What is 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 marketing such as gender, age, interests, and miscellaneous spending habits.*

Companies that deploy customer segmentation are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. Companies aim to gain a deeper approach of the customer they are targeting. Therefore, their aim has to be specific and should be tailored to address the requirements of each and every individual customer. Furthermore, through the data collected, companies can gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit. This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

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 behavioral patterns play a crucial role in determining the company direction towards addressing the various segments.

BASIS VALI SLIDE MA AA:-

Conducting best current customer segmentation research can have numerous other ancillary benefits, of course, but this guide will focus primarily on how it can impact the four cited above.

The most common ways in which businesses segment their customer base are:

1. **Demographic information**, such as gender, age, familial and marital status, income, education, and occupation.
2. **Geographical information**, which differs depending on the scope of the company. For localized businesses, this info might pertain to specific towns or counties. For larger companies, it might mean a customer’s city, state, or even country of residence.
3. **Psychographics**, such as social class, lifestyle, and personality traits.
4. **Behavioral data**, such as spending and consumption habits, product/service usage, and desired benefits.

There are three main approaches to market segmentation:

* *A priori* segmentation, the simplest approach, uses a classification scheme based on publicly available characteristics — such as industry and company size — to create distinct groups of customers within a market. However, a priori market segmentation may not always be valid, since companies in the same industry and of the same size may have very different needs.
* *Needs-based* segmentation is based on differentiated, validated drivers (needs) that customers express for a specific product or service being offered. The needs are discovered and verified through primary market research, and segments are demarcated based on those different needs rather than characteristics such as industry or company size.
* *Value-based*segmentation differentiates customers by their economic value, grouping customers with the same value level into individual segments that can be distinctly targeted.

## **What Are Companies That Use Segmentation?**

Numerous types of businesses use [market segmentation](https://www.investopedia.com/terms/m/marketsegmentation.asp) to optimize their ability to sell to a wide variety of consumers, including:

* Skincare, haircare, and beauty product manufacturers
* Car companies
* Clothing and apparel suppliers
* Banks and other [financial institutions](https://www.investopedia.com/terms/f/financialinstitution.asp)
* Television networks and media outlets

Businesses often [segment](https://www.investopedia.com/terms/s/segment.asp) the market based on key demographics such as age, gender, income level, or marital status. But, they also use more precisely defined categories to target specific groups. For example, the Whole Foods chain of high-end grocery stores (now owned by Amazon) targets high-income, educated city-dwellers who are health- or eco-conscious.

In terms of marketing strategy, segmentation within an audience or customer file can also be done by identifying the platforms on which specific audiences live. Targeted marketing companies such as Hubspot allow companies to segment their audiences, and send different versions of emails to specific subsets of customers based on their age or preferences.

## **Marketing Products to Different Groups of Customers**

One group of businesses that uses market segmentation to great effect are manufacturers of hair care, beauty, and other grooming products. For example, the razors marketed to men and women are fundamentally the same, but they have very different packaging and advertising messages. This is a perfect example of market segmentation. Based on research, companies devise different ways to sell products to men and women by using gender normative stereotypes around masculinity and feminity.

Walk into any drugstore, and you quickly notice that women's skincare, haircare, and grooming products are packaged in soft, gentle colors. Most often, the packaging is pink. The messaging used often refers to freshness, softness, or a carefree lifestyle. The women featured on the packaging are generally laughing or playfully smiling, embodying the effortless beauty to which many women aspire.

Conversely, the packaging for men's products is predominated by blacks, grays, reds, and oranges. The messaging focuses on strength, durability, and ruggedness. If a photograph is included, the subject is often a close-up of a granite-featured model with stubble, looking fiercely independent and brooding.

You may also notice the price difference between male and female products. [Market segmentation theory](https://www.investopedia.com/terms/m/market-segmentation-theory.asp) is not just about how to sell products, but also about how valuable those products are to different groups. In general, women are willing to spend more on self-care products due to a higher premium put on feminine beauty in today's society. Knowing this, companies tack on an extra dollar or two to products aimed at women.

Companies often want to be very careful not to be seen as charging higher prices to one group over another based on race, religion, ethnicity, or gender. Since anyone can easily purchase the cheaper, male-targeted product instead of the more-expensive pink product, women pay a premium for choosing pink over black – not for being female. This subtle distinction allows companies to avoid accusations of [price discrimination](https://www.investopedia.com/ask/answers/042415/what-are-different-types-price-discrimination-and-how-are-they-used.asp) without endangering the [bottom line](https://www.investopedia.com/terms/b/bottomline.asp).

