

‘garagens’, pois é possível que um apartamento não possua vaga. A Figura 26 apresenta o código para tal procedimento.

```
1 def imputer_media(dataset, coluna):
2     mask1 = dataset[coluna] == 0
3     mask2 = dataset[coluna] != 0
4     dataset.loc[mask1, coluna] = dataset.loc[mask2, coluna].mean()
5     return dataset
```

FIGURA 26: Código de imputação de valores da média.

Fonte: O Autor

Ao final, os dados ficaram disponíveis para manipulação, conforme demonstra a Tabela 2.

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

Finalizado o pré-processamento a próxima etapa foi a escolha do modelo de regressor.

Para definição desse modelo, foi necessário verificar, por meio de um diagrama de dispersão, a distribuição dos valores das variáveis no hiperplano, além do índice de correlação entre as variáveis independentes e a dependente, ou seja, a relação entre as variáveis ‘quartos’, ‘banheiros’, ‘garagens’ e ‘areas’ com a variável ‘valor’.

Na Figura 27, são apresentados os diagramas dos valores das variáveis independentes com relação a variável dependente.