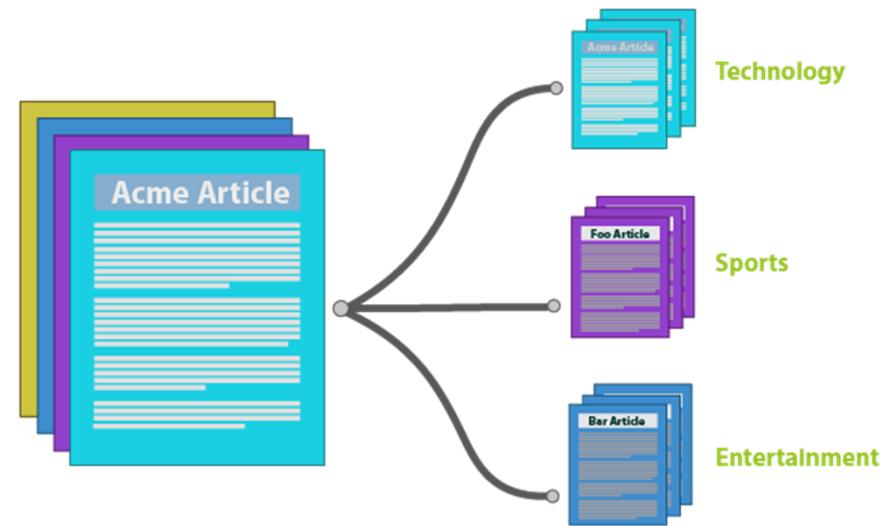


Applications

1. Document Classification

- Cluster documents in multiple categories based on **tags**, **topics** , and **the content** of the document.





2. Delivery Store Optimization

- Optimize the process of good delivery using truck drones by using a **combination of k-means** to find the optimal number of launch locations.