Using that example, the segmentation variables can be defined as the objective measures, factors, or characteristics that help you differentiate segments, whether they are needs- or value-based. In the above scenario, those variables focus on financial information, but they could just as well pertain to the customer’s reputation, online presence, or business model, depending on what is most relevant to the segment.

Developing variables and hypotheses is important for a variety of reasons, but its primary purpose is to provide a framework for the customer segmentation research process. Once you have established a clear hypothesis and the variables that you need to test, you can begin executing the intricate process that will help you identify your best current customer segments

WHY IT IS BENEFITICIAL? BENEFITS:-

Without a deep understanding of how a company’s best current customers are segmented, a business often lacks the market focus needed to allocate and spend its precious human and capital resources efficiently. Furthermore, a lack of best current customer segment focus can cause diffused go-to-market and product development strategies that hamper a company’s ability to fully engage with its target segments. Together, all of those factors can ultimately impede a company’s growth.

If best current customer segmentation is done right, however, the business benefits are numerous

1. Improving your whole product: Having a clear idea of who wants to buy your product and what they need it for will help you differentiate your company as the best solution for their individual needs. The result will be increased satisfaction and better performance against competitors. The benefits also extend beyond your core product offering, since any insights into your best customers will allow your organization to offer better customer support, professional services, and any other offerings that make up their whole product experience.
2. Focusing your marketing message: In parallel with improvements to the product, conducting a customer segmentation project can help you develop more focused marketing messages that are customized to each of your best segments, resulting in higher quality inbound interest in your product.
3. Allowing your sales organization to pursue higher percentage opportunities: By spending less time on less lucrative opportunities and more on your most successful segments, your sales team will be able to increase its win rate, cover more ground, and ultimately increase revenues.
4. Getting higher quality revenues:  Not all revenue dollars are created equal. Sales into the wrong segment can be more expensive to sell and maintain, and may have a higher churn rate or lower upsell potential after the initial purchase has been made. Staying away from these types of customers and focusing on better ones will increase your margins and promote the stability of your customer base.

CLUSTERING:-

## What is Clustering?

Clustering is a type of unsupervised learning method of machine learning. In the unsupervised learning method, the inferences are drawn from the data sets which do not contain labelled output variable. It is an exploratory data analysis technique that allows us to analyze the multivariate data sets.

Clustering is a task of dividing the data sets into a certain number of clusters in such a manner that the data points belonging to a cluster have similar characteristics. Clusters are nothing but the grouping of data points such that the distance between the data points within the clusters is minimal.

In other words, the clusters are regions where the density of similar data points is high. It is generally used for the analysis of the data set, to find insightful data among huge data sets and draw inferences from it. Generally, the clusters are seen in a spherical shape, but it is not necessary as the clusters can be of any shape.

It depends on the type of algorithm we use which decides how the clusters will be created. The inferences that need to be drawn from the data sets also depend upon the user as there is no criterion for good clustering.

## What are the types of Clustering Methods?

Clustering itself can be categorized into two types viz. Hard Clustering and Soft Clustering. In hard clustering, one data point can belong to one cluster only. But in soft clustering, the output provided is a probability likelihood of a data point belonging to each of the pre-defined numbers of clusters.

### Density-Based Clustering

In this method, the clusters are created based upon the density of the data points which are represented in the data space. The regions that become dense due to the huge number of data points residing in that region are considered as clusters.

The data points in the sparse region (the region where the data points are very less) are considered as noise or outliers. The clusters created in these methods can be of arbitrary shape. Following are the examples of Density-based clustering algorithms:

### DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN groups data points together based on the distance metric and criterion for a minimum number of data points. It takes two parameters – *eps* and *minimum points.*Eps indicates how close the data points should be to be considered as neighbors. The criterion for minimum points should be completed to consider that region as a dense region.

### OPTICS (Ordering Points to Identify Clustering Structure)

It is similar in process to DBSCAN, but it attends to one of the drawbacks of the former algorithm i.e. inability to form clusters from data of arbitrary density. It considers two more parameters which are core distance and reachability distance. Core distance indicates whether the data point being considered is core or not by setting a minimum value for it.

Reachability distance is the maximum of core distance and the value of distance metric that is used for calculating the distance among two data points. One thing to consider about reachability distance is that its value remains not defined if one of the data points is a core point.

### HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)

HDBSCAN is a density-based clustering method that extends the DBSCAN methodology by converting it to a hierarchical clustering algorithm.

