


# CLUSTERING







# O QUE É CLUSTERING?



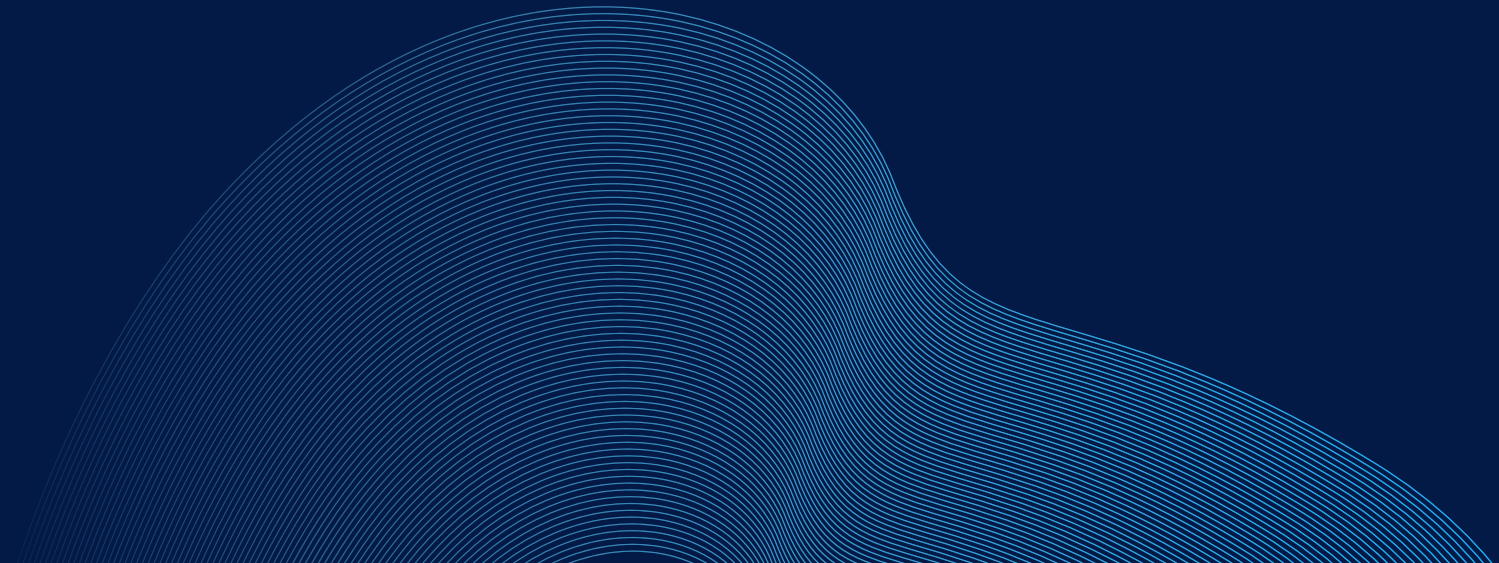
É uma técnica de aprendizado de máquina não supervisionado que tem como objetivo agrupar um conjunto de dados em subgrupos (ou "clusters") com base em suas similaridades.



Em termos simples, ele encontra padrões nos dados e agrupa os itens que são mais parecidos entre si, separando-os daqueles que são diferentes.



Imagine que você tem uma cesta de frutas misturadas, com laranjas, maçãs e bananas. Se você não sabe o nome de cada fruta, pode agrupá-las pela cor, forma, tamanho, doçura, etc.





