**Customer Segmentation**

Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to effective and appropriate marketing such as gender, age, interests, and miscellaneous spending habits.

In business-to-business marketing, a company might segment customers according to a wide range of factors, including:

* Industry
* Number of employees
* Products previously purchased from the company
* Location

In business-to-consumer marketing, companies often segment customers according to demographics that include:

* Age
* Gender
* Marital status
* Location (urban, suburban, rural)
* Life stage (single, married, divorced, empty-nester, retired, etc.)

Customer Segmentation is one pf the most important applications in unsupervised learnings.

The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioural patterns play a crucial role in determining the company direction towards addressing the various segments.

**Clustering**

The main goal behind cluster partitioning methods like k-means is to define the clusters such that the intra-cluster variation stays minimum.

**minimize(sum W(Ck)), k=1…k**

Where Ck represents the kth cluster and W(Ck) denotes the intra-cluster variation. With the measurement of the total intra-cluster variation, one can evaluate the compactness of the clustering boundary.

ELBOW METHOD  
define clusters such that the total intra-cluster variation [or total within-cluster sum of square (WSS)] is minimized. The total WSS measures the compactness of the clustering and we want it to be as small as possible.

SILHOUETTE METHOD

it measures the quality of a clustering. That is, it determines how well each object lies within its cluster. A high average silhouette width indicates a good clustering.

GAP STATISTIC

The gap statistic compares the total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data. The estimate of the optimal clusters will be value that maximize the gap statistic (i.e, that yields the largest gap statistic). This means that the clustering structure is far away from the random uniform distribution of points.

fviz\_nbclust() function [in factoextra R package]: It can be used to compute the three different methods [elbow, silhouette and gap statistic] for any partitioning clustering methods [K-means, K-medoids (PAM), CLARA, HCUT].