

05

## Faça como eu fiz na aula

Aprendemos que os algoritmos de Machine Learning trabalham com números, mas em algumas situações só ter números não é o suficiente, como no caso dos algoritmos de cluster. Ou seja, ainda é necessário tratar esses números, por exemplo, aplicando alguma técnica de normalização.

Vamos supor que uma determinada empresa possua um conjunto de dados chamado **consumo\_clientes.csv**, e quer executar um algoritmo de cluster para extrair alguns padrões dessa base. Porém, nessa base há colunas do tipo texto e números.

As colunas da base **consumo\_clientes.csv** são nomeadas da seguinte forma: `coluna_1` , `coluna_2` , `coluna_3` até `coluna_15` , mas há 5 colunas que não são do tipo número, no caso: `coluna_3` , `coluna_5` , `coluna_8` , `coluna_10` e `coluna_13` .

Desenvolva um script para tratar essa base de dados para que seja possível aplicar um algoritmo de cluster.