Evaluation of Clustering Algorithms

# **Clustering tendency**

# **Number of clusters**

Text

Description automatically generated

Diagram

Description automatically generated with low confidence

We want to trade off between these two WCSS.

Chart, line chart

Description automatically generated

After number of clusters=3, there is no significant reduce in the number of clusters.

Data-set

Chart, scatter chart

Description automatically generated

Graphical user interface, text

Description automatically generated

This sequence WCSS is decreasing with very big leaps in the first two steps.  
When each point it-self is a separate cluster (i.e 6 clusters) , we have WCSS=0.

Chart, line chart

Description automatically generated

Cluster=2 🡪 a sub-optimal solution since from cluster-2 to 3 has a giant leap.  
Cluster=3 🡪 optimal solution since there is no much leap.

# **Clustering quality**

Diagram

Description automatically generated

Cohesive and Separation 🡪 <https://cmparlettpelleriti.github.io/CohandSep.html>

Text

Description automatically generated

# **Validity Index**

Validity Index should be minimum for a good cluster

A piece of paper with writing on it

Description automatically generated

A piece of paper with writing on it

Description automatically generated

A picture containing text, whiteboard

Description automatically generated

# **Gain**

A piece of paper with writing on it

Description automatically generated with medium confidence

A white paper with writing on it

Description automatically generated with low confidence

# **Silhouette Coefficient**

Graphical user interface, text, application, email

Description automatically generated

References

<https://www.youtube.com/watch?v=FqIGui0rwh4>

<https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>