Sl No	Terminologies	Clustering (Un - Supervised)		Dimension Reduction
		K-Means Clustering	Hierarchial Clustering	Principal Component Analysis
	Key Concepts	**Centroid** **Convergence based on Eucledian Distance** **Elbow Curve (WSS - Within Sum of Squares --> No of Clusters vs WSS)**	**Dendrogram: Hierarchy created from the smallest to largest distance.** **It's of 2 types:** **Agglomerative: Bottoms up / Smallest to Largest distance (preferred)** **Divisive: Top Down**	**Dimension Reduction Technique**
1	**Category**	No Target Variable	No Target Variable	No Target Variable
2	**Linear/Non-Linear**	Linear Data Only. Spectral Clustering for Non Linear Data	Linear Data Only	Linear Data Preferred
3	**Method**	Distance Based	Distance (proximity) Based	
4	**Scaling**	Required	Required	Required
5	**Correlation Check**	Recommended	Recommended	Not Needed
6	**Outlier Treatment**	Required	Required	Required
7	**H/W Requirement**	Low	High	
8	**Reliability**	Data cannot be interpreted due to complexity		
9	**Comments**	# Lazy Learners: Re-Training happens for every prediction # **Works well with Large Datasets** # Results are not scalable and not reproducible as centroids are created randomly # Once clustering is done, we can use the cluster column for classification algorithms	# **Works only with small datasets** but results are scalable and reproducible # High Time Complexity of order n^2 # Rarely Used	# Used for vast datasets # Used in Preprocessing before applying ML Model
10	**Use Cases**	Market Segmentation, Document Clustering	Hard Disk Organization, Social Network Data Storage	Image Compression, Facial Expression Data