

Relatório de Análise V

Tratamento de Dados Faltantes

In [1]:

```
import pandas as pd
```

In [2]:

```
dados = pd.read_csv("dados/aluguel_residencial.csv", sep=";")
dados.head(10)
```

Out[2]:

	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	NaN	NaN
2	Apartamento	Centro	1	0	0	15	800.0	390.0	20.0
3	Apartamento	Higienópolis	1	0	0	48	800.0	230.0	NaN
4	Apartamento	Vista Alegre	3	1	0	70	1200.0	NaN	NaN
5	Apartamento	Cachambi	2	0	0	50	1300.0	301.0	17.0
6	Casa de Condomínio	Barra da Tijuca	5	4	5	750	22000.0	NaN	NaN
7	Casa de Condomínio	Ramos	2	2	0	65	1000.0	NaN	NaN
8	Apartamento	Centro	1	0	0	36	1200.0	NaN	NaN
9	Apartamento	Grajaú	2	1	0	70	1500.0	642.0	74.0

In [3]:

```
dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22580 entries, 0 to 22579
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Tipo        22580 non-null  object
1   Bairro      22580 non-null  object
2   Quartos     22580 non-null  int64
3   Vagas       22580 non-null  int64
4   Suites      22580 non-null  int64
5   Area        22580 non-null  int64
6   Valor       22571 non-null  float64
7   Condominio  20765 non-null  float64
8   IPTU        15795 non-null  float64
dtypes: float64(3), int64(4), object(2)
memory usage: 1.6+ MB
```

In [4]:

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

Out[4]:

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
58	Apartamento	Barra da Tijuca	2	1	1	70	NaN	970.0	68.0
1492	Apartamento	Leme	2	0	0	75	NaN	878.0	NaN
1683	Casa	Campo Grande	3	4	3	363	NaN	NaN	NaN
2012	Apartamento	Botafogo	2	0	0	95	NaN	1010.0	170.0
2034	Apartamento	Copacabana	2	0	0	72	NaN	850.0	NaN
4941	Casa	Campo Grande	3	2	1	100	NaN	NaN	NaN

	Tipo	Bairro	Quartos	Vagas	Suites	Area	Valor	Condominio	IPTU
8568	Apartamento	Leme	2	0	1	75	NaN	878.0	NaN
8947	Apartamento	Glória	3	0	1	135	NaN	910.0	228.0
9149	Apartamento	Gávea	3	1	1	105	NaN	880.0	221.0

```
In [5]: dados.dropna(subset= ['Valor'], inplace=True)
```

```
In [6]: dados[dados['Condominio'].isnull()].shape[0]
```

```
Out[6]: 1813
```

```
In [7]: selecao = (dados['Tipo'] == "Apartamento") & (dados['Condominio'].isnull())
dados = dados[~selecao]
```

```
In [8]: dados.fillna(0, inplace=True)
```

```
In [9]: dados.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 21826 entries, 0 to 22579
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 [10]: dados.to_csv("dados/aluguel_residencial.csv", sep = ";", index=False)
```

```
In [ ]:
```