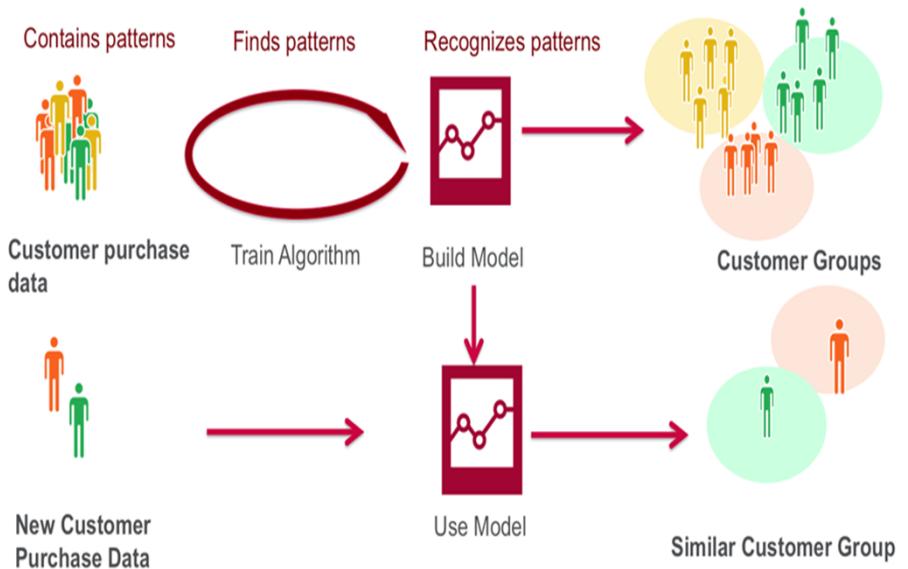
3. Identifying Crime Localities

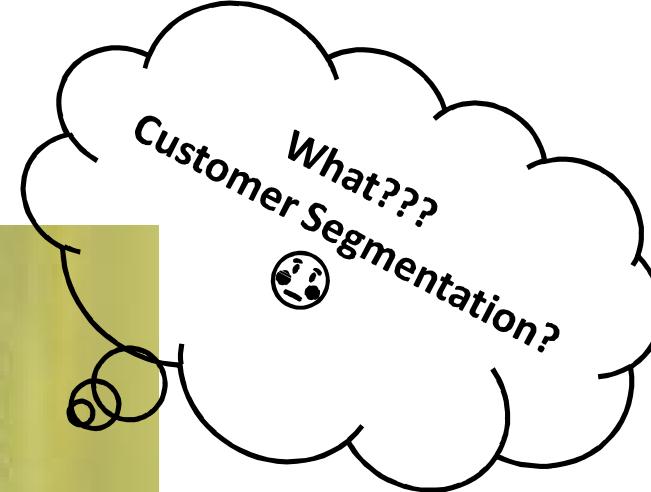
- With data related to crimes available in specific **localities in a city, the category of crime, the area of the crime**, and the association between the two can give quality insight into crime-prone areas within a city or a locality.



4. Customer Segmentation

- Clustering helps marketers improve their **customer base, work on target areas, and segment customers** based on purchase history, interests, or activity monitoring.
- The classification would help the company target **specific clusters of customers** for specific campaigns.





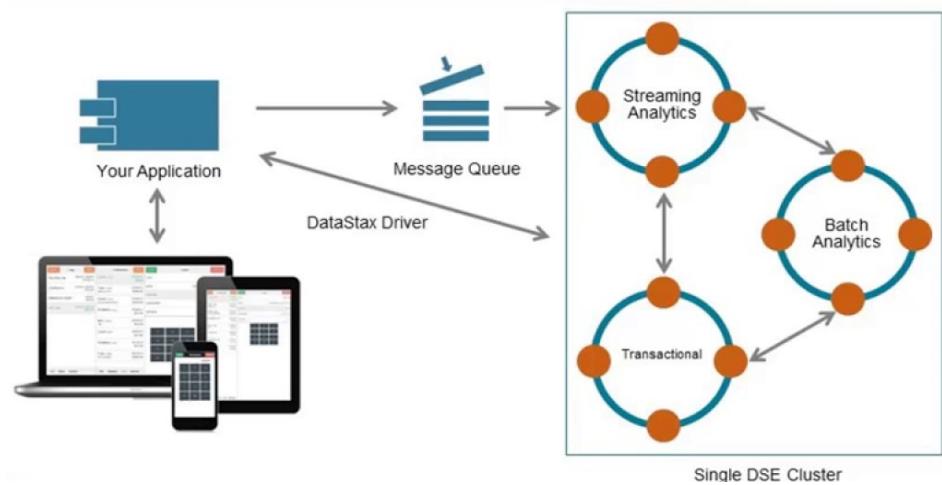
- Netflix **recommends** several other similar shows to watch **by Simply clustering observations together based on similarity.**

Because you watched Pine Gap



5. Insurance Fraud Detection

- Utilizing past historical data on fraudulent claims, it is possible to isolate new claims based on its proximity to **clusters** that indicate **fraudulent patterns**.
- Since insurance fraud can potentially have a multi-million dollar impact on a company, **the ability to detect frauds is crucial**.



