

# **Hierarchical Clustering**

## **Market Segmentation, Targeting and Positioning**

**RIDA AZMI**

### **Introduction & Business Problem :**

Businesses may not be in a position to satisfy all of their customers, every time. It may prove difficult to meet the exact requirements of each individual customer. People do not have identical preferences, so rarely does one product completely satisfy everyone. Many companies may usually adopt a strategy that is known as target marketing. This strategy involves dividing the market into segments and developing products or services to these segments. A target marketing strategy is focused on the customers' needs and wants. Hence, a prerequisite for the development of this customer-centric strategy is the specification of the target markets that the companies will attempt to serve. The marketing managers who may consider using target marketing will usually break the market down into groups and to target the most profitable segments. They may adapt their marketing mix elements, including; products, prices, channels, and promotional tactics to suit the requirements of individual consumers. In sum, this chapter explains the three stages of target marketing, including; market segmentation (ii) market targeting and (iii) market positioning.

### **Problem Description**

A market segment is a group of individuals, groups or organisations who may share the same interests, traits and characteristics. The consumer segments may have similar needs, wants and expectations. Therefore, businesses should ask themselves which segments should they serve? To answer this question, the businesses must determine the most appropriate ways to distinguish and to differentiate their segments. Once the segments have been identified they must customise their offerings to satisfy each and every one of them.

Market segmentation is the actual process of identifying segments of the market and the process of dividing a broad customer base into sub-groups of consumers consisting of existing and prospective customers. Market segmentation is a consumer-oriented process and can be applied to almost any type of market. In dividing or segmenting markets, researchers typically look for shared characteristics such as common needs, common interests, similar lifestyles or even similar demographic profiles. So, market segmentation assumes that different segments require different marketing programmes, as diverse customers are usually targeted through different offers, prices, promotions, distributions or some combination of marketing variables. For example, Southwest Airlines' single-minded focus on the short-haul, point-to-point, major-city routes, allowed them to prosper as their competitors floundered. The airline's focus on specific segments allowed them to do a better job of deciding what their target segment really valued (for example, convenience, low price, on-time departures and arrivals, among other things).

Once the customer segments have been identified and profiled, the marketer must decide which segment to target. Diverse customers will have different expectations. For instance, there may

be customers who will value a differentiated, high quality service, whilst others may be more price-sensitive. Notwithstanding, not all firms have the resources to serve all customers in an adequate manner. Trying to serve the entire market could be a recipe for disaster. The overall aim of segmentation is to identify high-yield segments. These are likely to be the most profitable groups of customers, or may hold potential for growth. Hence, the most lucrative segments will usually become target markets

## Target Audience:

Once the market segmentation has been completed, the company should be aware of the needs and wants of its selected segments. It is in the interest of the business to identify any untapped needs in the marketplace, as there could be customers who may not be adequately served by competitors. It is then necessary to identify the most profitable segments and to decide which segments will be served. There are three market coverage alternatives which can be applied; undifferentiated marketing; differentiated marketing and concentrated marketing.

## Success Criteria:

The clustering technique can be very handy when it comes to unlabeled data. Since most of the data in the real-world is unlabeled and annotating the data has higher costs, clustering techniques can be used to label unlabeled data. In this case, we will use the clustering technique in order to segment customers targets into 4 groupes

## Description of input Data

The dataset for this problem can be downloaded from the following link:

[https://drive.google.com/open?id=18Dsja5\\_1jRY1GnWoORXFFKGTJhylk6rJ](https://drive.google.com/open?id=18Dsja5_1jRY1GnWoORXFFKGTJhylk6rJ)

Our dataset has five columns: CustomerID, Genre, Age, Annual Income, and Spending Score. this is suffiscent to simulate targeting in marketing strategies and with that type of a=data we can easily segment our customer

The data look like this

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Table 1 : sample of ra data

## Data cleaning

Our dataset has five columns: CustomerID, Genre, Age, Annual Income, and Spending Score. To view the results in two-dimensional feature space, we will retain only two of these five columns. We can remove CustomerID column, Genre, and Age column. We will retain the Annual Income (in thousands of dollars) and Spending Score (1-100) columns. The Spending Score column signifies how often a person spends money in a mall on a scale of 1 to 100 with 100 being the highest spender. Execute the following script to filter the first three columns from our dataset:

## Exploratory Data Analysis

Our dataset has five columns: CustomerID, Genre, Age, Annual Income, and Spending Score. To view the results in two-dimensional feature space, we will retain only two of these five columns. We can remove CustomerID column, Genre, and Age column. We will retain the Annual Income (in thousands of dollars) and Spending Score (1-100) columns. The Spending Score column signifies how often a person spends money in a mall on a scale of 1 to 100 with 100 being the highest spender. Execute the following script to filter the first three columns from our dataset:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Table 2 : sample of dataset

## Clustering analysis

we need to know the clusters that we want our data to be split to. We will again use the scipy library to create the dendrograms for our dataset.

Basic definition :

Hierarchical clustering is a type of unsupervised machine learning algorithm used to cluster unlabeled data points. Like K-means clustering, hierarchical clustering also groups together the data points with similar characteristics. In some cases the result of hierarchical

and K-Means clustering can be similar. Before implementing hierarchical clustering using Scikit-Learn, let's first understand the theory behind hierarchical clustering.

