**A Study In Market Segmentation**

What do organizations need the IT departments for? Why do big corporates conduct surveys ? What the process of implementing certain policies that often don’t make sense when viewed vicariously ?

The answers to all these questions can be obtained by understanding the process of Market Segmentation. There are 10 crucial steps to resolve a segmentation problem, out of which I am here to explain the part that deals with identifying the segment, which arguably is the most crucial step in market segmentation.

**Bsc Data Science**

**Pratyush Rao, Intern at Feynn Labs**

**The basics**

To understand the concepts of how market segmentation is performed it is only logical to answer the primary questions. To truly understand any new concept it is crucial to lay a foundations first.

**What is Market Segmentation ?**

Disintegrating the market from the standpoint of an organization and implementing distinct measures in each of the disintegrated segments with similar behavior, to maximize the organizations objective is known as Market Segmentation.

The individual segments are identified such that, individual elements in a segment behave similar to their neighbors inside the same segment, and distinctively from members of the other segments. However it is important to be careful before we blindly implement segmentation techniques to resolve any business problem. Here is a comprehensive introduction as to when and how we is it advisable to implement market segmentation.

Step 1: Knowing When Not To Implement Market Segmentation.

Identifying patterns in your data is crucial because this gives the organization an insight into what kind of patterns can be capitalized upon to better attain the company’s objective, however, it is not necessary that implementing Market Segmentation in your organization will always be profitable. To identify useful patterns you require data and to ensure those patterns are not simply a co-incidence and a one time occurrence you need large volumes of it. It is essential to weigh your options before you implement segmentation techniques.

Market segmentation is not a data but a committed marriage, It requires investment of time money and labor. It should only applied if the organization would be in a loss implementing it. If there is any alternative which is simpler and less expensive, the latter should always be preferred. Market Segmentation requires huge volumes of data, over a long period of time to understand seasonal or cyclic patterns in your data. For example applying segmentation for data that is collected during the Diwali period would give incorrect profit estimates because Customers tend to spend more magnanimously during Festivals.

It also requires and in-depth market research and analysis and co-ordination with people from different organizational units and involvement of the senior management of the organization. Unwillingness to accept absurd ideas that the data recommends, by the senior management, is irresponsible and could incur losses as Market segmentation is expensive.

A structure in the organization along with the willingness to keep and open mind and actively participate with the analysis teams is essential to yield successful results from Segmentation.

Above all, a resolute sense of purpose and dedication is required, tempered by patience and a willingness to appreciate the inevitable problems which will be encountered in implementing the conclusions.

Step 2: Identifying the correct Segments.

The data often received in real world scenarios to implementation segmentation upon comes with several variables. For example, a car sales company would get different types customers who will demand a certain type of car will have a certain budget for expenditure, will have a certain age while buying the car, will be of a certain gender and will have certain preferences and biases towards buying a certain type of car. As a salesman you would like to know what age groups and genders of car owners are more prone to spend more at your organization so you can define a target audience for yourself. So in this case you are segmenting on the basis of Age and Gender. This is known as Socio-demographic Segmentation, a topic which we will discuss moving ahead later.

The purpose of the previous example was to demonstrate how crucial deciding a segmentation variable is. If we for example selected an individual’s name for segmentation the profits would not show any change (might even reduce profits because it is likely we will implement policies like “don’t sell cars to people who’s names start with R” )

There are two cases an analyst will encounter while selecting segmenting variables, one more often than the other.

The first one is Common Sense Implementation, in which there is only one segmentation variable and hence requires no further analysis to segment. You simply elect the individual segmentation variable and then understand the traits of the classes inside the segment and implement policies to maximize profit.

However it is highly unlikely to encounter a real world problem with the common sense implementation of market segmentation. The other type is data-driven segmentation. This type of segmentation consists of compiling a collection of segmentation variables and then creating segments out of it and studying the descriptor variables of each of those provided segments. This helps in better understanding the behavioral traits of your clustered segments. In this approach you are likely to encounter several useless variables that will create unnecessary noise in your segments, so the first step is selecting a set of segmentation variables by consulting different organizational units. The conventionally accepted number of segmenting variables is 6.

There are two stages that of electing segmenting variables known as the Knock out stage and the attractiveness stage.

In the knock out stage, organizational units collectively decide upon the variables that are necessary by eliminating the ones that are certainly noisy or are more likely to be considered as a descriptor variable.