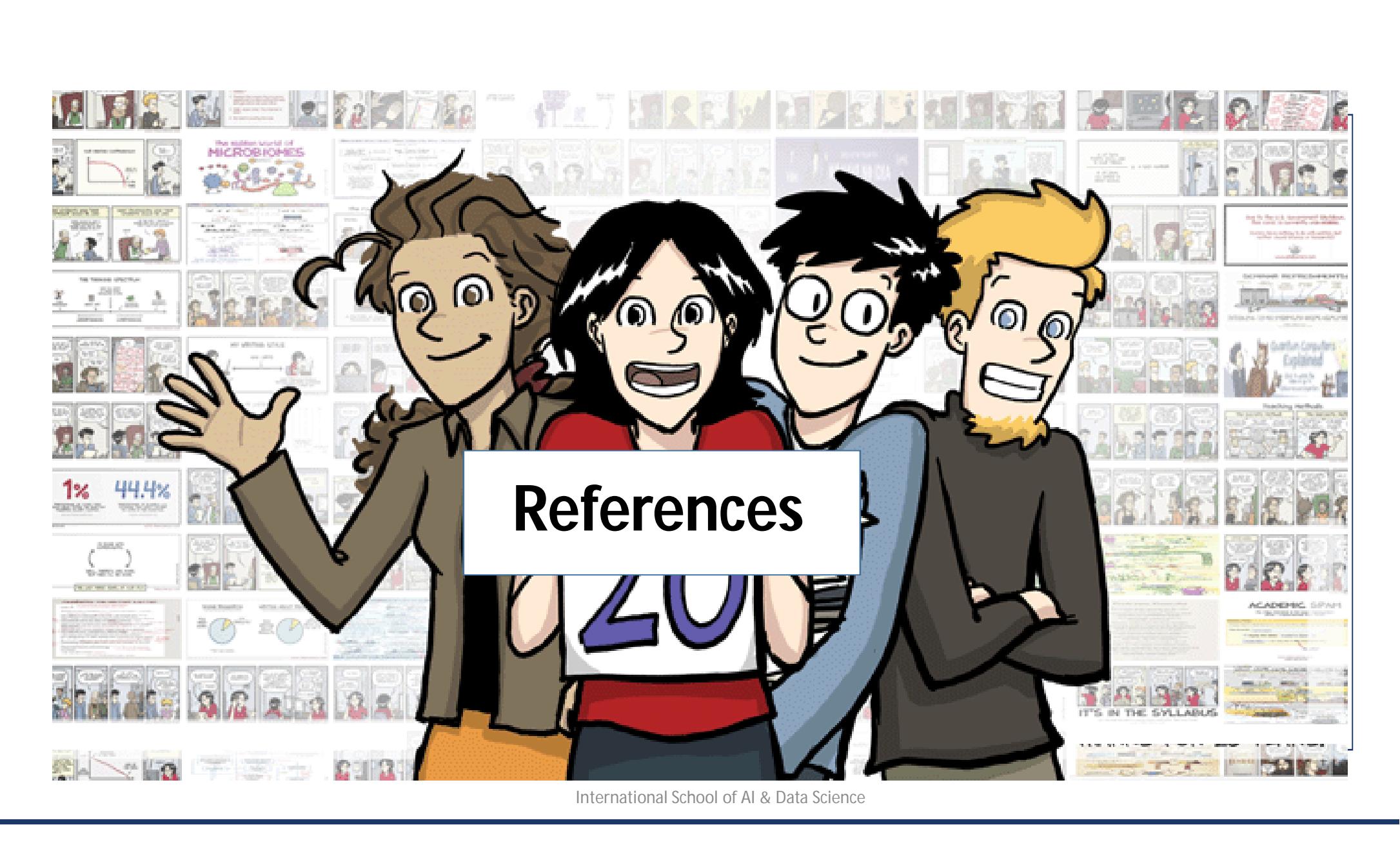
AGENDA

- What is Clustering?
- Unsupervised Learning
- Why Clustering?
- Types of Clustering
 - Partitioning Clustering
- K Means Clustering
- Challenges in K Means Clustering
- Elbow Method
- Euclidean Distance
- Illustration of K Means algorithm
- Applications of K Means

References



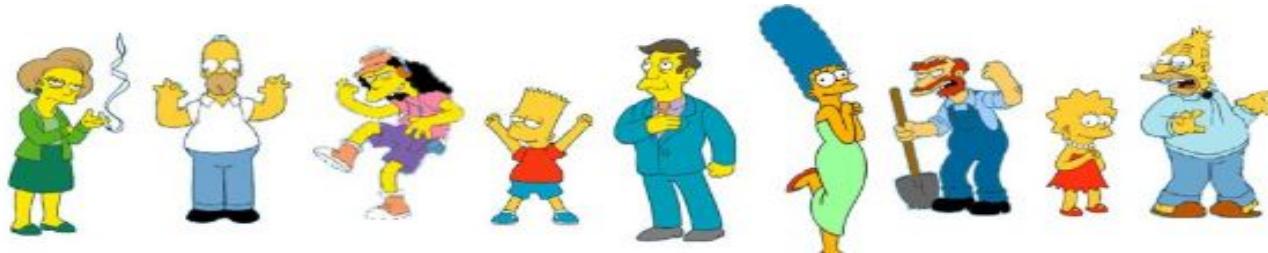
- **Hierarchical Clustering**
 - Agglomerative Clustering
 - Divisive Clustering
- Applications
- Density Based Clustering
- Distance metrics
 - Manhattan
 - Minkowski
 - Mahalanobis



