

Relatório de Análise VIII

Identificando e Removendo Outliers

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
plt.rc('figure', figsize = (14,6))
```

```
In [2]: dados = pd.read_csv("dados/aluguel_residencial.csv", sep = ";")
```

```
In [3]: dados[dados['Valor'] >= 500000]
```

Out[3]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU	Valor m2	Ag
7629	Apartamento	Barra da Tijuca	1	1	0	65	600000.0	980.0	120.0	9230.77	Apart
10636	Casa de Condomínio	Freguesia (Jacarepaguá)	4	2	3	163	800000.0	900.0	0.0	4907.98	
12661	Apartamento	Freguesia (Jacarepaguá)	2	2	1	150	550000.0	850.0	150.0	3666.67	Apart
13846	Apartamento	Recreio dos Bandeirantes	3	2	1	167	1250000.0	1186.0	320.0	7485.03	Apart
15520	Apartamento	Botafogo	4	1	1	300	4500000.0	1100.0	0.0	15000.00	Apart

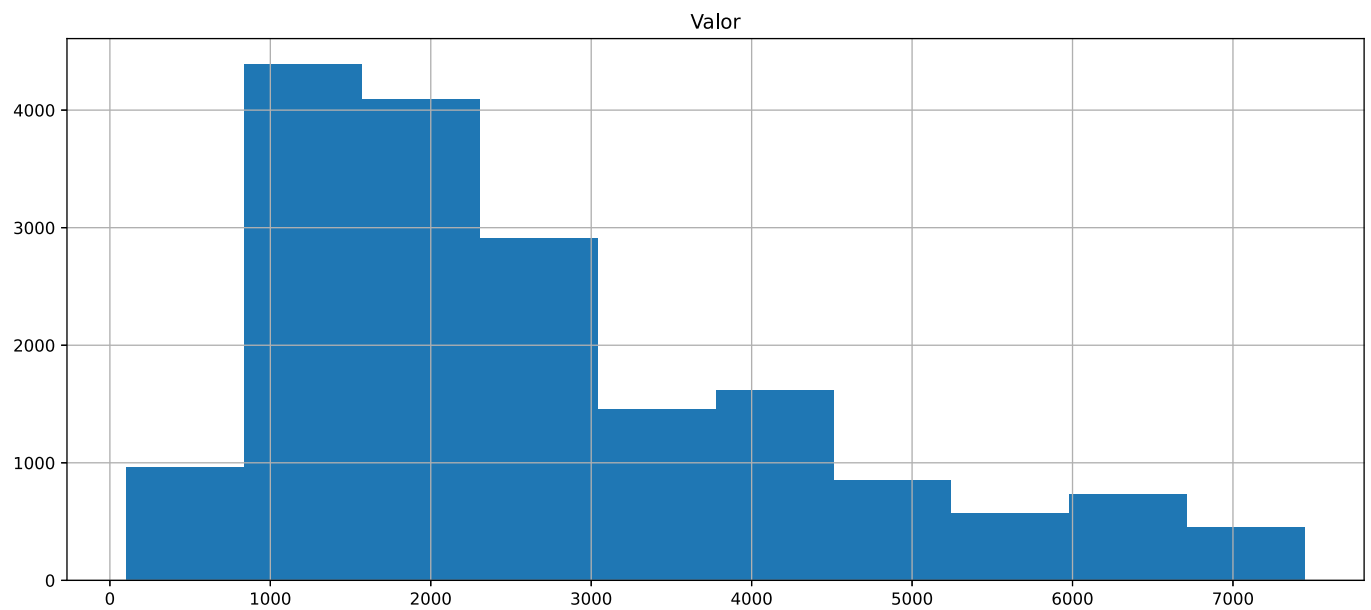
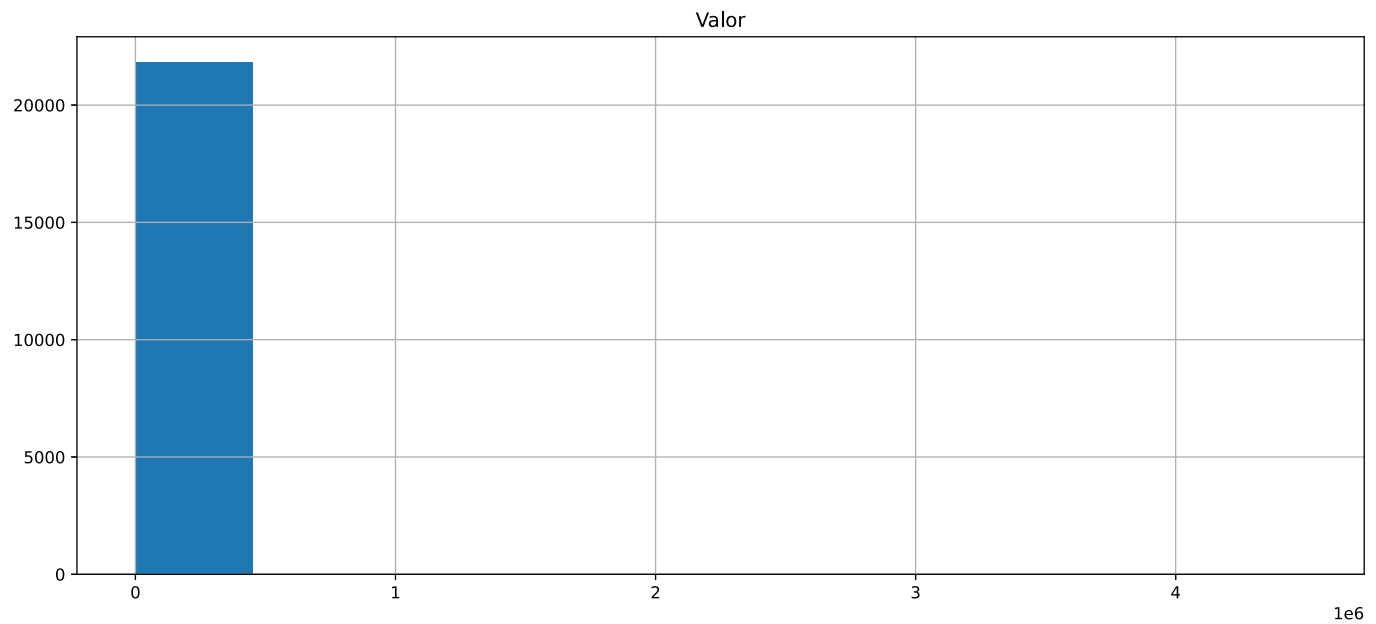
```
In [4]: valor = dados['Valor']
```