The variables that are elected in the knock out stage then go through an attractiveness phase where each of the segmenting variables are assigned a weightage depending on the future growth prospects in them.

**Types of Segmentation**

While deciding upon segmentation variables one will encounter several categories. Deciding which one to keep and which one to eliminate varies depending upon the objective the organization hopes to attain. However a general pattern can be observed to categorize a set of segmentation variables that can help with electing a set of variables knowing the nature of the problem.

1. Geographical Segmentation: By far the simplest and the most effective type is the geographical segmentation. It is simple and it is effective, Organizations obviously need to divide their policies based on geographical variables. For example customers in Kashmir would like sales on Heaters whereas Customers in Mumbai would want sales on Air conditioners. Geographical location plays an important role in monitoring that sales of an organization and it is most of the times effective, but the assumption that people living in the same geographical location exhibit similar geographical traits is dangerous and can cost an organization.
2. Socio-Demographic Segmentation: Segmentation based on Gender, Age, Income and Education comes under this categorization. Segmenting based on these factors do give better insights than simply assuming similarity because of Geographic locations but according to Haley it only explains 5% of the variance in customer behaviors and Yankelovich and Meer claim that it is not a strong segmentation variable.

Both of the above character traits are easy to determine as they are physical and primary character traits, the following however are not.

1. Behavioral Segmentation: Segmentation based upon behavioral patterns is a marketing strategy that involves dividing consumers into distinct groups based on their behavior, actions, and habits. This type of segmentation focuses on how consumers interact with a product or service, rather than demographic or psychographic characteristics. The problem however with this type of segmentation is that the behavior you study and analyze will be coming from the data that exists in your customer database and will not be ready to segment if your database changes.
2. Psychographic Segmentation: Segmentation based upon psychographic characteristics is a marketing strategy that involves dividing consumers into distinct groups based on their attitudes, values, beliefs, interests, aspirations, and preferences. Benefit segmentation proposed by Haley is the most popular form of psychographic market segmentation, followed by lifestyle segmentation based on people's activities, opinions, and interests. However, this approach has challenges, including identifying relevant psychographic characteristics, avoiding stereotyping, and accurately targeting specific consumer groups due to the subjective nature of these characteristics.

Step 3: Collecting data.

Segmentation of the market is crucial for targeted marketing. There are primarily three ways to collect data for segmentation: Surveys, Internal data, and Experimental data. Surveys have the following problems:

1. Biased data: One of the most significant problems with surveys is the potential for biased data. Respondents may unconsciously skew their responses to fit social norms or their own beliefs, leading to inaccurate segmentation.
2. Selection of variables: Another problem with surveys is the selection of variables. Only important variables should be considered, but selecting too many variables can cause respondent fatigue. It is crucial to consider variables that have a significant impact on the target market, such as demographics, psychographics, and behavioral traits.
3. Response options: The accuracy of responses can be compromised if the response options are not clear. It is recommended to use binary, metric or visual analogue scale options to avoid confusion. Visual analogue scales are useful for capturing responses that are not easily quantifiable, such as emotions. Binary and metric options are preferred as they have a clear distance measure, unlike ordinal options.
4. Response styles: Biases like acquiescence bias (saying yes to everything) should be avoided. Other response styles like extreme response style (tendency to choose extreme options) or social desirability bias (tendency to give socially desirable answers) can also affect the accuracy of segmentation. Careful consideration should be given to the response styles to ensure the accuracy of the responses.

However, surveys can provide several benefits for segmentation such as:

1. Segmentation variables: Surveys can help in selecting the most important variables for segmentation.
2. Identifying segments: Surveys can help in identifying new market segments.
3. Understanding behavioral traits: Surveys can help in understanding the behavioral traits of the segments.

Internal data and Experimental data are other sources of data that can be used for segmentation. Internal data is data that the company already has, and it can provide valuable insights into consumer behavior. It can be used to create more accurate and detailed segments based on the company's internal information.

Experimental data is data that is collected through experiments, and it can be used to test hypotheses and validate segmentation variables. This type of data is used mostly in laboratory and research circles like Biotechnology or fields of biology. Top of Form

Step 5: Extracting Segments.

Whether or not distinct non-intersecting clusters are formed or not depends on what type of segmenting variables the organizational units have chosen for the data. If there exists a segmenting pattern in the underlying data, and the segmenting variables are chosen according to classes that cause this pattern, then the segments elected will be distinct. However, if there are noisy parameters that are chosen, then the segments will not be as segregated as you’d expect them to be.