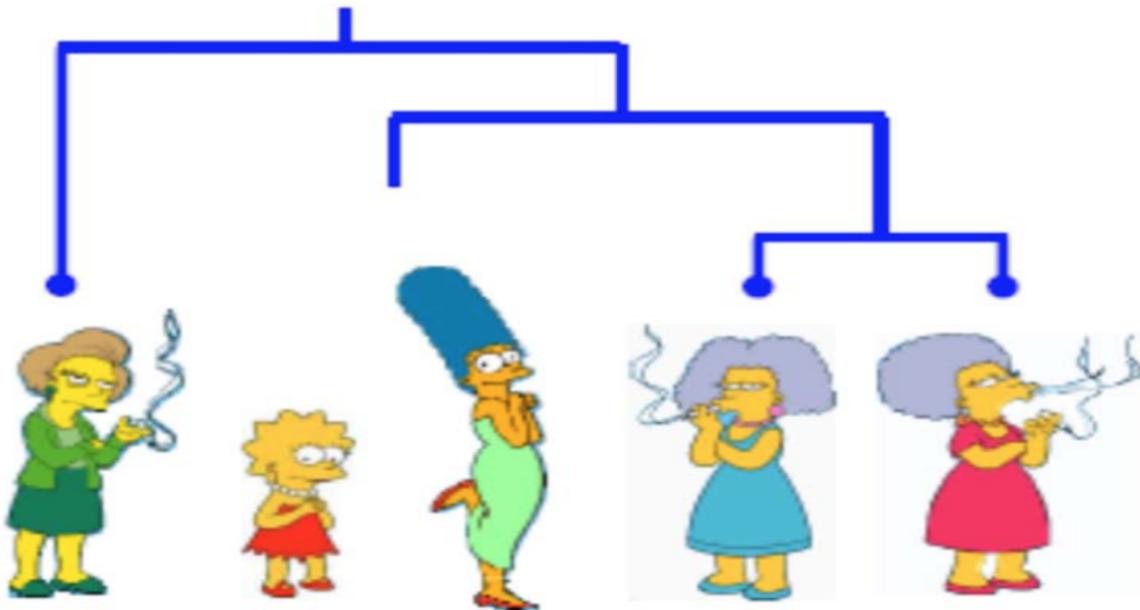
References

Hierarchical Clustering

- Produces a set of **nested clusters** organised as a hierarchical tree
- It can be visualised as **a dendrogram**.
 - A **tree like diagram** that records the sequences of **merges or splits**.



Dendrogram



Hierarchical clustering

- Hierarchical clustering is of two types :
- **Agglomerative**(bottom - up approach)
 - Start with the points as individual clusters.
 - At each step, **merge the closest pair of clusters** until only one cluster(or k clusters) left.
- **Divisive**(top - down approach)
 - Start with one , all inclusive cluster.
 - At each step, **split a cluster until each cluster** contains a point.