```
In [5]: Q1 = valor.quantile(.25)
Q3 = valor.quantile(.75)
IIQ = Q3 - Q1
limite_inferior = Q1 - 1.5 * IIQ
limite_superior = Q1 + 1.5 * IIQ
```

```
In [6]: selecao = (valor >= limite_inferior) & (valor <= limite_superior)
dados_new = dados[selecao]
```

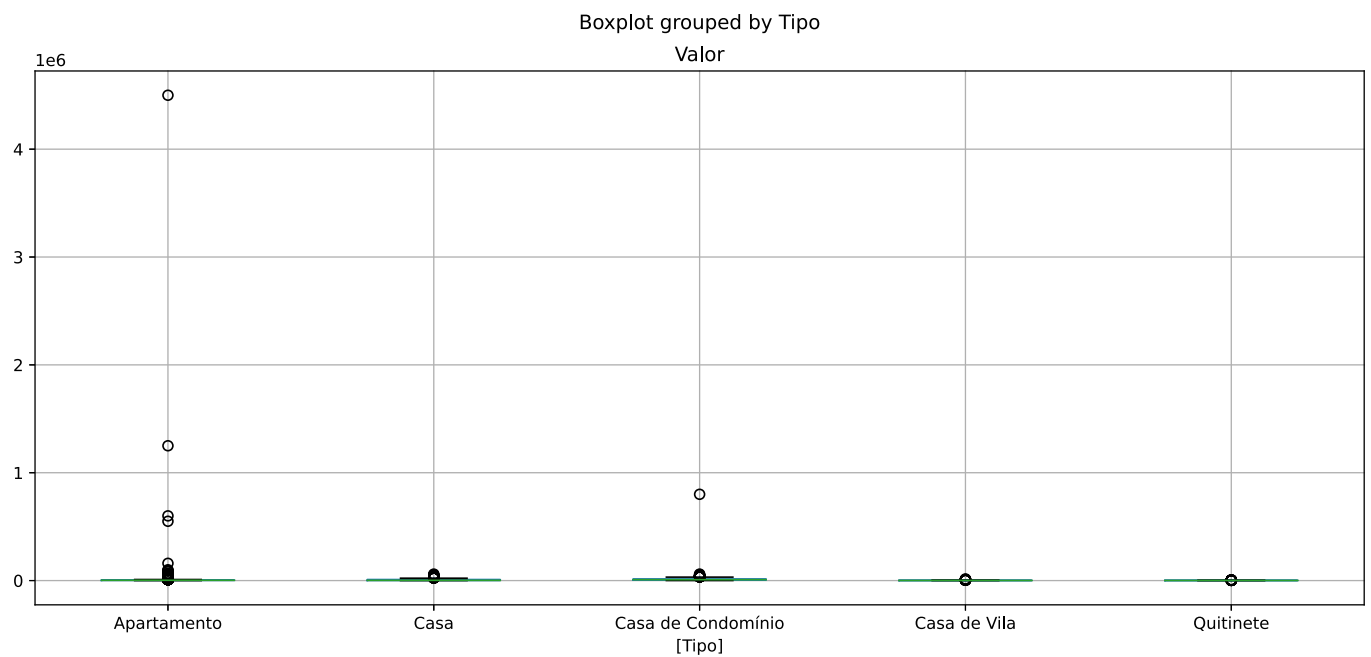
```
In [7]: dados.hist(['Valor'])
dados_new.hist(['Valor'])
```

Out[7]: array([[<AxesSubplot:title={'center':'Valor'}>]], dtype=object)