* Clustering is the most common method used for segmenting.
* KMeans clustering won't always work, for example, in spiral datasets. It identifies grouped clusters based on regions.
* A better performance on spiral shapes can be obtained by the Single Linkage Hierarchical clustering algorithm.
* It's important to note that this doesn't mean that the Single Linkage Algorithm is better, only an indication that different algorithms work better on different types of data.

If consumer data is well-structured and well-separated, distinct market segments exist, and tendencies of different algorithms matter less. However, if the data is not well-structured, the tendency of the algorithm influences the solution substantially. In such situations, the algorithm will impose a structure that suits the objective function of the algorithm.

Types of methods:

* Distance-based methods: These methods use the proximity of data points to determine clusters. Examples include:
  + KMeans clustering
  + Hierarchical clustering
  + DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
  + OPTICS (Ordering Points To Identify the Clustering Structure)
* Model-based methods: These methods formulate a concise stochastic model to extract segments. Examples include:
  + Gaussian Mixture Models (GMM)
  + Latent Class Analysis (LCA)
  + Finite Mixture Models (FMM)

Important to note that no single best algorithm exists. It’s all trial, error, and logical inferences drawn from what you know about the dataset that help you choose the most appropriate algorithm that yields the best results to form those segments.

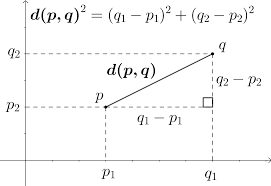
Binary Segmentation: Segmentation involved with using binary segmentation variables.

* Binary segmentation involves using binary segmentation variables to segment data.
* Two methods of binary segmentation include:
  + Symmetric binary segmentation: where ones and zeros have equivalent values.
  + Asymmetric binary segmentation: where 1 is more valuable for clustering, but 0 doesn't indicate anything. For example, a person being interested in horseback riding is more useful information to the tourism industry than a person who does not like horseback riding.Top of Form

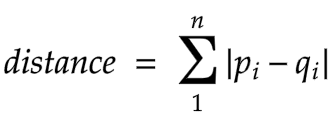
Distance based methods:

Three types of distances:

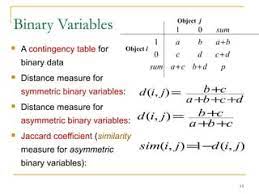
1. Euclidean distance: Calculates the straight-line distance between two points in space. Assumes that all dimensions of the data are equally important. It is the most commonly used distance metric.



1. Manhattan distance: Calculates the distance between two points in a grid based on the sum of the absolute differences of their coordinates. Ignores the diagonal distance between two points. Is often used when dealing with sparse datasets.



1. Asymmetric binary distance: Calculates the distance between two binary vectors based on the number of elements that differ between them. Can be used when comparing objects that do not have a natural notion of distance.



Euclidean and Manhattan distances:

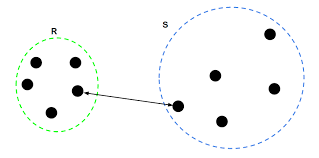
* Both treat all dimensions of data equally
* They take a sum over all dimensions of squared or absolute differences
* Data needs to be standardized if dimensions are not on the same scale

Hierarchical Clustering:

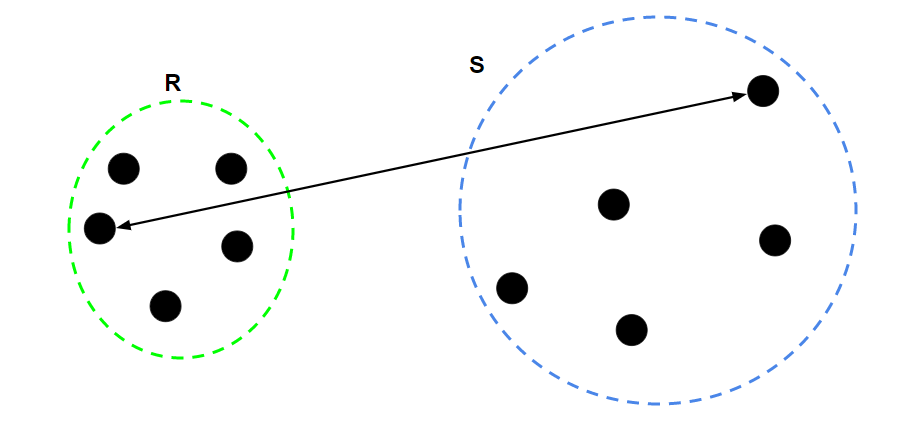
* Minimum number of segments that can be formed is 1, with the entire data set being part of one whole super cluster
* Maximum number that can be formed is N, where N is the number of possible available data points, with each observation being its own segment
* Clustering algorithms aim to find the number of segments K that lies between the two extremes
* Two types: Divisive clustering and Agglomerative clustering
* Both result in a sequence of nested partitions

