Unsupervised Learning

* Independent Variables, Target Variable 🡪 Supervised Learning
* Independent Variables 🡪 Unsupervised Learning
* Not used for Target Variable Predictions
* Patterns in the data is found
* Multiple operations are done including: Clustering, Segmenting, Hashing, Recommendation…

# Clustering

* Data which is similar to other data must be clubbed in one group or cluster
* Most famous Clustering Approaches:
  + K-Means Clustering
  + Hierarchical Clustering

## K-Means Clustering

* Randomly find n cluster points
* All points which are closer to each cluster points are grouped together
* Find the centroid of each cluster
* Repeat the process till there is no migration of any of the data points into any of the clusters
* Sum of all data points (X-Axis) divided by total number of points and Sum of all data points (Y-axis) divided by total number of points will give us the coordinates of Centroid. {(x1+x2+x3+x4+x5)/5}, {(y1+y2+y3+y4+y5)/5}Diagram

  Description automatically generated

# Hierarchical Clustering

* Agglomerative HC
* Divisive HC – Opposite of Agglomerative

**Agglomerative HC**

Diagram

Description automatically generated

**Dendrogram**

A picture containing line chart

Description automatically generated