### Hierarchical Clustering

Hierarchical Clustering groups (Agglomerative or also called as Bottom-Up Approach) or divides (Divisive or also called as Top-Down Approach) the clusters based on the distance metrics. In Agglomerative clustering, each data point acts as a cluster initially, and then it groups the clusters one by one.

Divisive is the opposite of Agglomerative, it starts off with all the points into one cluster and divides them to create more clusters. These algorithms create a distance matrix of all the existing clusters and perform the linkage between the clusters depending on the criteria of the linkage. The clustering of the data points is represented by using a dendrogram. There are different types of linkages: –

o   **Single Linkage**: – In single linkage the distance between the two clusters is the shortest distance between points in those two clusters.

o   **Complete Linkage**: – In complete linkage, the distance between the two clusters is the farthest distance between points in those two clusters.

o   **Average Linkage**: – In average linkage the distance between the two clusters is the average distance of every point in the cluster with every point in another cluster.

Read: [Common Examples of Data Mining.](https://www.upgrad.com/blog/most-common-examples-of-data-mining/)

### Fuzzy Clustering

In fuzzy clustering, the assignment of the data points in any of the clusters is not decisive. Here, one data point can belong to more than one cluster. It provides the outcome as the probability of the data point belonging to each of the clusters. One of the algorithms used in fuzzy clustering is Fuzzy c-means clustering.

This algorithm is similar in process to the K-Means clustering and it differs in the parameters that are involved in the computation like fuzzifier and membership values.

### Partitioning Clustering

This method is one of the most popular choices for analysts to create clusters. In partitioning clustering, the clusters are partitioned based upon the characteristics of the data points. We need to specify the number of clusters to be created for this clustering method. These clustering algorithms follow an iterative process to reassign the data points between clusters based upon the distance. The algorithms that fall into this category are as follows: –

o   **K-Means Clustering: –**K-Means clustering is one of the most widely used algorithms. It partitions the data points into k clusters based upon the distance metric used for the clustering. The value of ‘k’ is to be defined by the user. The distance is calculated between the data points and the centroids of the clusters.

The data point which is closest to the centroid of the cluster gets assigned to that cluster. After an iteration, it computes the centroids of those clusters again and the process continues until a pre-defined number of iterations are completed or when the centroids of the clusters do not change after an iteration.

It is a very computationally expensive algorithm as it computes the distance of every data point with the centroids of all the clusters at each iteration. This makes it difficult for implementing the same for huge data sets.

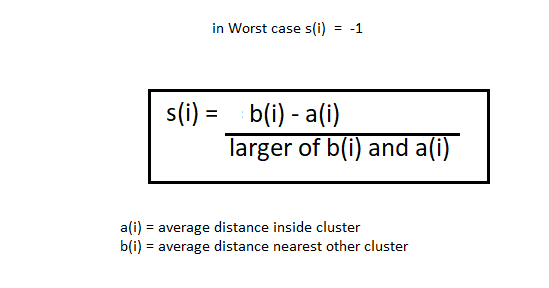
ELBOW METHOD: -

**The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters. By default, the distortion score is computed, the sum of square distances from each point to its assigned center.**

**SILLHOUTE METHOD: -**

**The silhouette method is somewhat different. The elbow method it also picks up the range of the k values and draws the silhouette graph. It calculates the silhouette coefficient of every point. It calculates the average distance of points within its cluster a (i) and the average distance of the points to its next closest cluster called b (i).**

Now, we have the values of a (i) and b (i). we will calculate the silhouette coefficient by using the below formula.



**Abstract: - The Gap statistic is a standard method for determining the number of clusters in a set of data. The Gap statistic standardizes the graph of log(Wk), where Wk is the within-cluster dispersion, by comparing it to its expectation under an appropriate null reference distribution of the data.**

* **It is an empirical method to find out the best value of k. it picks up the range of values and takes the best among them. It calculates the sum of the square of the points and calculates the average distance.**
* **When the value of k is 1, the within-cluster sum of the square will be high. As the value of k increases, the within-cluster sum of square value will decrease.**

**ADVANTAGES: -**

**Benefits of Customer Segmentation**

Number of advantages to it:

* Helps identify least and most profitable customers, thus helping the business to concentrate marketing activities on those most likely to buy your products or services
* Helps build loyal relationships with customers by developing and offering them the products and services they want
* Helps improve customer service
* Helps maximize use of your resources
* Helps improve or tweak products to meet customer requirements
* Helps increase profit by keeping costs down