Three types of linkages:

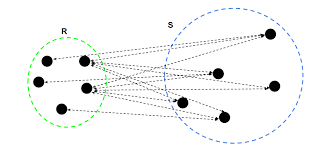
* Single linkage: Distance between the closest points belonging to two different sets of clusters.



* Complete linkage: Distance between two farthest points of the sets.



* Average linkage: Average distance between all the points in the two sets.



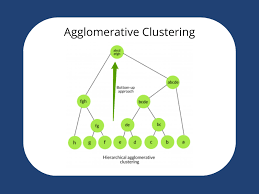
These are the three types of distances that hierarchal clustering algorithms use. You can either have Single linkage, Complete linkage or Average linkage distances. To be noted that this is an example of Agglomerative clustering.

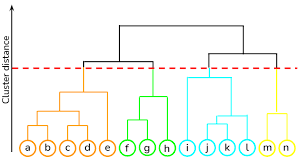
The steps for conducting agglomerative clustering is as follows:

The first step of course is declaring all the individual consumers as Individual clusters. The idea is to converge into one single supercluster after the process is complete. After a dendogram is ready we can decide upon the ideal number of market segments by observing the charts.

After declaring all the consumers as individual clusters we compute all possible distances from a point say xi with the rest of the clusters, (which at the time is n – 1, assuming the number of observations is n ) and then pick the point with the smallest distance with the point xi. This process is then repeated for all the points and we end up with n/2 clusters after the first iteration.

Now how the distance between clusters is computed is based upon the type of method you choose. If you choose singly linked method then the distances computed are from two closest points from two different clusters. If you choose Complete linked clustering then it’s two of the farthest points.





Hierarchical clusters are expressed through dendrograms: A dendrogram is a tree diagram . The root of the tree represents the one-cluster solution where one market segment contains all consumers. The leaves of the tree are the single observations and branches in-between correspond to the hierarchy of market segments formed at each step of the procedure. Dendrograms are often recommended as a guide to select the number of market segments, but rarely provide guidance of this nature because the data sets underlying the analysis are not well-structured enough.

Partitioning Methods

KMeans: An Iterative method used to determine the optimal k Market segments, where K is predetermined by the users. In this processs the input K is provided by the analyst and the algorithm elects K random data points from the set of records, which are declared as centroids. In the next step, each data record’s Euclidean distance is computed from all the individual centroids and the centroid which is the closest to the record is assigned a segment of belonging to that centroid. After repeating this for every single data point, the centroids are re-assigned by the Mean of the elected segments, and the entire process is repeated again, until the co-ordinates of the Segment centroids do not change.

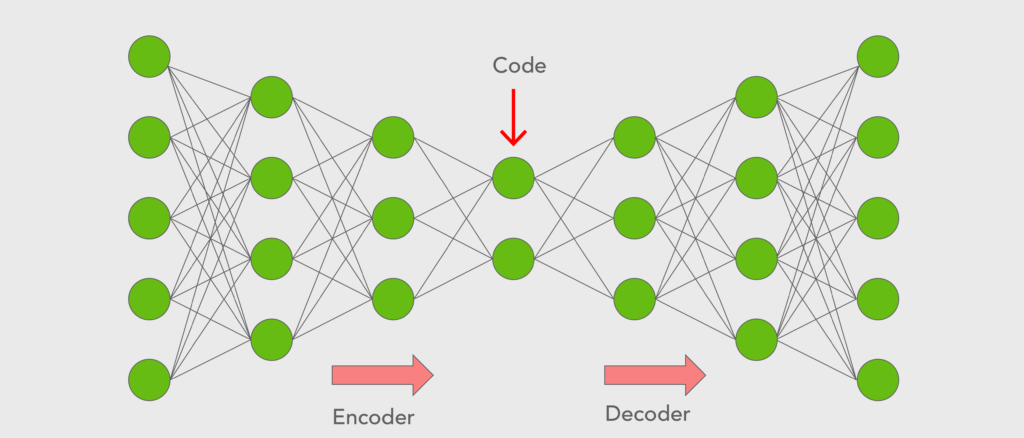
K-Medians is algorithm which uses the same procedure but instead of Euclidean distance it uses Manhattan distance and instead of re-assigning the segment centroids with the means of the newly updated segments it assigns the Medians.

