



**Universidade Federal do Pampa**

# **Aprendizado de Máquina**

## **Trabalho 2 - K-means X PAM**

Jian Furquim - 161152063

# Agenda:

- Introdução;
- Ambiente;
- Metodologia;
- Resultados;
- Conclusão.

# Introdução:

Este trabalho tem como objetivo principal a realização a implementação dos algoritmos K-means e PAM, a fim de realizar uma comparação entre ambos.

Para tanto foi disponibilizado em aula quatro datasets com pontos em um plano cartesiano que serviram para realizar o estudo.

# Ambiente:

Para a execução dos testes foi usada a seguinte máquina:

- Intel® Core™ i3-6006U CPU @ 2.00GHz × 4
- 7,7 GiB RAM
- Ubuntu 20.04.1 LTS



**matplotlib**

# Metodologia:

Os algoritmos foram implementados de forma que são executados da seguinte maneira:

```
$ python3 kmeans data.dat 3 7
```

Onde:

- kmeans - É o algoritmo que está sendo usado (kmeans ou PAM);
- data.dat - É o dataset com os dados a serem tratados;
- 3 - É o numero de clusters ( k ) a serem usados;
- 7 - É a quantidade de vezes que cada clustering troca de posição(iterações).

# Metodologia:

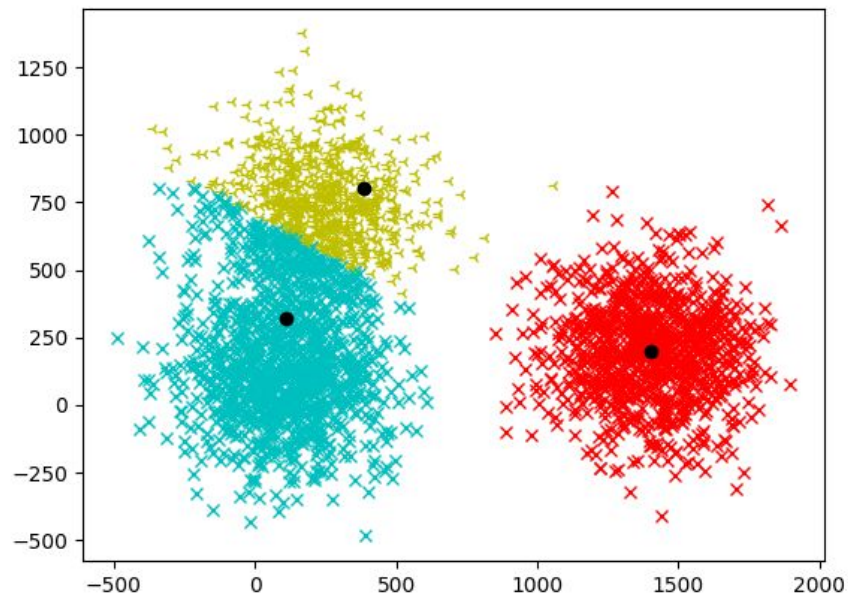
Para fins de comparação os dois algoritmos foram executados:

- Usando 3, 4, 5 e 7 clusters
- Usando 7 troca de posições (iterações).
- Todos rodaram nos datasets:
  - data1.dat
  - data2.dat
  - data3.dat

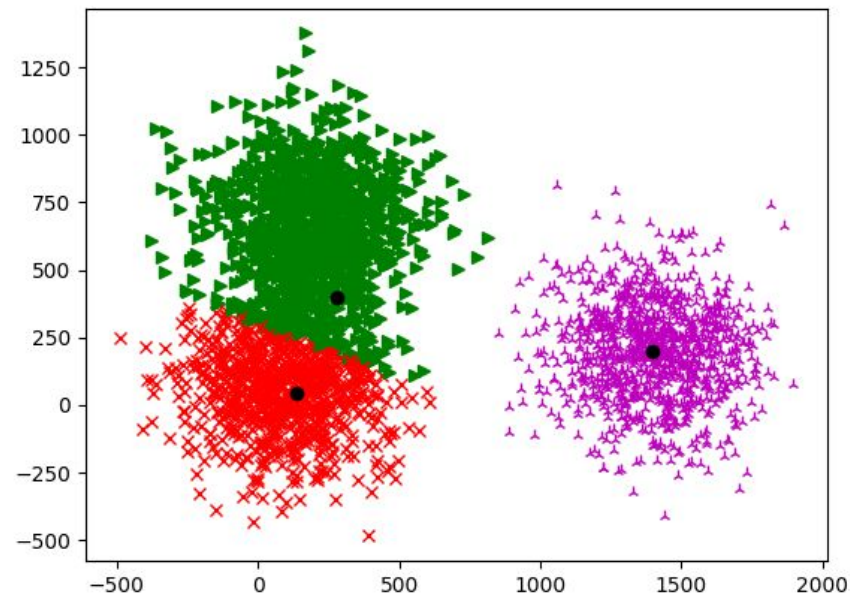
# Resultados:

Dataset data1.dat com 3 clusters

K-means



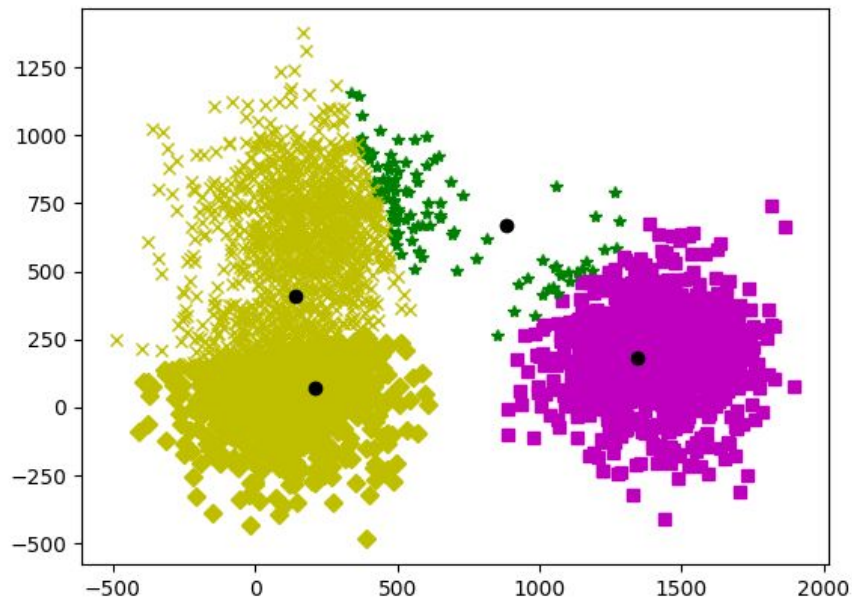
PAM



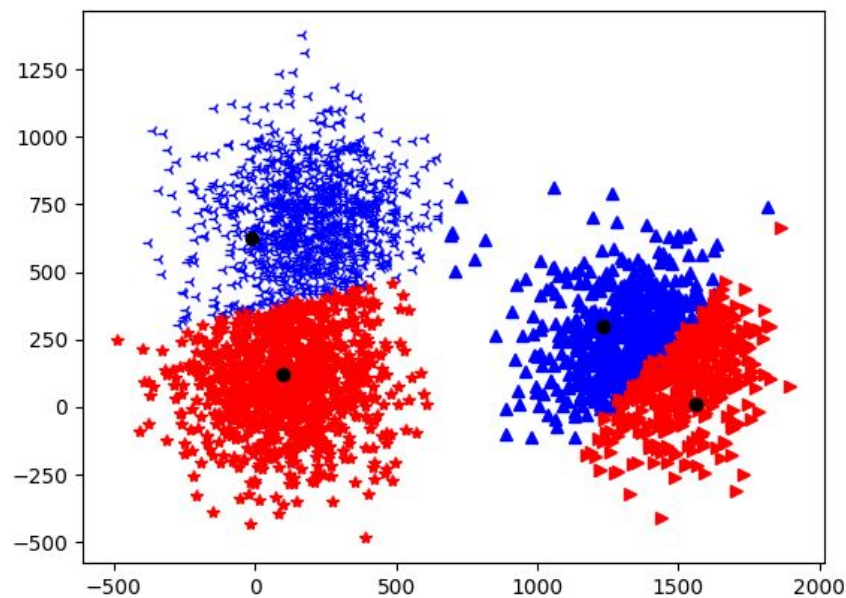
# Resultados:

Dataset data1.dat com 4 clusters

K-means



PAM

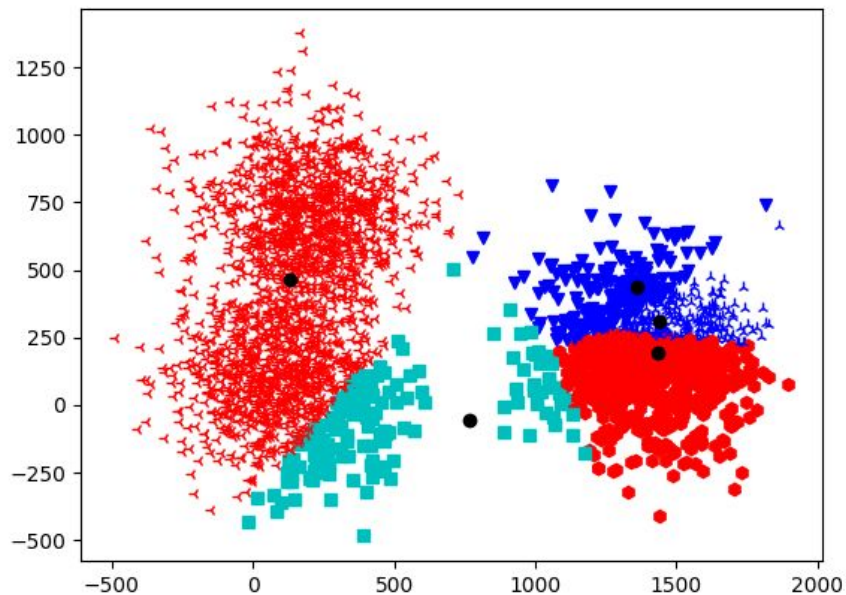




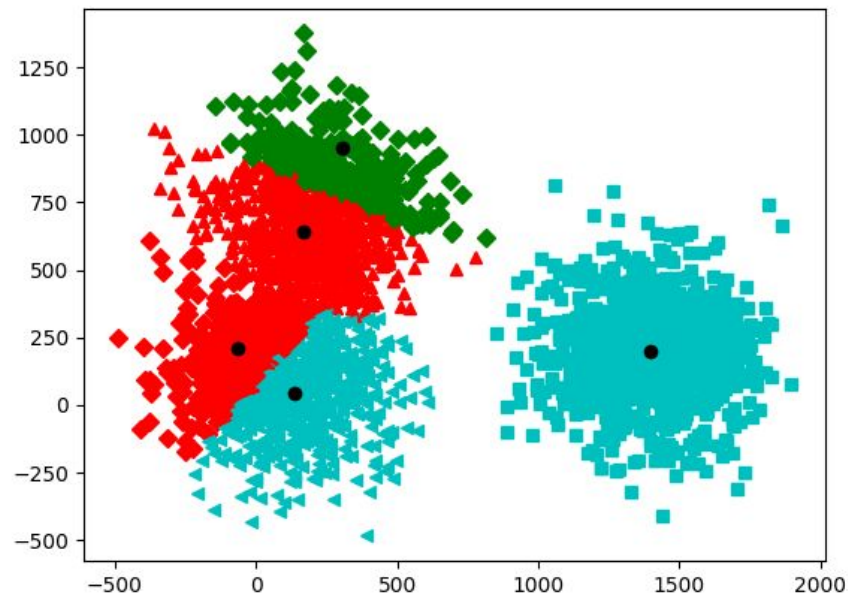
# Resultados:

Dataset data1.dat com 5 clusters

K-means



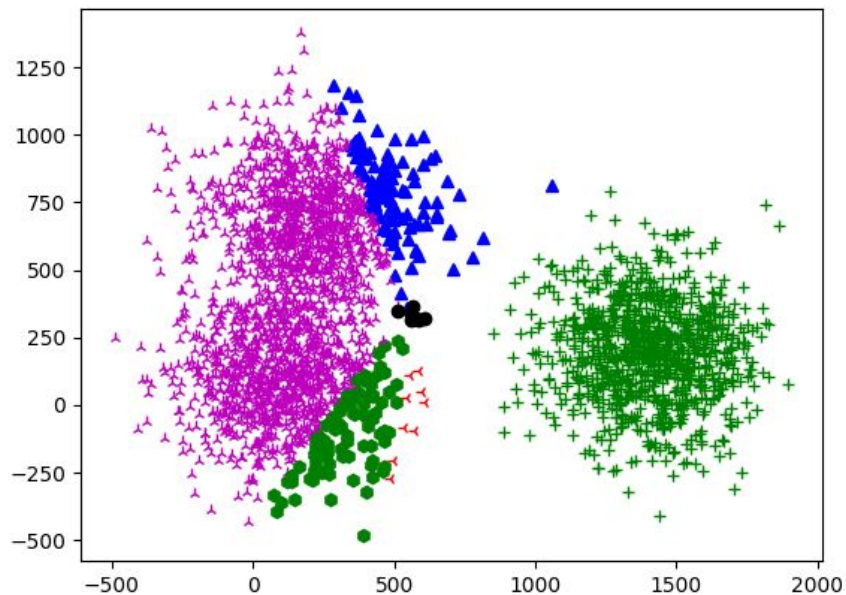
PAM



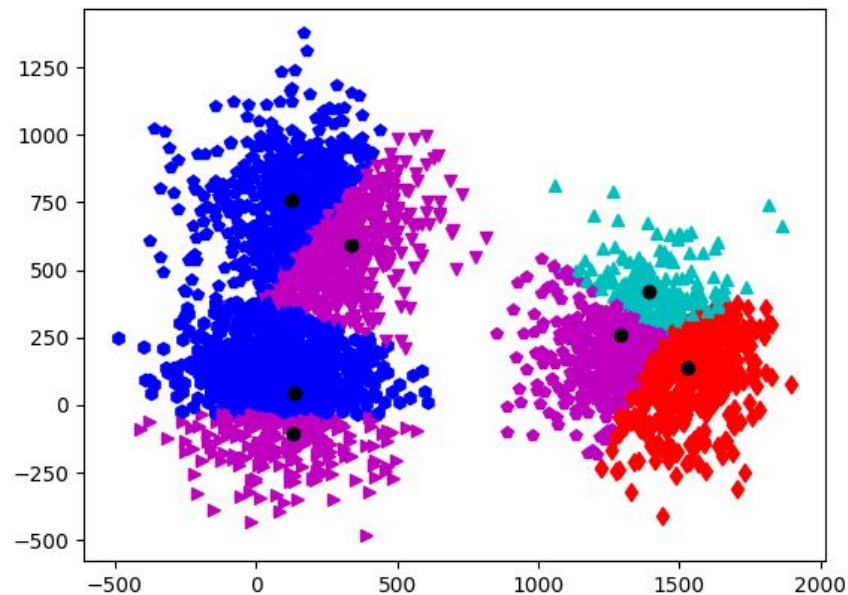
# Resultados:

Dataset data1.dat com 7 clusters

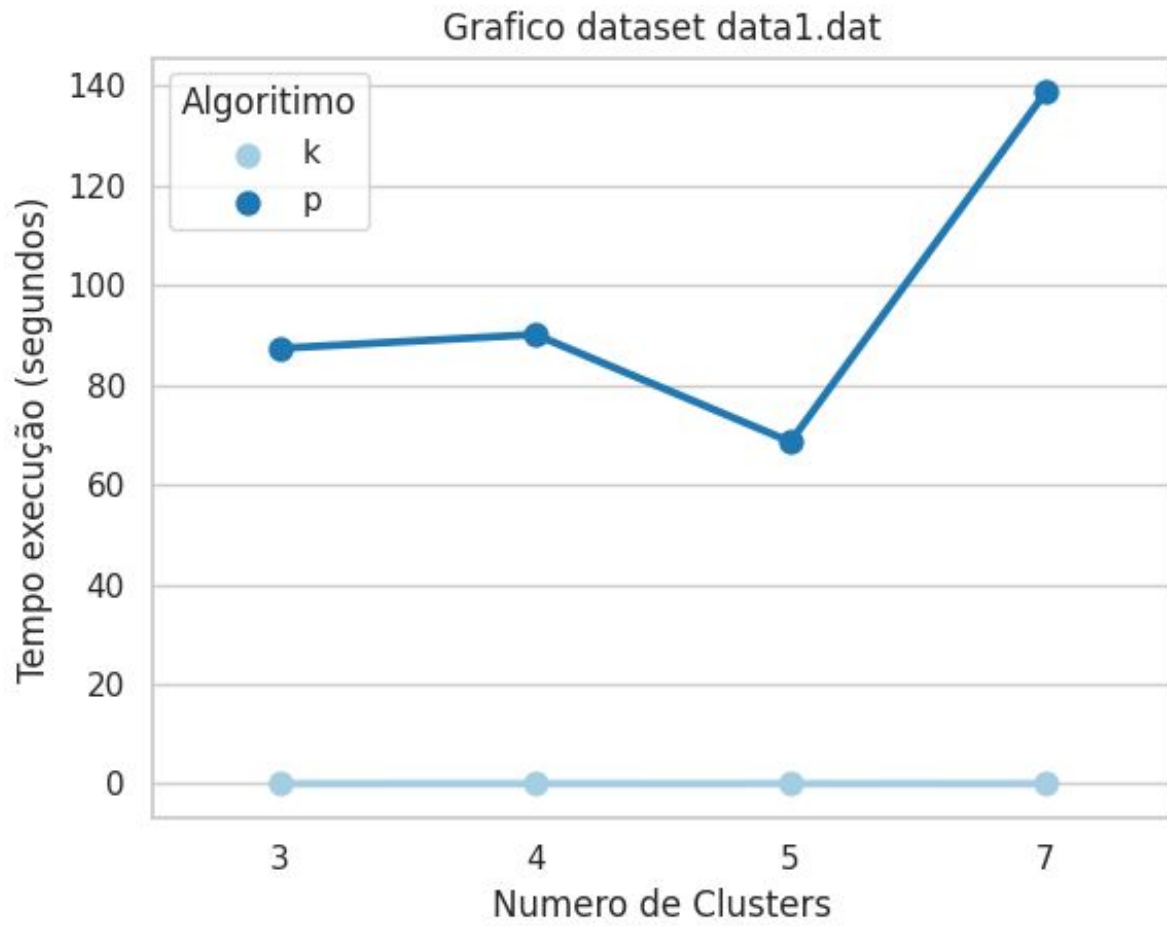
K-means



PAM



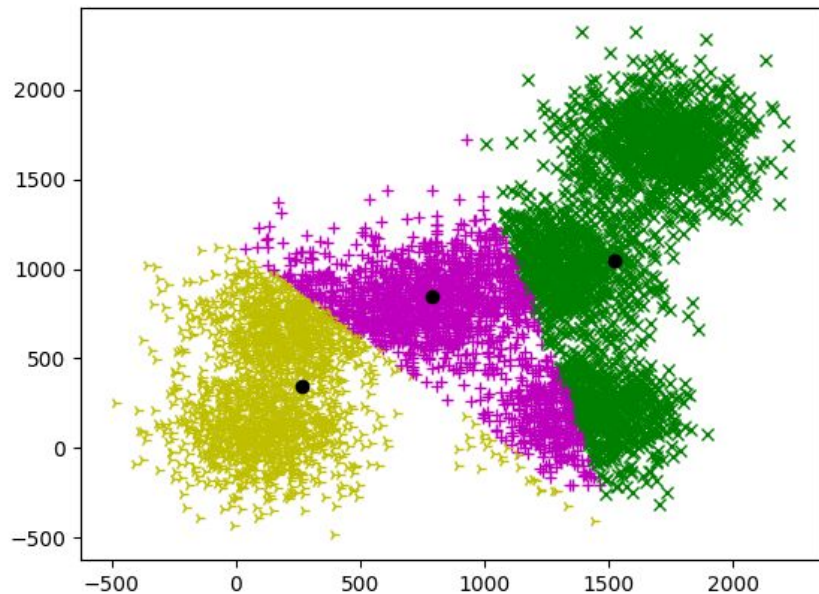
# Resultados:



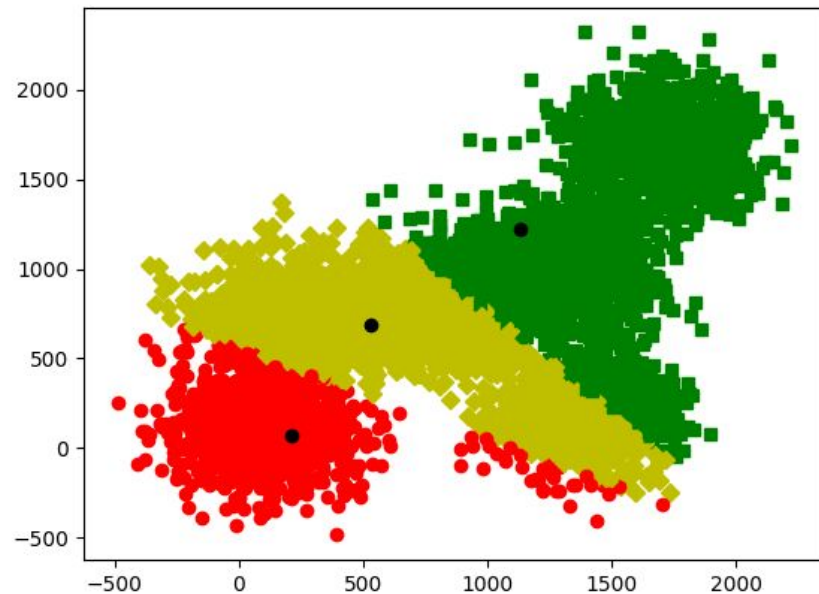
# Resultados:

Dataset data2.dat com 3 clusters

K-means



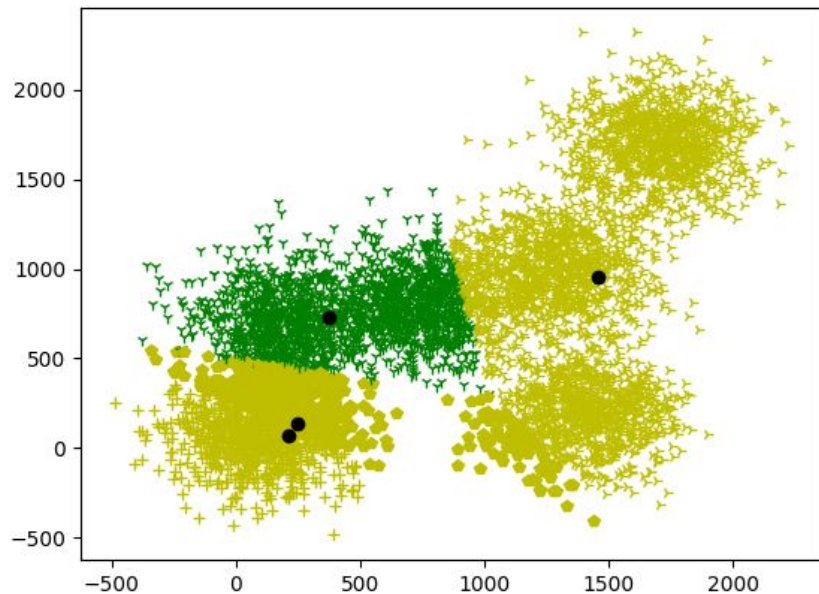
PAM



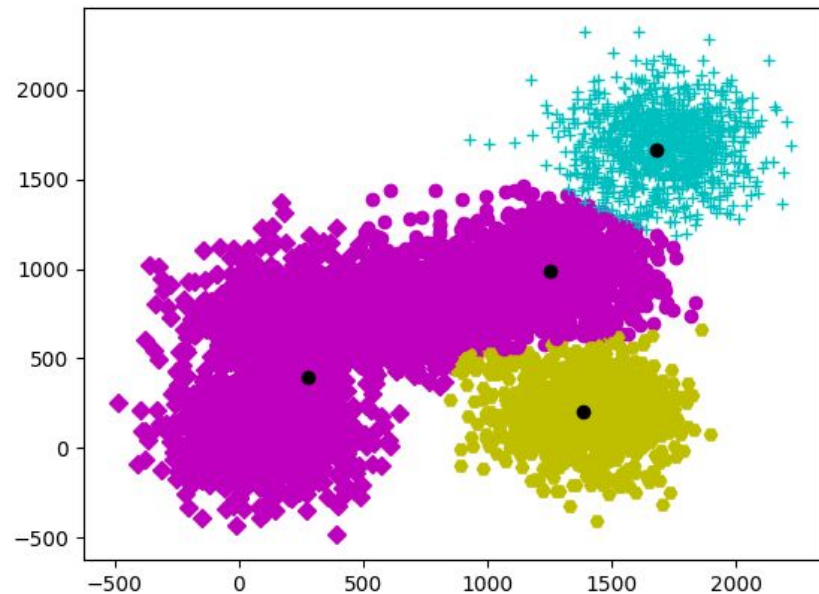
# Resultados:

Dataset data2.dat com 4 clusters

K-means



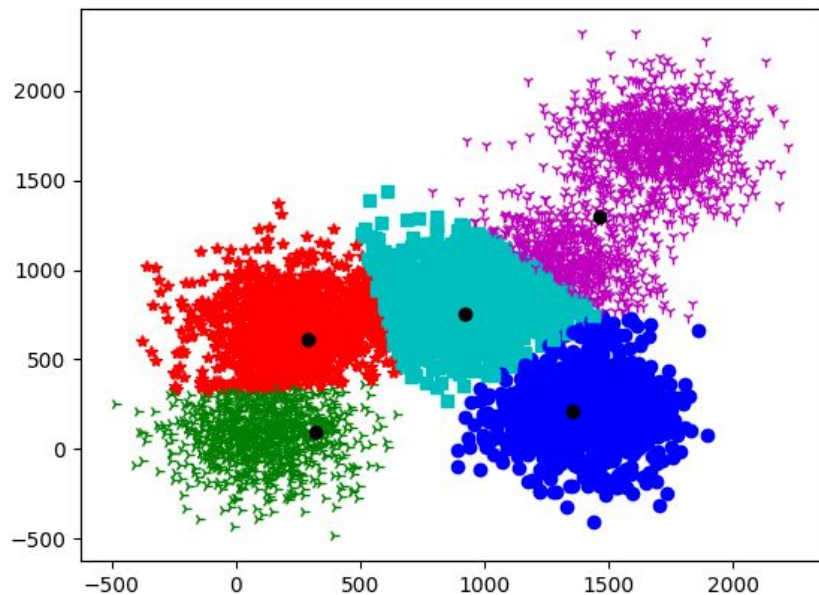
PAM



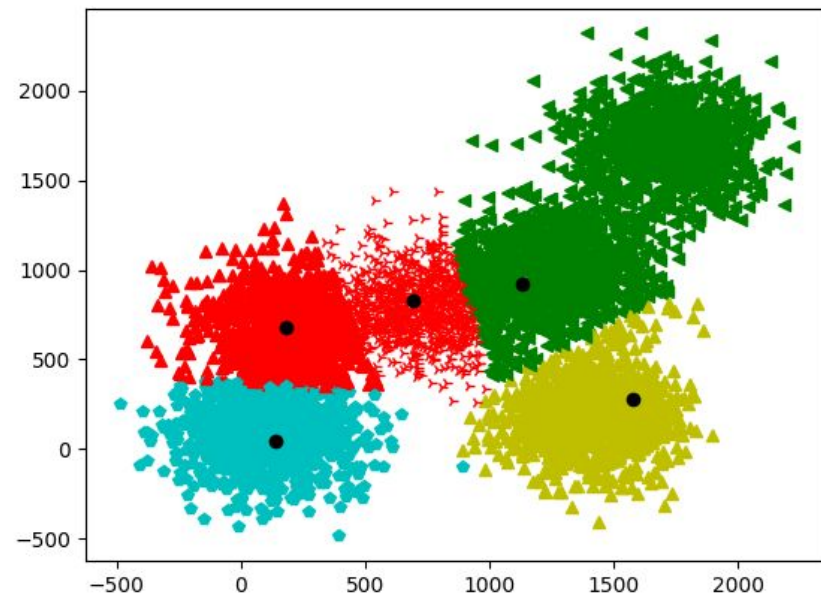
# Resultados:

Dataset data2.dat com 5 clusters

K-means



PAM

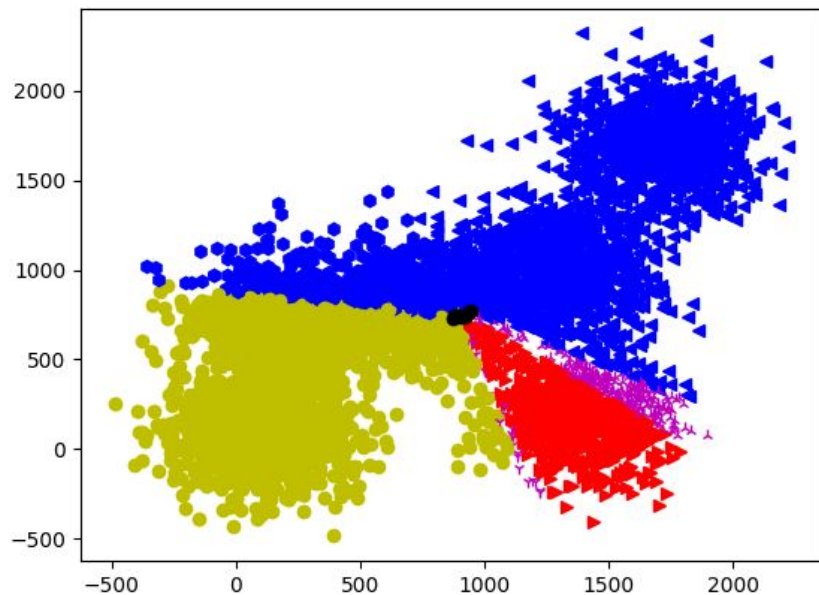




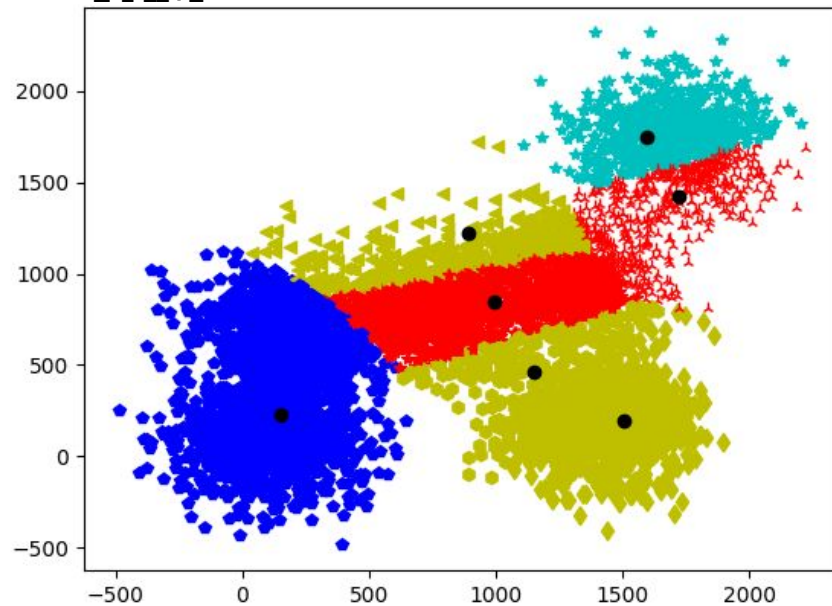
# Resultados:

