



# Clustering Project

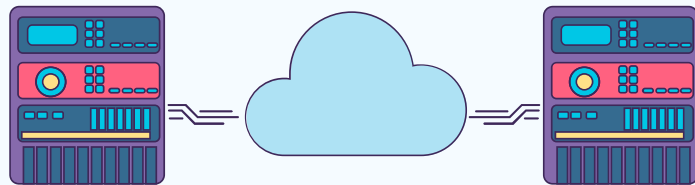
# Introdução

Dataset escolhido: *Wine Data Set*

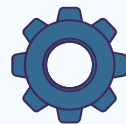
Na filtragem das informações, foram escolhidas duas colunas:

- *Alcohol*
- *Malic Acid*

	Alcohol	Malic acid
0	14.23	1.71
1	13.20	1.78
2	13.16	2.36
3	14.37	1.95
4	13.24	2.59

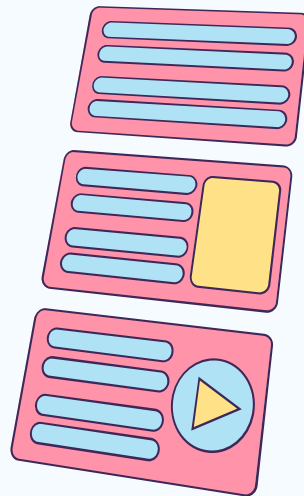


# Fundamentos



Para a análise e tratamento dos dados, os modelos a seguir foram utilizados:

- **KMeans**
  - *Elbow Method*
  - Silhouette Values
- **KMedoids**
- **DBSCAN**



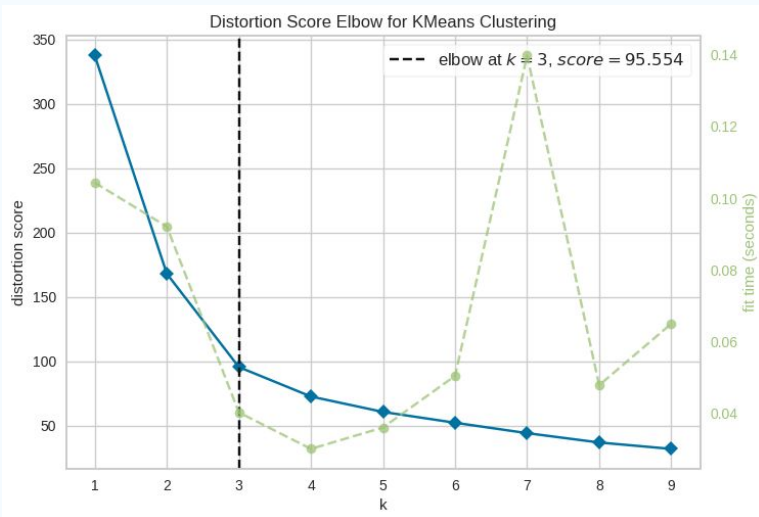
# Metodologia

Em um primeiro momento, o Dataset foi preparado através dos métodos da biblioteca *Pandas*.  
Separamos as colunas necessárias que seriam utilizadas durante o processo