## Theory of Hierarchical Clustering

There are two types of hierarchical clustering: Agglomerative and Divisive. In the former, data points are clustered using a bottom-up approach starting with individual data points, while in the latter top-down approach is followed where all the data points are treated as one big cluster and the clustering process involves dividing the one big cluster into several small clusters.

In this article we will focus on agglomerative clustering that involves the bottom-up approach.

## Steps to Perform Hierarchical Clustering

Following are the steps involved in agglomerative clustering:

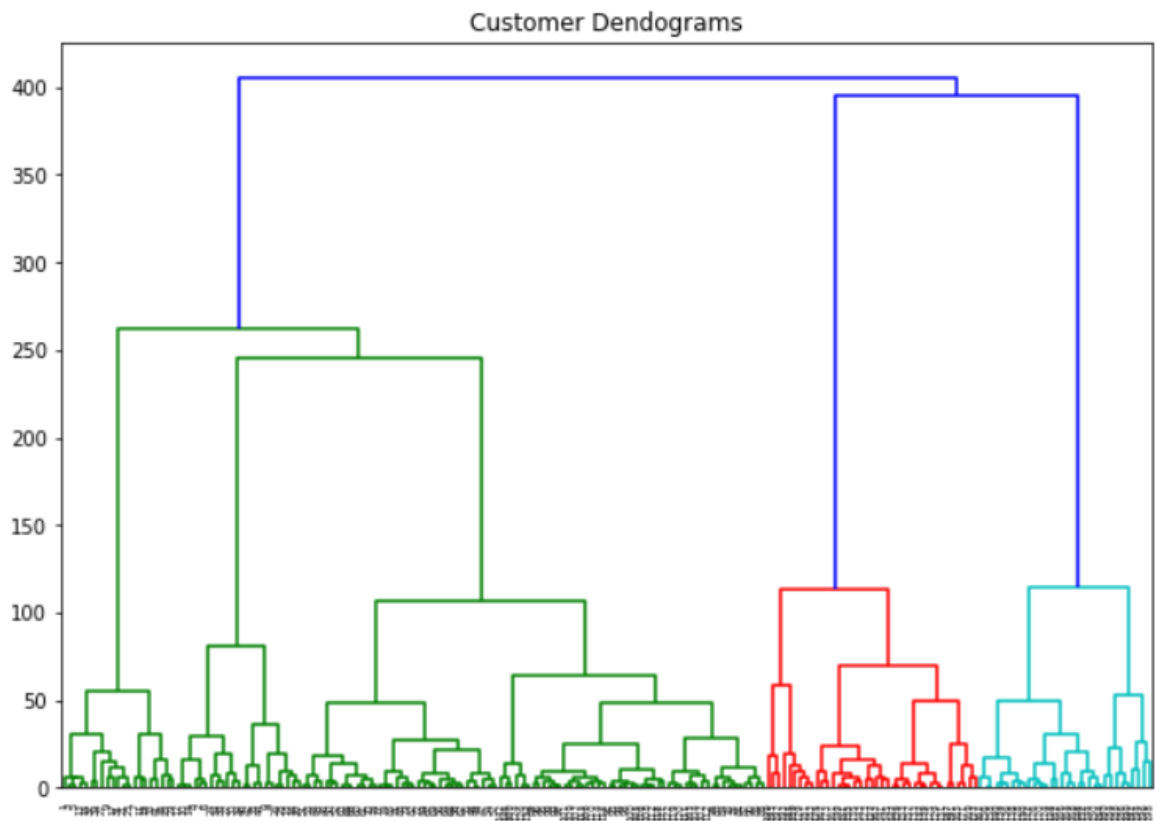
1. At the start, treat each data point as one cluster. Therefore, the number of clusters at the start will be  $K$ , while  $K$  is an integer representing the number of data points.
2. Form a cluster by joining the two closest data points resulting in  $K-1$  clusters.
3. Form more clusters by joining the two closest clusters resulting in  $K-2$  clusters.
4. Repeat the above three steps until one big cluster is formed.
5. Once single cluster is formed, dendrograms are used to divide into multiple clusters depending upon the problem. We will study the concept of dendrogram in detail in an upcoming section.

There are different ways to find distance between the clusters. The distance itself can be Euclidean or Manhattan distance. Following are some of the options to measure distance between two clusters:

1. Measure the distance between the closest points of two clusters.
2. Measure the distance between the farthest points of two clusters.
3. Measure the distance between the centroids of two clusters.
4. Measure the distance between all possible combination of points between the two clusters and take the mean.

## Role of Dendrograms for Hierarchical Clustering

In the last section, we said that once one large cluster is formed by the combination of small clusters, dendrograms of the cluster are used to actually split the cluster into multiple clusters of related data points. Figure 1 one shows result of clustering using input data :



**Figure 1 :** plotted clustering result using `scipy.cluster.hierarchy`

## Conclusion

The clustering technique can be very handy when it comes to unlabeled data. Since most of the data in the real-world is unlabeled and annotating the data has higher costs, clustering techniques can be used to label unlabeled data.