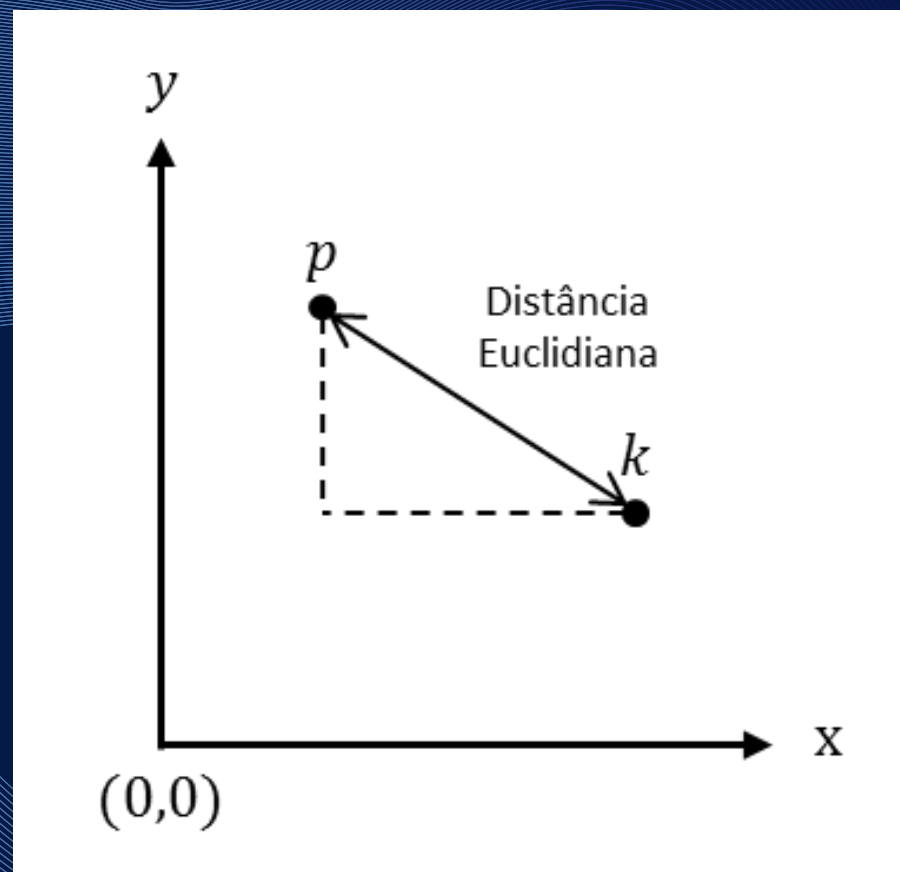
# POR QUE É IMPORTANTE?

- ✦ **Segmentação de clientes:** agrupar consumidores com perfis semelhantes e personalizar campanhas.
- ✦ **Análise de mercado:** identificar produtos comprados juntos e otimizar vendas.
- ✦ **Deteccção de anomalias:** reconhecer fraudes, falhas ou dados atípicos.
- ✦ **Redução de dimensionalidade:** simplificar análises e modelos ao agrupar variáveis correlacionadas.
- ✦ **Bioinformática:** estudar genes com padrões de expressão semelhantes, auxiliando na compreensão de funções e doenças.

