**Notebook 1: Introduction to Clustering and Dimensionality Reduction**

**Purpose**: Foundation. Explain concepts, use cases, overview of methods, and a first clustering demo.

**Content**:

* What is clustering? (Markdown + simple illustrations)
* Supervised vs unsupervised learning
* Real-world applications of clustering:
  + Customer segmentation
  + Anomaly detection
  + Image compression
  + Medical diagnosis
* Overview of clustering methods:
  + K-means
  + Hierarchical clustering
  + DBSCAN
* Overview of dimensionality reduction methods:
  + PCA
  + t-SNE
* Evaluation metrics:
  + Silhouette score
  + Davies-Bouldin index
  + Visual inspection (2D projection)
* Hands-on:
  + Simple K-means on 2D toy dataset (make\_blobs)
  + Visualisation of clusters
  + Evaluate with silhouette score

**🔹 Notebook 2: K-Means Clustering**

**Content**:

* How K-means works (steps, objective function)
* Initialisation methods (e.g. k-means++, random)
* Choosing the number of clusters:
  + Elbow method
  + Silhouette analysis
* Hands-on:
  + Apply K-means to synthetic dataset
  + Apply K-means to real-world dataset (e.g. Iris or Mall Customers)
  + Try different number of clusters, graph results
* Visualisations:
  + Clusters and centroids
  + Silhouette plots
* Limitations of K-means (e.g. spherical assumption, sensitivity to scale)

**🔹 Notebook 3: Hierarchical Clustering**

**Content**:

* Types: Agglomerative vs Divisive
* Linkage methods: single, complete, average, Ward
* Dendrograms
* Hands-on:
  + Apply Agglomerative Clustering
  + Visualise dendrogram
  + Use scipy.cluster.hierarchy and sklearn
  + Compare linkage methods
* Real-world dataset (e.g. seeds or gene expression)
* Visual comparison with K-means

**🔹 Notebook 4: DBSCAN and Density-Based Clustering**

**Content**:

* Motivation for density-based methods
* How DBSCAN works: eps and min\_samples
* Strengths (arbitrary shapes, noise handling)
* Weaknesses (parameter sensitivity)
* Hands-on:
  + DBSCAN on non-globular clusters (e.g. make\_moons)
  + Tuning eps and min\_samples, visualising effects
  + Compare with K-means
  + Use silhouette or other metrics
* Real-world example: anomaly detection or spatial data

**🔹 Notebook 5: Dimensionality Reduction: PCA**

**Content**:

* Why reduce dimensions?
* PCA intuition (variance, orthogonal components)
* Mathematical idea (optional visual summary)
* Hands-on:
  + PCA on high-dimensional dataset
  + Scree plot, explained variance
  + Visualise in 2D (before and after clustering)
  + Combine PCA + K-means for better performance

**🔹 Notebook 6: Dimensionality Reduction: t-SNE**

**Content**:

* Intuition behind t-SNE
* Difference from PCA (non-linear, local structure)
* Perplexity and learning rate
* Hands-on:
  + t-SNE on high-dimensional data (e.g. digits)
  + Visualise clusters in 2D
  + Combine with K-means or DBSCAN
* Warnings about misuse (over-interpretation)

**🔹 Notebook 7 (Optional): Clustering Evaluation and Comparison**

**Content**:

* Silhouette Score
* Davies-Bouldin index
* Calinski-Harabasz index
* Adjusted Rand Index (for labelled datasets)
* Side-by-side clustering comparison:
  + Run K-means, Agglomerative, DBSCAN on same data
  + Evaluate and compare scores
  + Visualise results