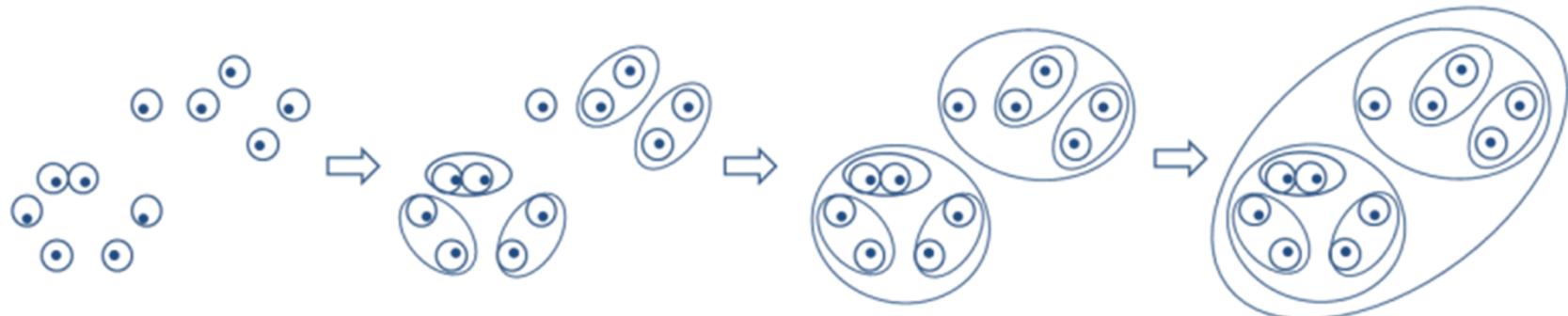
AGENDA

- What is Clustering?
- Unsupervised Learning
- Why Clustering?
- Types of Clustering
 - Partitioning Clustering
- K Means Clustering
- Challenges in K Means Clustering
- Elbow Method
- Euclidean Distance
- Illustration of K Means algorithm
- Applications of K Means

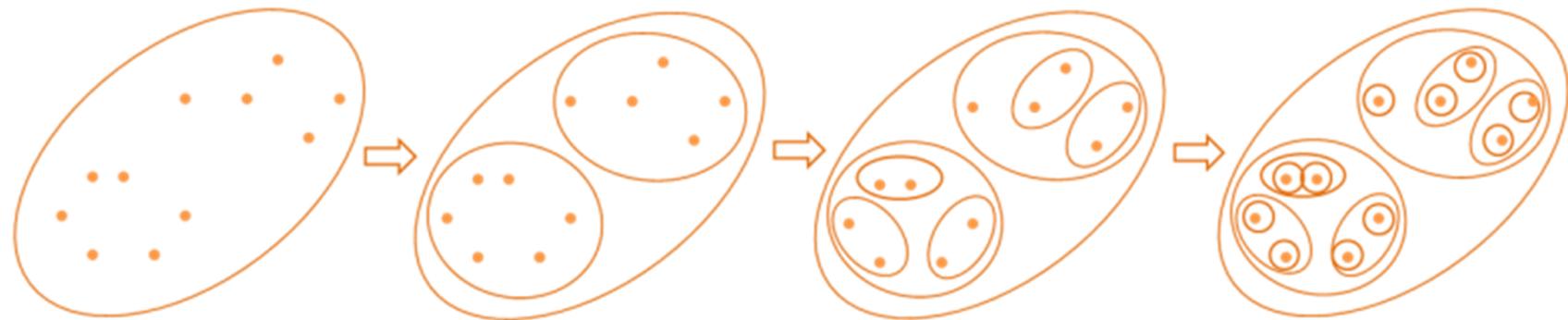
References

- **Hierarchical Clustering**
 - Agglomerative Clustering
 - Divisive Clustering
- Applications
- Density Based Clustering
- Distance metrics
 - Manhattan
 - Minkowski
 - Mahalanobis