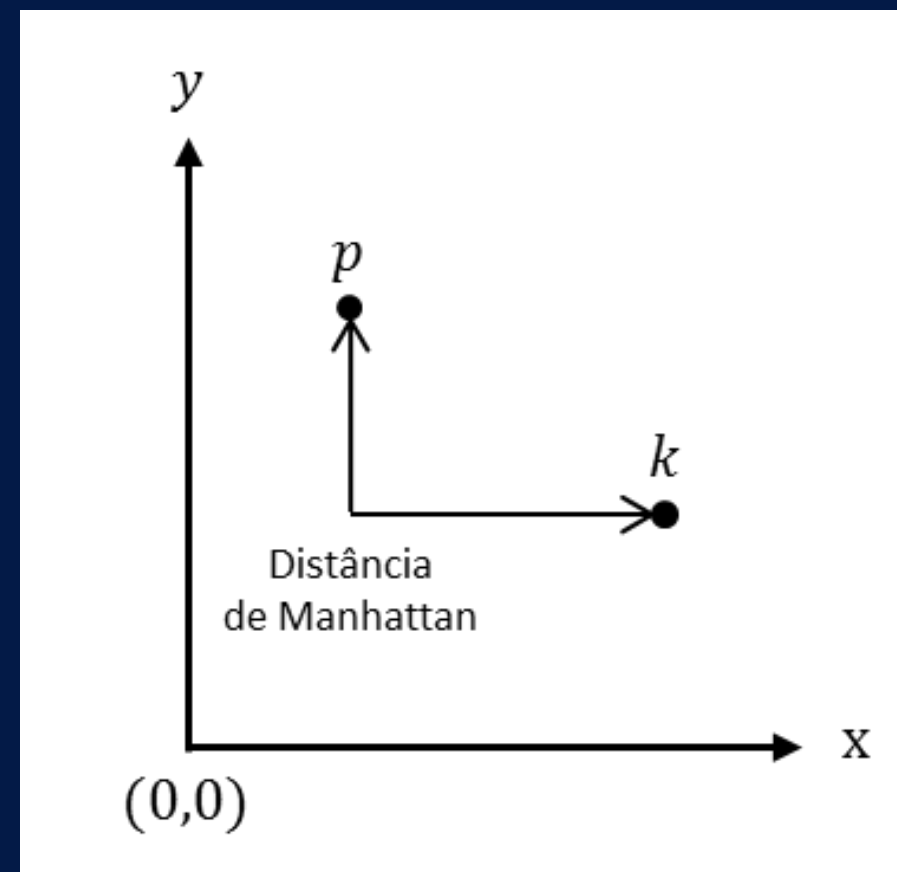


# COMO A SEMELHANÇA É MEDIDA?

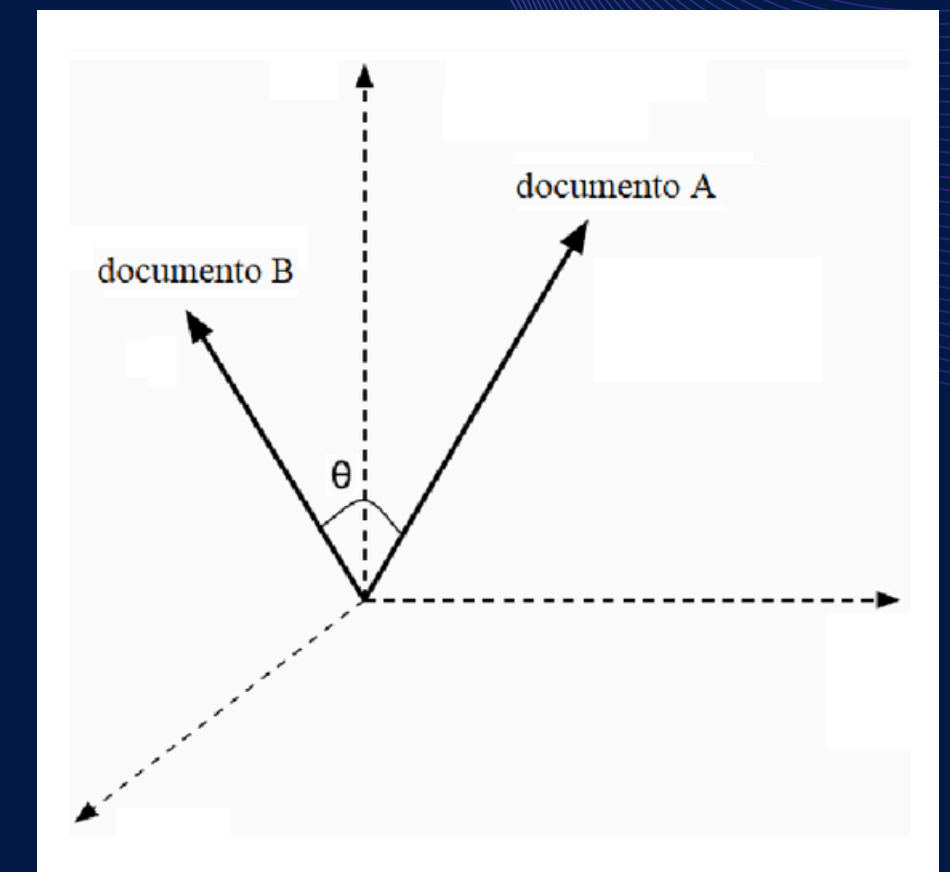
**Distância Euclidiana**



**Distância Manhattan**



**Semelhança do Cosseno**

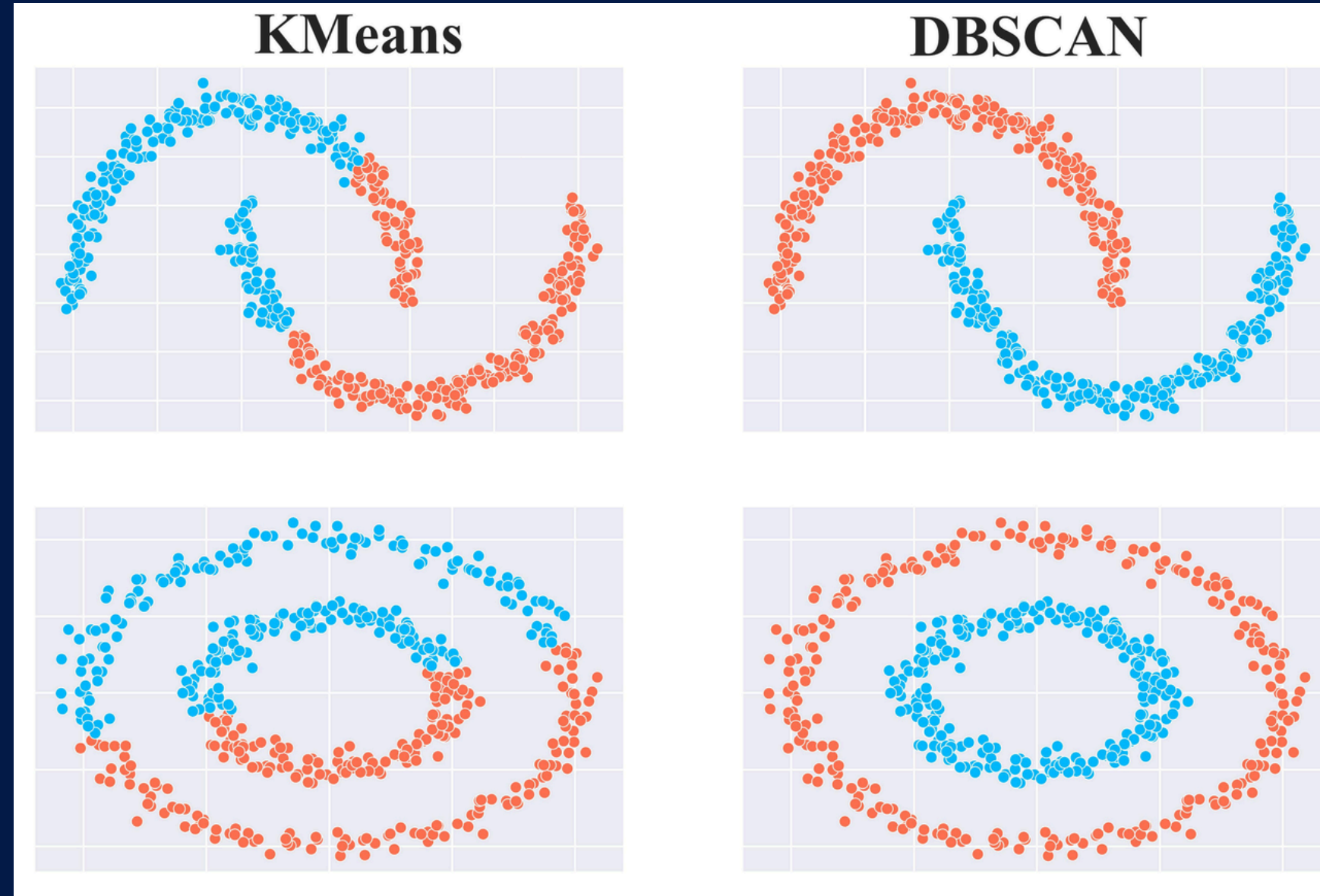


# K-MEANS

- ✦ Selecciona o número de clusters
- ✦ Inicializa os centróides
- ✦ Calcula a distância de um ponto até os centróides
- ✦ Atribui o ponto para o cluster do centróide mais próximo
- ✦ Recalcula o centróide de cada cluster como a média de seus pontos
- ✦ Volta pro 3º passo até os centróides não mudarem mais (ou mudarem muito pouco)



# QUANDO NÃO USAR ESSE ALGORITMO





# QUANDO USAR ESSE ALGORITMO

