Inertia

- Inertia is the sum of squared distances to the closest cluster center (centroid).
- As the number of clusters increases, inertia decreases because data points are closer to their assigned centroids.
- The 'elbow' point indicates an optimal number of clusters where adding more clusters yields diminishing returns in reducing inertia.
- Lower inertia indicates tighter clusters.

Elbow method

- Used to identify optimal k clusters
- Identified where inertia stops decreasing sharply
- Alternatively, use 2nd differential of inertia values to find optimal k.
- It is possible to use silhouette score to identify optimal k.

Silhouette score

- Measures how similar an object is to its own cluster compared to other clusters.
- A higher score indicates better-defined clusters. The plot helps identify the optimal number of clusters where the silhouette score is maximized
    - Silhouette Score for a single data point is the **ratio of the difference between inter-cluster and intra-cluster distances to their maximum**.
    - However, this is extremely computationally expensive.
    - If you have N data points, you will end up calculating $\mathbf{N^2}$ pairwise distances.
    - This gets extremely slow for large data sets and requires a lot of memory.

Calinski-Harabasz Score

- Higher is better → Indicates better defined clusters
- Calinski-Harabasz Score evaluates cluster validity by measuring the ratio of between-cluster variance to within-cluster variance.
- A higher score indicates better-defined clusters. The plot helps identify the optimal number of clusters where this score is maximized.

Davies-Bouldin Score

- Lower is better → Indicates more distinct clusters
- The Davies-Bouldin Score assesses cluster quality by measuring the average similarity ratio of each cluster with its most similar cluster.
- A lower score indicates better-defined clusters. The plot helps identify the optimal number of clusters where this score is minimized.

PCA

- PCA is a dimensionality reduction method.
- PC1 is a linear combinations of original features, with different weights (loadings) attached to each feature (feature space where variance is largest)
- PC2 : orthogonal to PC1, captures second highest variance direction.
- Explained variance ratio : proportion of total variance explained by each principal component

Ex. $PC1 = 0.6 \times Feature1 + 0.3 \times Feature2 - 0.7 \times Feature3$