Relatório de Analise V

Tratamento de dados

In [1]:

```
import pandas as pd
```

In [2]:

```
dados = pd.read_csv('dados/aluguel_residencial.csv',sep=';')
```

In [3]:

```
dados.head(10)
```

Out[3]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
0	Quitinete	Copacabana	1	0	0	40	1700.0	500.0	60.0
1	Casa	Jardim Botânico	2	0	1	100	7000.0	0.0	0.0
2	Apartamento	Centro	1	0	0	15	800.0	390.0	20.0
3	Apartamento	Higienópolis	1	0	0	48	800.0	230.0	0.0
4	Apartamento	Cachambi	2	0	0	50	1300.0	301.0	17.0
5	Casa de Condomínio	Barra da Tijuca	5	4	5	750	22000.0	0.0	0.0
6	Casa de Condomínio	Ramos	2	2	0	65	1000.0	0.0	0.0
7	Apartamento	Grajaú	2	1	0	70	1500.0	642.0	74.0
8	Apartamento	Lins de Vasconcelos	3	1	1	90	1500.0	455.0	14.0
9	Apartamento	Copacabana	1	0	1	40	2000.0	561.0	50.0

In [4]:

```
dados.isnull()
```

Out[4]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
21821	False	False	False	False	False	False	False	False	False
21822	False	False	False	False	False	False	False	False	False
21823	False	False	False	False	False	False	False	False	False
21824	False	False	False	False	False	False	False	False	False
21825	False	False	False	False	False	False	False	False	False

21826 rows × 9 columns

In [5]:

```
dados.notnull()
```

Out[5]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
0	True	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True	True
...	...	...	...	...	...	...	...	...	...
21821	True	True	True	True	True	True	True	True	True
21822	True	True	True	True	True	True	True	True	True
21823	True	True	True	True	True	True	True	True	True
21824	True	True	True	True	True	True	True	True	True
21825	True	True	True	True	True	True	True	True	True

21826 rows × 9 columns

In [6]:

```
dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21826 entries, 0 to 21825
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Tipo        21826 non-null  object 
 1   Bairro      21826 non-null  object 
 2   Quartos     21826 non-null  int64  
 3   Vagas       21826 non-null  int64  
 4   Suites      21826 non-null  int64  
 5   Area        21826 non-null  int64  
 6   Valor       21826 non-null  float64
 7   Condominio  21826 non-null  float64
 8   IPTU        21826 non-null  float64
dtypes: float64(3), int64(4), object(2)
memory usage: 1.5+ MB
```

In [7]:

```
dados['Valor'].isnull()
```

Out[7]:

```
0        False
1        False
2        False
3        False
4        False
         ...
21821    False
21822    False
21823    False
21824    False
21825    False
Name: Valor, Length: 21826, dtype: bool
```

In [8]:

```
dados[dados['Valor'].isnull()]
```

Out[8]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU

In [9]:

```
A = dados.shape[0]
dados.dropna(subset=['Valor'], inplace=True)
B = dados.shape[0]
A - B
```

Out[9]:

```
0
```

In [10]:

```
dados[dados['Valor'].isnull()]
```

Out[10]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU

Tratamento de Dados Faltantes (Continuação)

In [11]:

```
dados[dados['Condominio'].isnull()].head()
```

Out[11]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU

In [12]:

```
dados[dados['Condominio'].isnull()].shape[0]
```

Out[12]:

```
0
```

In [13]:

```
selecao = (dados['Tipo'] == 'Apartamento') & (dados['Condominio'].isnull())
```

In [14]:

```
selecao
```

Out[14]:

```
0        False
1        False
2        False
3        False
4        False
         ...
21821    False
21822    False
21823    False
21824    False
21825    False
Length: 21826, dtype: bool
```

In [15]:

```
A = dados.shape[0]
dados = dados[~selecao]
B = dados.shape[0]
A - B
```

Out[15]:

```
0
```

In [16]:

```
dados[dados['Condominio'].isnull()].shape[0]
```

Out[16]:

```
0
```

In [17]:

```
# dados.fillna(0, inplace= True)
```

In [18]:

```
dados = dados.fillna({'Condominio':0,'IPTU':0})
```

In [19]:

```
dados
```

Out[19]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
0	Quitinete	Copacabana	1	0	0	40	1700.0	500.0	60.0
1	Casa	Jardim Botânico	2	0	1	100	7000.0	0.0	0.0
2	Apartamento	Centro	1	0	0	15	800.0	390.0	20.0
3	Apartamento	Higienópolis	1	0	0	48	800.0	230.0	0.0
4	Apartamento	Cachambi	2	0	0	50	1300.0	301.0	17.0
...	...	...	...	...	...	...	...	...	...
21821	Apartamento	Méier	2	0	0	70	900.0	490.0	48.0
21822	Quitinete	Centro	0	0	0	27	800.0	350.0	25.0
21823	Apartamento	Jacarepaguá	3	1	2	78	1800.0	800.0	40.0
21824	Apartamento	São Francisco Xavier	2	1	0	48	1400.0	509.0	37.0
21825	Apartamento	Leblon	2	0	0	70	3000.0	760.0	0.0

21826 rows × 9 columns

In [20]:

```
dados[dados['Condominio'].isnull()].shape[0]
```

Out[20]:

```
0
```

In [21]:

```
dados[dados['IPTU'].isnull()].shape[0]
```

Out[21]:

```
0
```

In [22]:

```
dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 21826 entries, 0 to 21825
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Tipo        21826 non-null  object 
 1   Bairro      21826 non-null  object 
 2   Quartos     21826 non-null  int64  
 3   Vagas       21826 non-null  int64  
 4   Suites      21826 non-null  int64  
 5   Area        21826 non-null  int64  
 6   Valor       21826 non-null  float64
 7   Condominio  21826 non-null  float64
 8   IPTU        21826 non-null  float64
dtypes: float64(3), int64(4), object(2)
memory usage: 1.7+ MB
```

In [23]:

```
dados.to_csv('dados/aluguel_residencial.csv',sep=';',index=False)
```

In [24]:

```
dados = pd.read_csv('dados/aluguel_residencial.csv',sep=';')
dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21826 entries, 0 to 21825
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Tipo        21826 non-null  object 
 1   Bairro      21826 non-null  object 
 2   Quartos     21826 non-null  int64  
 3   Vagas       21826 non-null  int64  
 4   Suites      21826 non-null  int64  
 5   Area        21826 non-null  int64  
 6   Valor       21826 non-null  float64
 7   Condominio  21826 non-null  float64
 8   IPTU        21826 non-null  float64
dtypes: float64(3), int64(4), object(2)
memory usage: 1.5+ MB
```