```
In [8]: dados.boxplot(['Valor'], by = ['Tipo'])
```

```
Out[8]: <AxesSubplot:title={'center': 'Valor'}, xlabel='[Tipo]'
```



```
In [9]: grupo_tipo = dados.groupby('Tipo')['Valor']
        Q1 = grupo_tipo.quantile(.25)
```

```

Q3 = grupo_tipo.quantile(.75)
IIQ = Q3 - Q1
limite_inferior = Q1 - 1.5 * IIQ
limite_superior = Q3 + 1.5 * IIQ

```

In [10]:

```

dados_new = pd.DataFrame()
for tipo in grupo_tipo.groups.keys():
    eh_tipo = dados['Tipo'] == tipo
    eh_dentro_limite = (dados['Valor'] >= limite_inferior[tipo]) & (dados['Valor'] <= limite_superior[tipo])
    selecao = eh_tipo & eh_dentro_limite
    dados_selecao = dados[selecao]
    dados_new = pd.concat([dados_new, dados_selecao])

```

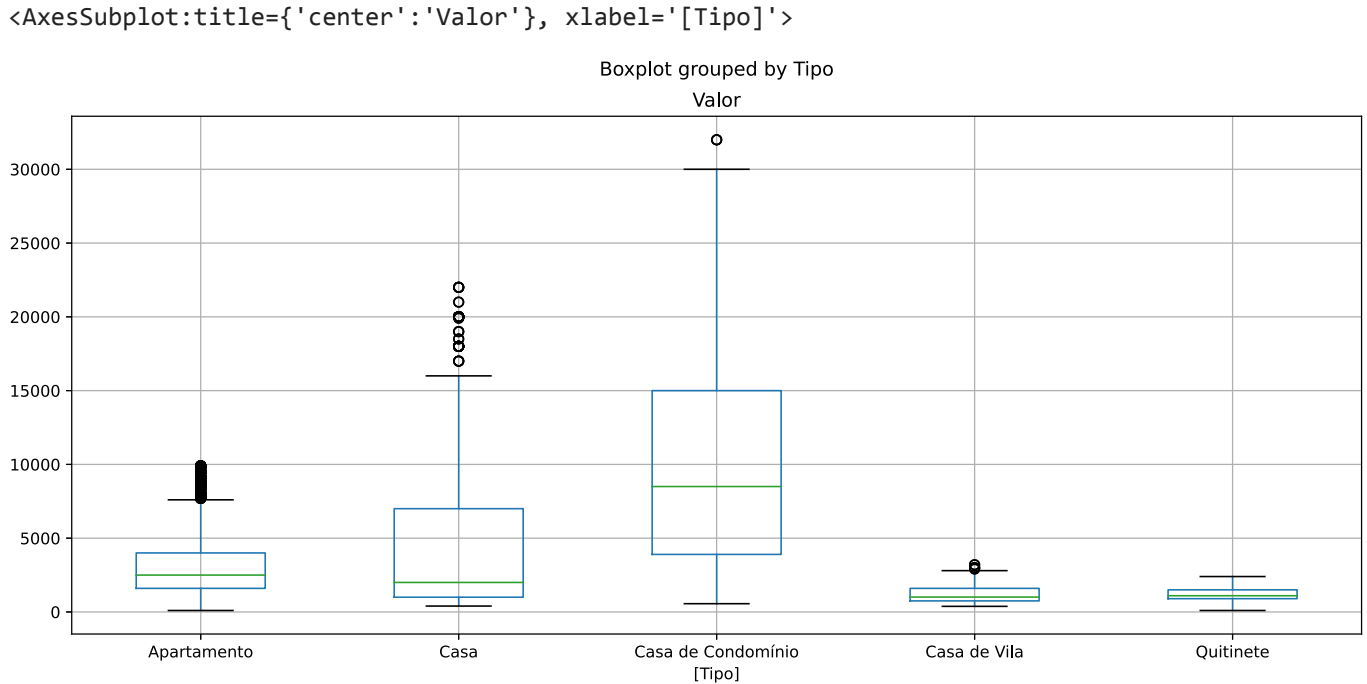
In [11]:

```

dados_new.boxplot(['Valor'], by = ['Tipo'])

```

Out[11]:



In [12]:

```

dados_new.to_csv('dados/aluguel_residencial_sem_outliers.csv', sep = ';', index = False)

```