

# K-Means



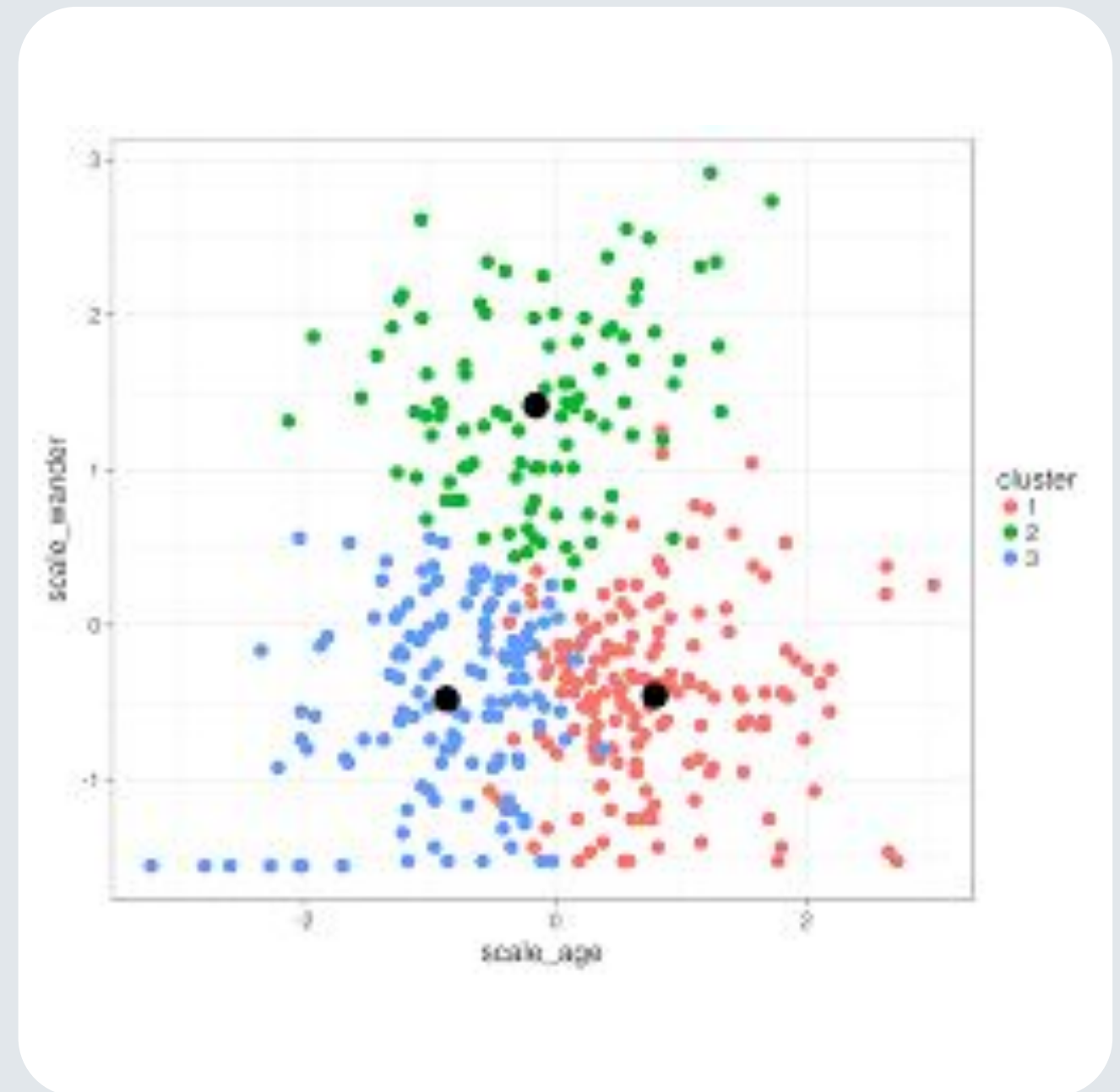
Por que?

Processamento que usa a distância entre os pontos e a centroide do cluster. É um algoritmo não supervisionado e por isso não precisa de um dataset de treinamento e testes.

Matematicamente como acontece?

Ele considera a distância entre dois pontos para saber a qual cluster o ponto pertence, sempre calculando a menor distância entre o ponto e o centroide.

$$C = \frac{1}{A_C} \sum_{i=0}^n \mathbf{c}_i a_i$$





## Vantagens e desvantagens

As vantagens deste algoritmo é que ele não precisa ser previamente treinado.

E uma das desvantagens é que não é indicado quando existem muitas variáveis ou haja outliers na base de dados.

## Na prática

Na prática, a biblioteca Scikit Learn possui modelos prontos para serem treinados, possuindo uma documentação bem escrita para estudo.