# Descobrendo o K Value

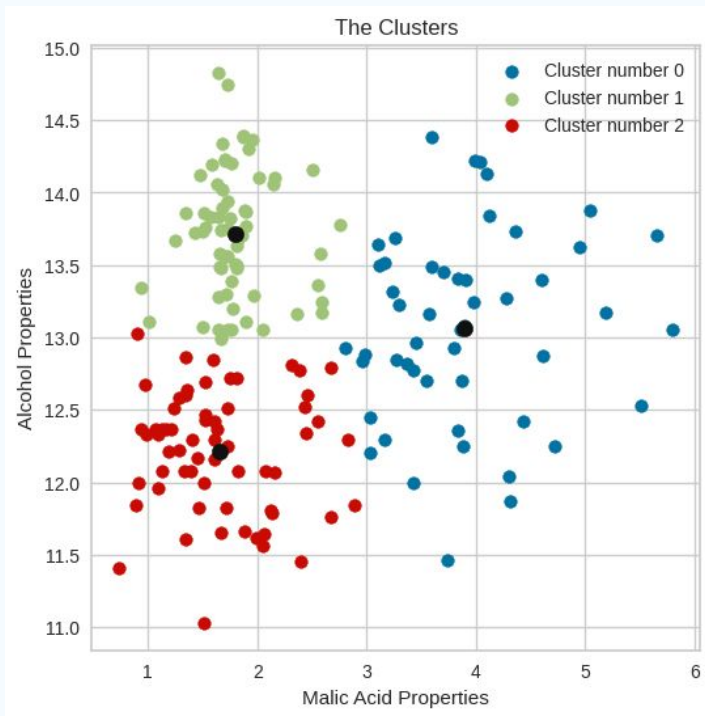
## Elbow Method



## Silhouette Method

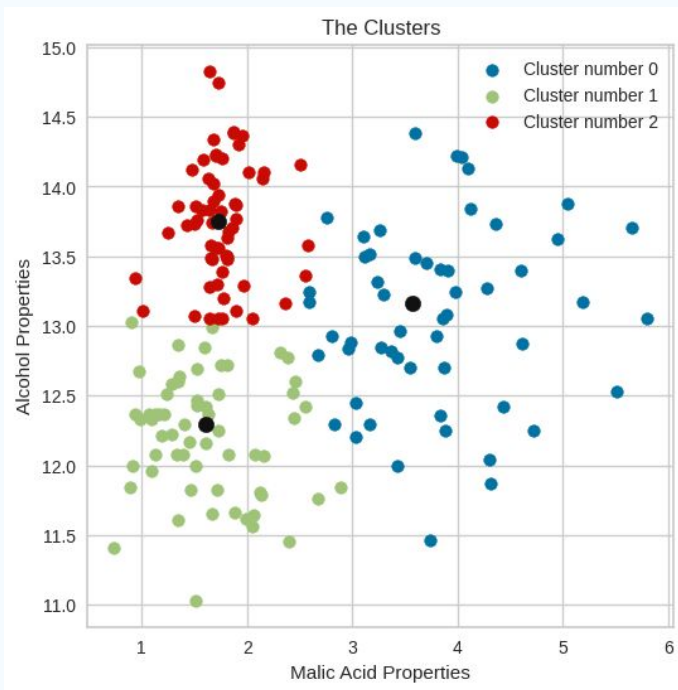
```
The cluster number 2 has a silhouette value equal to 0.4774557823729399
The cluster number 3 has a silhouette value equal to 0.4805357240626079
The cluster number 4 has a silhouette value equal to 0.45639433450218314
The cluster number 5 has a silhouette value equal to 0.44554528655948905
The cluster number 6 has a silhouette value equal to 0.4239584509070041
The cluster number 7 has a silhouette value equal to 0.3830845962295741
The cluster number 8 has a silhouette value equal to 0.3771401689976096
The cluster number 9 has a silhouette value equal to 0.3859616300126425
```

# K Mean



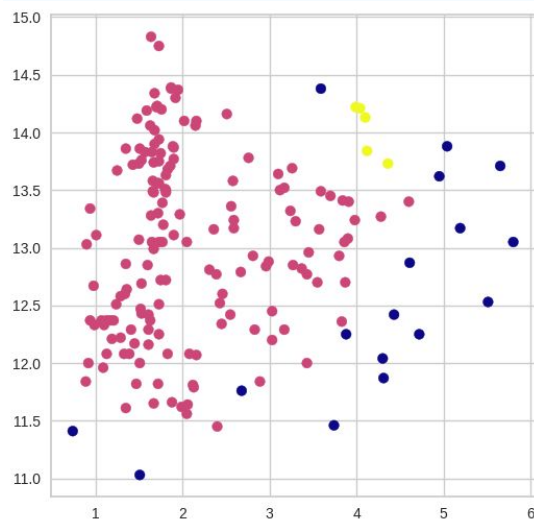
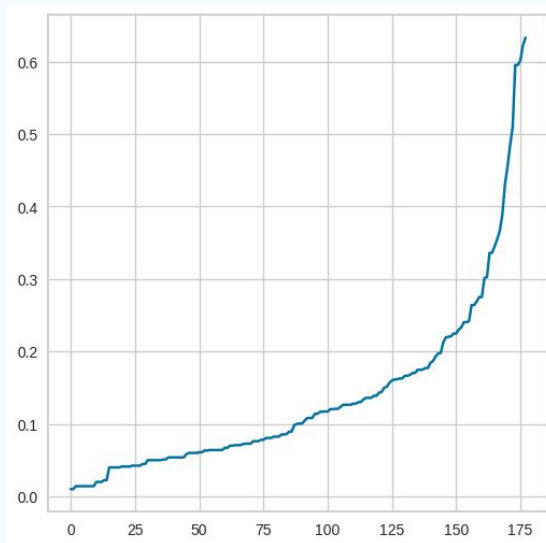
Ao descobrir o valor adequado para aplicar o algoritmo *K Mean* no Dataset, os clusters foram formados a partir dos dados fornecidos

# K Medoids



O método K Medoids é similar na avaliação de clusters, mas apresenta diferenças na sua metodologia e no posicionamento dos centróides

# DBSCAN



- Gráfico para descobrir o valor de epsilon

- Clusters formados utilizando o método DBSCAN



- KMeans e KMedoids podem não ser métodos eficazes

- Espaços *intercluster* e *intracluster* são péssimos!

- Utilizar o DBSCAN pode ser um método mais efetivo