The Neural gas Algorithm: The Neural Gas Algorithm doesn’t just adjust the primary centroid towards the consumer but also adjusts the secondary centroid by moving that closer to the user as well. This adjustment is known as Neural Gas Algorithm.

Before we delve into Model based clustering one more method that can be used for identifying familiar market segments that should be discussed are Autoencoders.

Autoencoders are neural networks that are trained to predict the input without using the trivial multiply by one solution. One common question that crosses people’s minds is why would anyone ever do that? Predicting the input is useless because we already have the input.

The answer lies in the architecture of the Neural Network.



Autoencoders consist of an Encoder and a Decoder. As you already know they predict the input variables using the input variable without the Trivial Multiply by one solution. The Encoder part converts the input numbers into smaller coded data which can be interpreted by the decoder to predict the input variables again. This technique is usually used for Image data to compress and decompress large image files or to extract useful features from the images and eliminate the noise.

But one useful attribute of autoencoders are that they can clear noise from your data. Which means that if you provide it with a data point that doesn’t behave like the rest of the data it will have large RECONSTRUCTION ERROR, making it an outlier. This type of technique can therefore be used in Clustering algorithms or for Anomaly detection.

Model Based Methods: The model based methods are based on Two primary assumptions

* 1. There exists a certain number of perfect market segments each of which have their own distribution.
  2. Each segment has a specific set of characteristics that makes it distinct compared to the rest of them.

Model based learning uses these two primary assumptions to compute the probability of a data point belonging to all the possible segments, and the one where the data point is most likely to belong to, is considered as the segment from which the record is coming from.

It assumes a type of distribution for the segments and then uses maximum likelihood to estimate the mean and co-variance of the segment that maximize the probability of getting a particular data record. After iteratively identifying the correct mean and Co-variance the model is trained and can be used to predict based on the probability of belonging to a certain segment.

Examples are GMM, VGMM and DPGMM.

Two types of Model based learning algorithms:

* 1. Finite Mixture of distributions: In finite mixture of distribution, instead of assuming one single distribution for all of the segments, it assumes that all the segments have their own individual distribution different from the rest. It uses the expected maximum (EM) algorithm to determine the co-variance and the mean of the distribution and after assignment classifies the records based on clusters.
  2. Finite Mixture of Regressions: In this method we use something called a response variable and assume that the data comes from a different type of regression models. Each segment following it’s own regression model of course. The data points that can provide the best predictions for the response variables in a segment are all clumped together into the same segment. The data points that increase oppose the co-variance are clumped out of the cluster.

Algorithms with Integrated Variable selection:

Feature selection is a tedious process but obtaining the proper features for segmentation is crucial to reduce noise. It becomes simple, when the segmentation variables are metric to elect features and disregard the irrelevant ones, but when it comes to binary variables it becomes difficult to choose the useful ones and eliminate the noise. To counter this issue we have the following two algorithms.

1. Biclustering: Binary data is harder to deal with than metric data due to reasons aforementioned. The 1’s in the segmentation variable might be useful information but the zeroes could act as noise as they don’t convey any relevant information. To counter this biclustering algorithm was introduced.

The biclustering algorithm is specifically designed to cluster binary segmentation variables, It does it by applying matrix consumer switching operations on the rows of a dataset to clump a maximum number of ones in the top left corner of the dataset.

Once the data records are clumped together the records at the top left corner are assigned to be in the same cluster and removed from the matrix. Then the same process Is repeated on the rest of the dataframe, to provide with similar clusters.

1. Variable Segmentation for based Clustering for Binary data (VSBD) algorithm: Just like the biclustering algorithm, variable segmentation for binary data algorithm removes noise from the existing data. It does it by iteratively computing the inertia of the clustering algorithm and finding the set of columns that have the least inertia.

It does this by first electing a subset of the entire dataset to reduce computational complexity, and elects the best combination of V Segmentation variables from all possible combinations. Once it has that, it keeps adding new columns to the dataset that reduce the inertia of the clustering, ending in the best subset of columns that give us clusters. The Kmeans cluster of course takes a K value for input but to compute the value of K, Lance ratio or Ratowsky’s ratio can be used.