Agglomerative Hierarchical Clustering

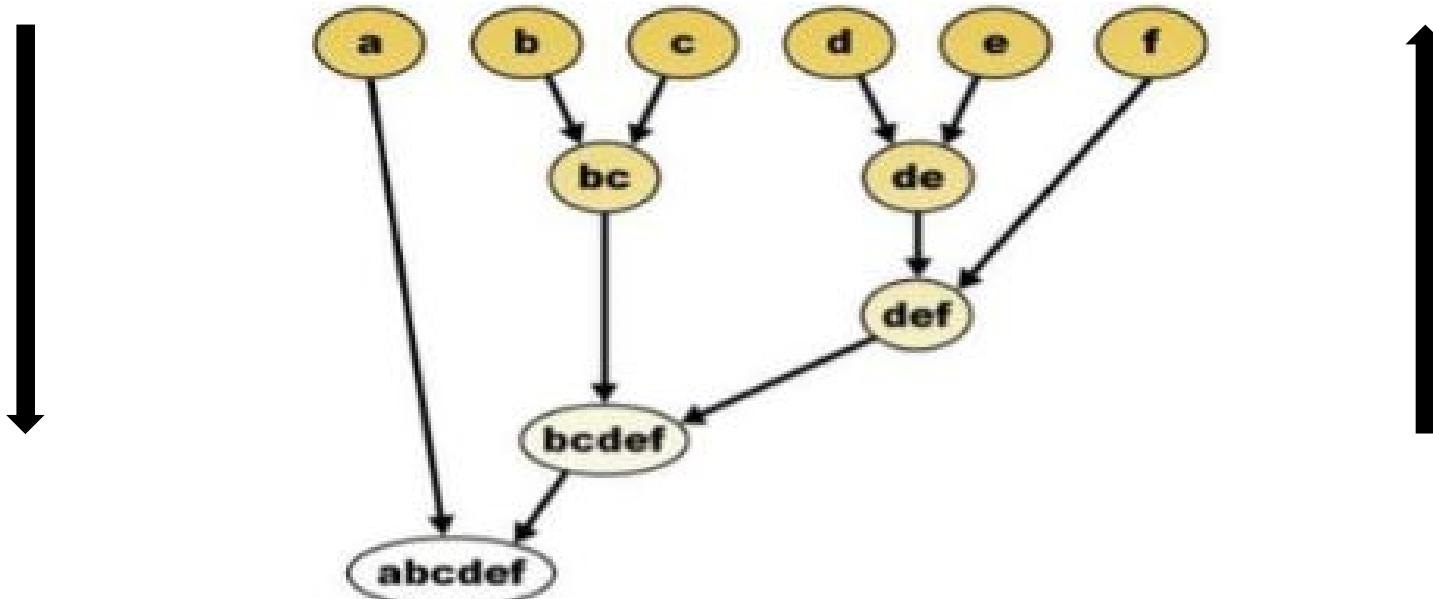


Divisive Hierarchical Clustering



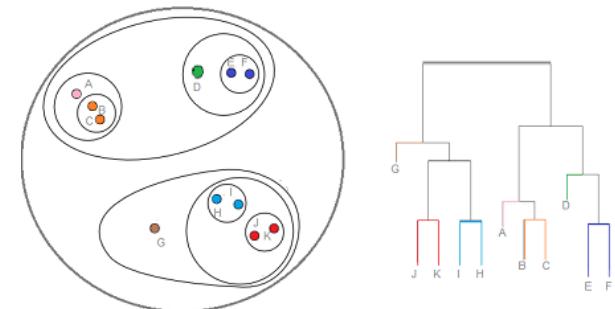
Agglomerative

Divisive



When to use Hierarchical clustering?

- Hierarchical clustering **outputs a hierarchy**, i.e a structure that is **more informative** than the unstructured set of flat clusters returned by k-means.
- Therefore, it is **easier to decide** on the number of clusters by **looking at the dendrogram**.
- This type of clustering allows us to choose **any distance metric**.
 - K means restricted us to Euclidean distance.



AGENDA

- What is Clustering?
- Unsupervised Learning
- Why Clustering?
- Types of Clustering
 - Partitioning Clustering
- K Means Clustering
- Challenges in K Means Clustering
- Elbow Method
- Euclidean Distance
- Illustration of K Means algorithm
- Applications of K Means