Dataset data2.dat com 7 clusters

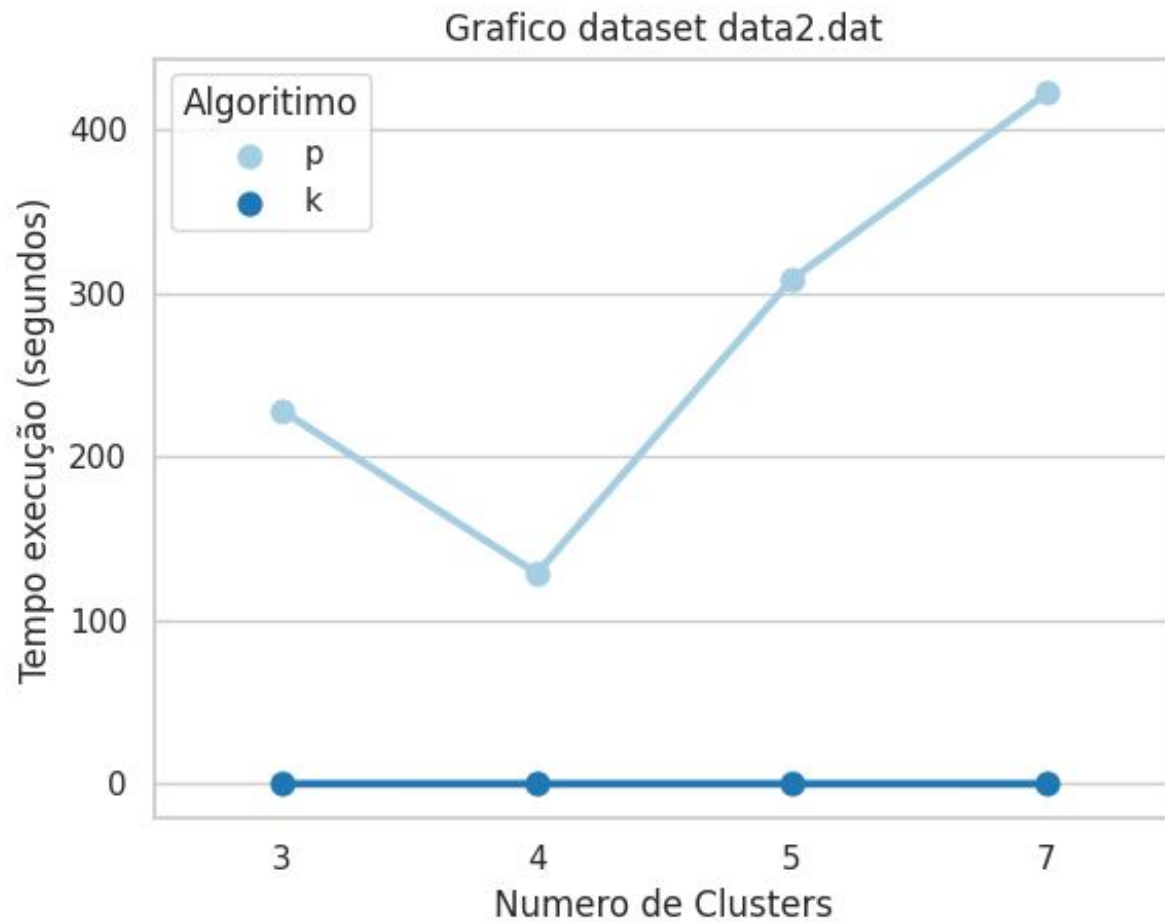
K-means



PAM



# Resultados:

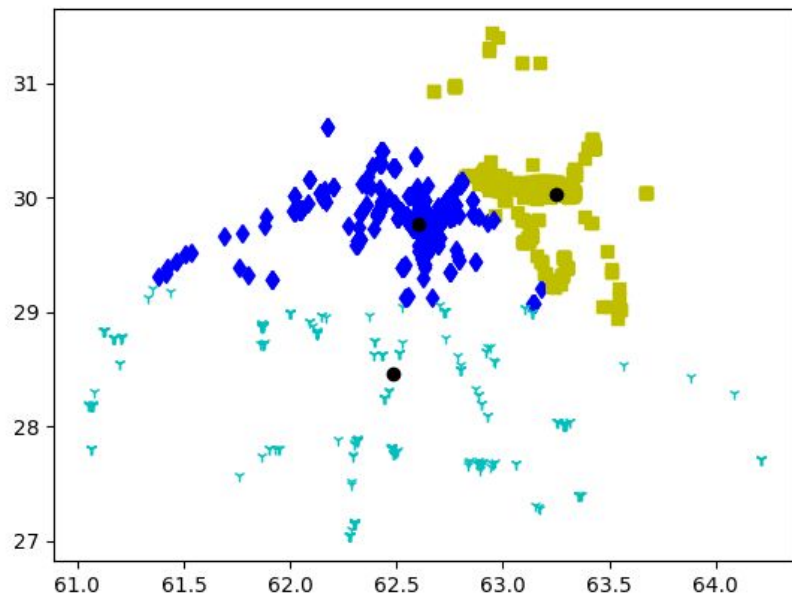




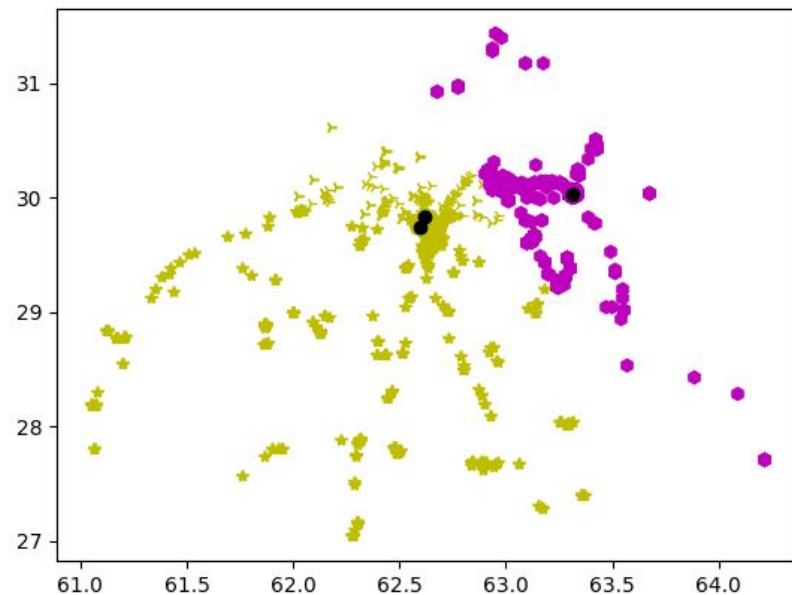
# Resultados:

Dataset data3.dat com 3 clusters

K-means



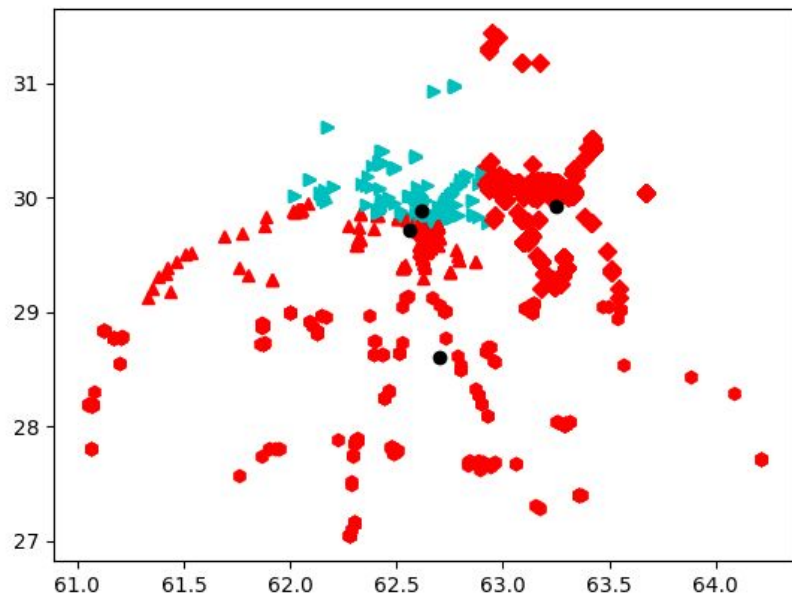
PAM



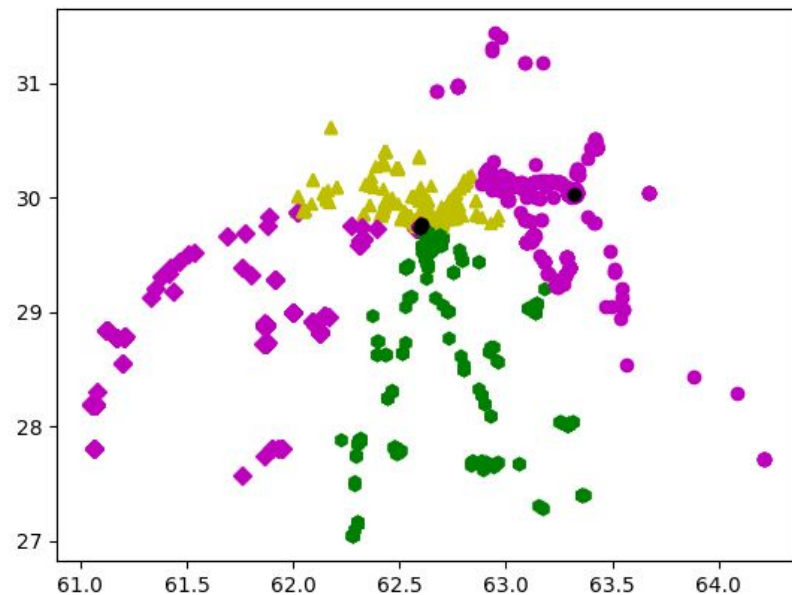
# Resultados:

Dataset data3.dat com 4 clusters

K-means



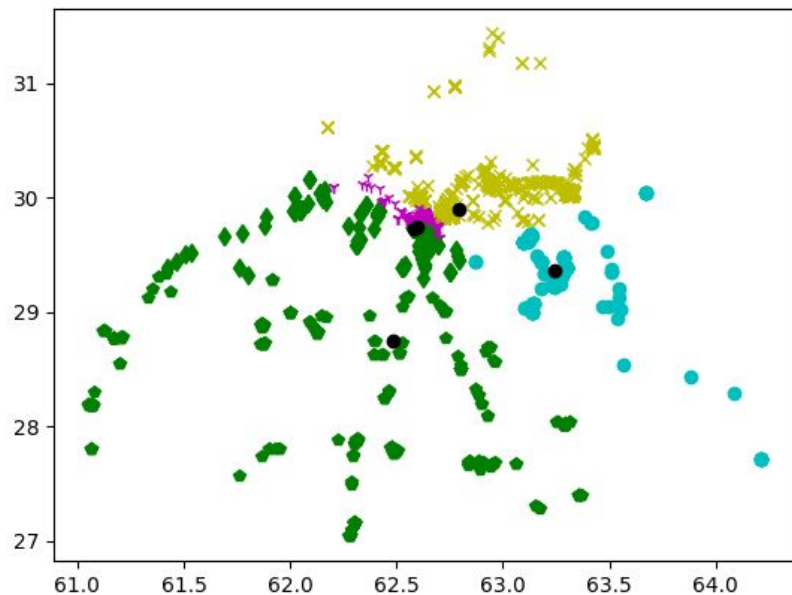
PAM



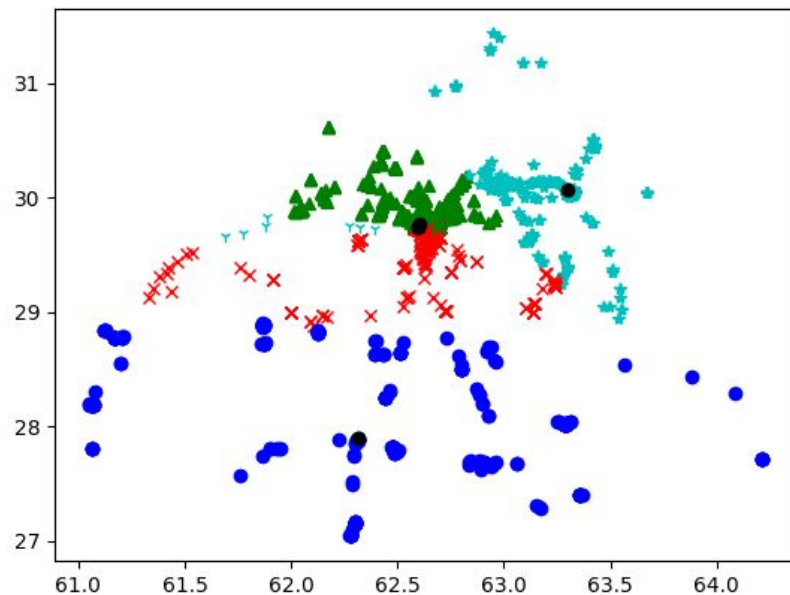
# Resultados:

Dataset data3.dat com 5 clusters

K-means



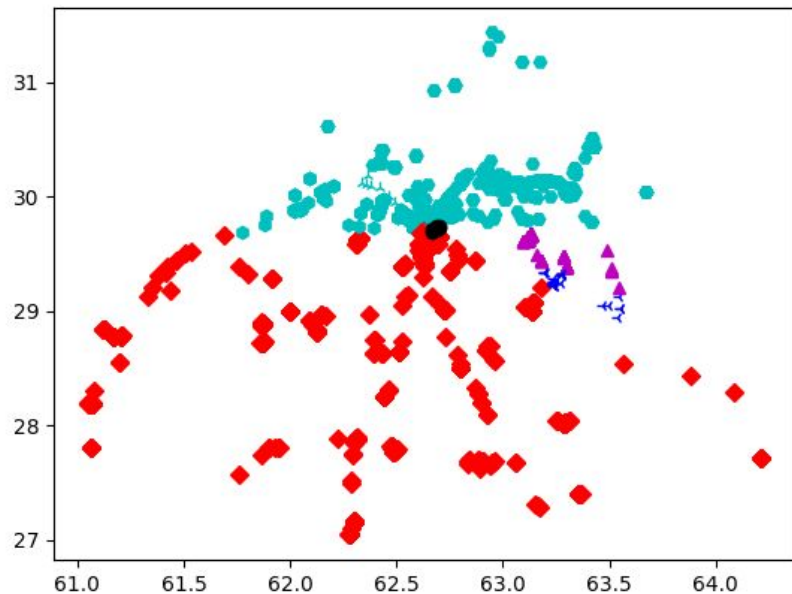
PAM



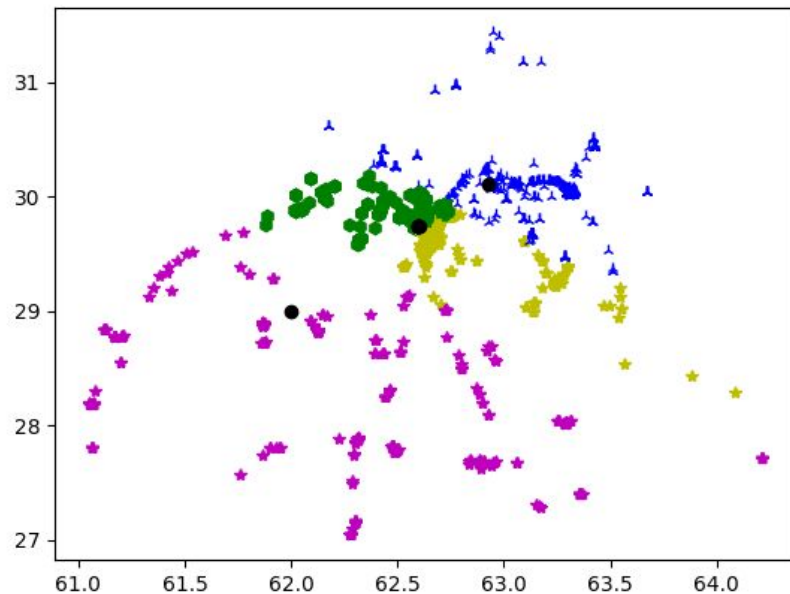
# Resultados:

Dataset data3.dat com 7 clusters

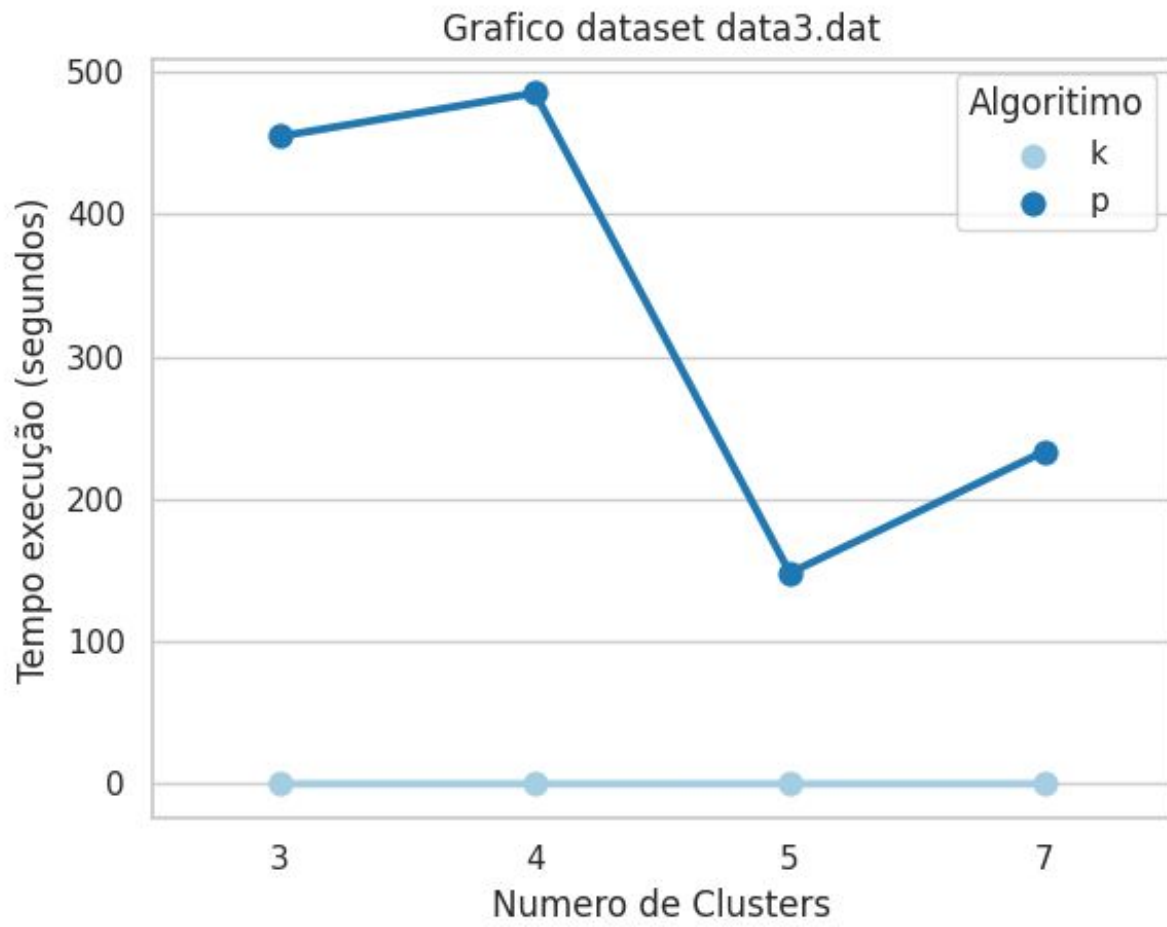
K-means



PAM



# Resultados:



# Conclusão:

- Em relação a tempo o k-means é sempre melhor que o PAM
- Em relação a número de trocas:
  - Se o numero de clusters for alto o PAM tem um melhor desempenho.
  - Se o dataset tiver muitos pontos no mesmo lugar ambos os algoritmos tem um desempenho ruim
- A precisão do PAM é melhor que a do k-means, visto que o PAM usa um ponto existente no dataset como medoid.
- Com um número de execuções (média de tempo) maior a precisão dos resultados seria melhor.

**Obrigado !**

Duvidas ?