sim (1) ou Não (0).
  9. piscina: Sim (1) ou Não (0).
  10. condominio: Transformação Escore-Z.
  11. bairro\_Barcelona: Sim (1) ou Não (0).





Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

- 12. bairro\_Boa\_Vista: Sim (1) ou Não (0).
- 13. bairro\_Centro: Sim (1) ou Não (0).
- 14. bairro\_Ceramica: Sim (1) ou Não (0).
- 15. bairro\_Fundacao: Sim (1) ou Não (0).
- 16. bairro\_Jardim\_Sao\_Caetano: Sim (1) ou Não (0).
- 17. bairro\_Maua: Sim (1) ou Não (0).
- 18. bairro\_Nova\_Gerty: Sim (1) ou Não (0).
- 19. bairro\_Olimpico: Sim (1) ou Não (0).
- 20. bairro\_Osvaldo\_Cruz: Sim (1) ou Não (0).
- 21. bairro\_Santa\_Maria: Sim (1) ou Não (0).
- 22. bairro\_Santa\_Paula: Sim (1) ou Não (0).
- 23. bairro\_Santo\_Antonio: Sim (1) ou Não (0).

- Atributo Alvo
  - preco: em R\$.

## Aluno

Ao final foram obtidos 3918 objetos contendo 27 atributos, sendo:

- Candidatos a Atributos Preditores (26):
  1. 'area\_total',
  2. 'quartos',
  3. 'banheiros',
  4. 'vagas',
  5. 'academia',
  6. 'churrasqueira',
  7. 'elevador',
  8. 'salao\_festa',
  9. 'piscina',
  10. 'condominio',
  11. 'preco',
  12. 'bairro\_Barcelona',
  13. 'bairro\_Boa\_Vista',
  14. 'bairro\_Centro',
  15. 'bairro\_Ceramica',
  16. 'bairro\_Fundacao',
  17. 'bairro\_Jardim\_Planalto',



Curso: Bacharelado em Ciência da Computação

Disciplina: Inteligência Artificial – 7º N

Prof. Dr. Ivan Carlos Alcântara de Oliveira

18. 'bairro\_Jardim\_Sao\_Caetano',
19. 'bairro\_Maua',
20. 'bairro\_Nova\_Gerti',
21. 'bairro\_Nova\_Gerty',
22. 'bairro\_Olimpico',
23. 'bairro\_Osvaldo\_Cruz',
24. 'bairro\_Santa\_Maria',
25. 'bairro\_Santa\_Paula',
26. 'bairro\_Santo\_Antonio',
27. 'bairro\_Sao\_Jose'

- Atributo Alvo
  - preco: em R\$.