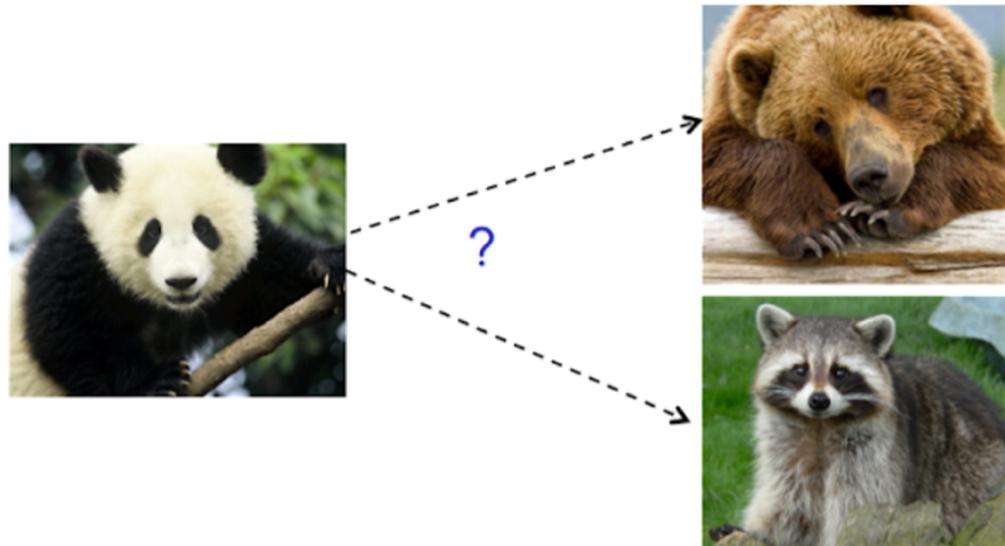
References

- Hierarchical Clustering
 - Agglomerative Clustering
 - Divisive Clustering
- Applications
 - Density Based Clustering
 - Distance metrics
 - Manhattan
 - Minkowski
 - Mahalanobis

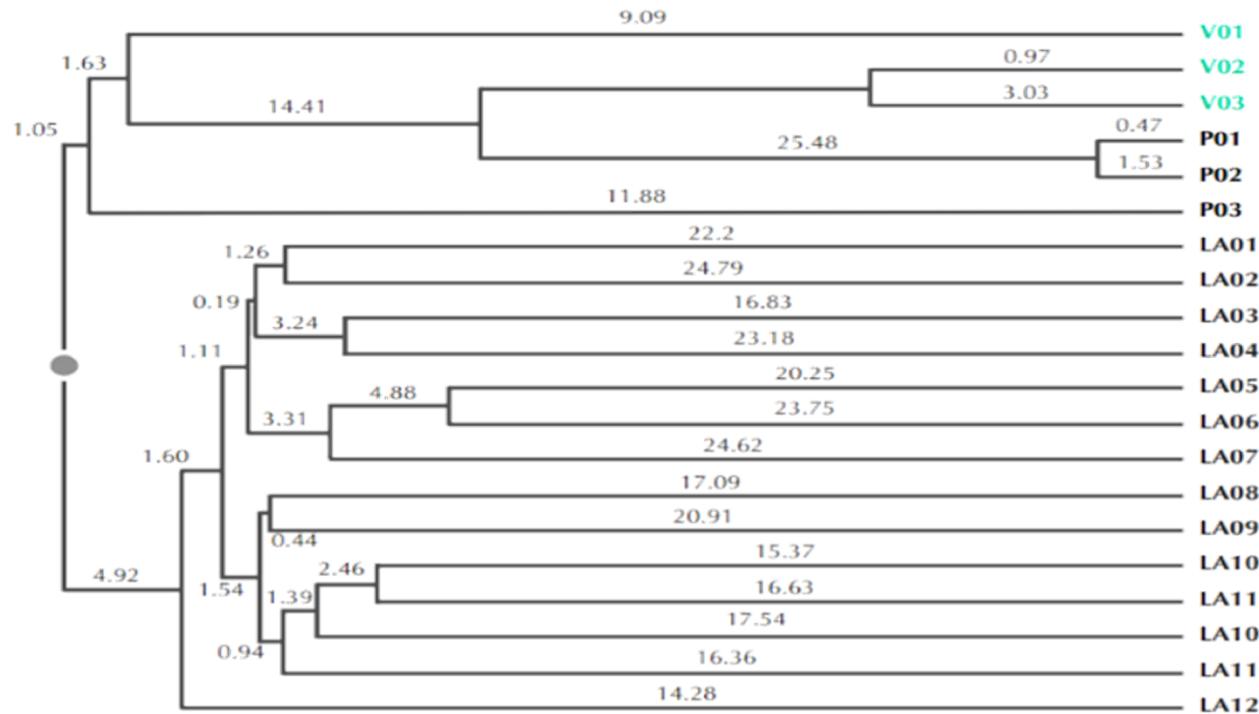


Applications

- Charting Evolution through Phylogenetic trees.



- Tracking Viruses through Phylogenetic trees



- A similar study was also done for finding the animal that gave the humans the SARS virus

