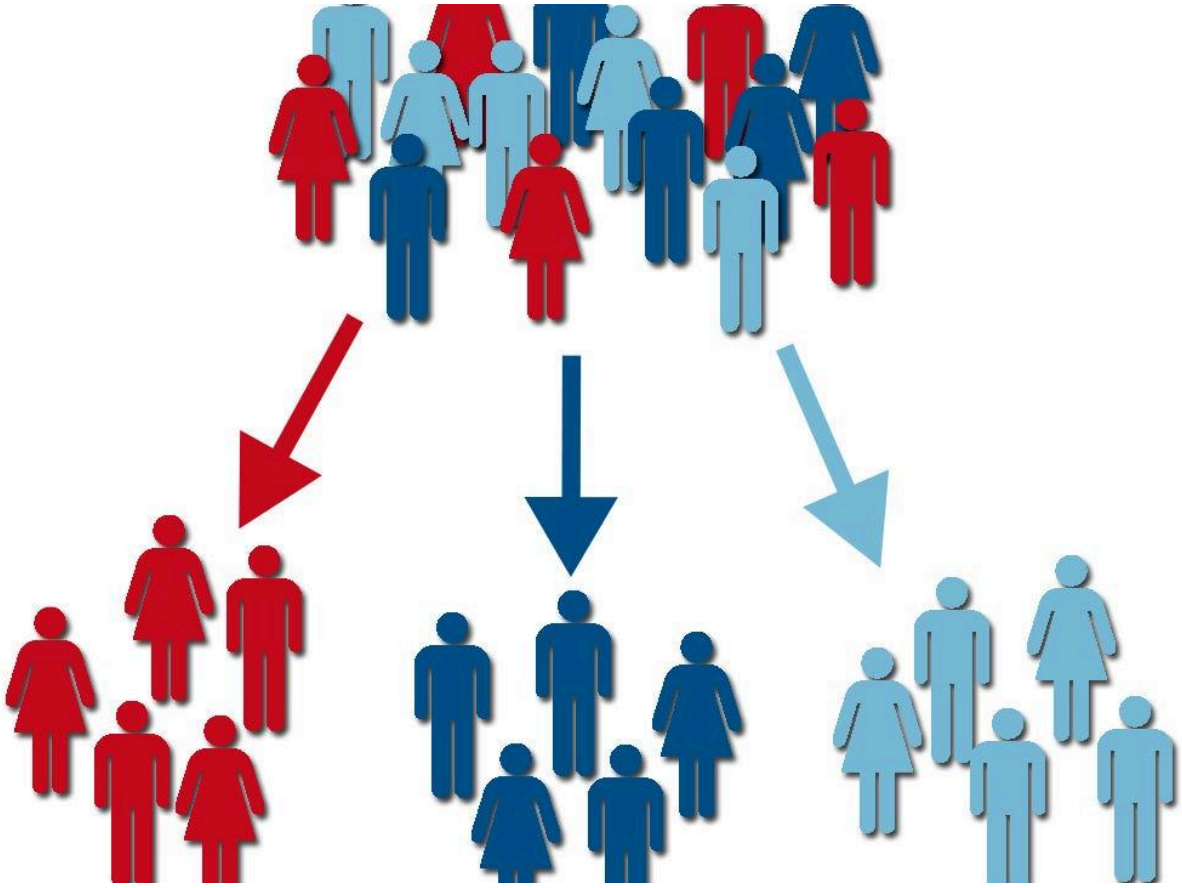
Market Segmentation – A Case Study



PANKAJ

04.04.2024

ABSTRACT

This document provides an overview of a widely-used marketing approach called Market Segmentation (MS), which involves categorizing an audience into smaller groups based on common traits. It discusses the essential aspects of MS, including its theoretical foundations and practical application through Python programming.

1 . Market Segmentation

1.1 **What is Market Segmentation**

Market Segmentation involves dividing a broad market into smaller groups with similar characteristics such as age, interests, and behavior. By categorizing customers based on these shared traits, businesses can create targeted marketing strategies that resonate more effectively with each group's specific needs and preferences.

1.2 Why is it important?

Market Segmentation is necessary as:

- It can help you to target just the people most likely to become customers of your company or consumers of your content/product.
- By understanding your market segments, you can leverage this targeting in product, sales, and marketing strategies.
- It can also power your product development cycles by informing how you create product offerings for different segments like men vs. women or high income vs. low income.

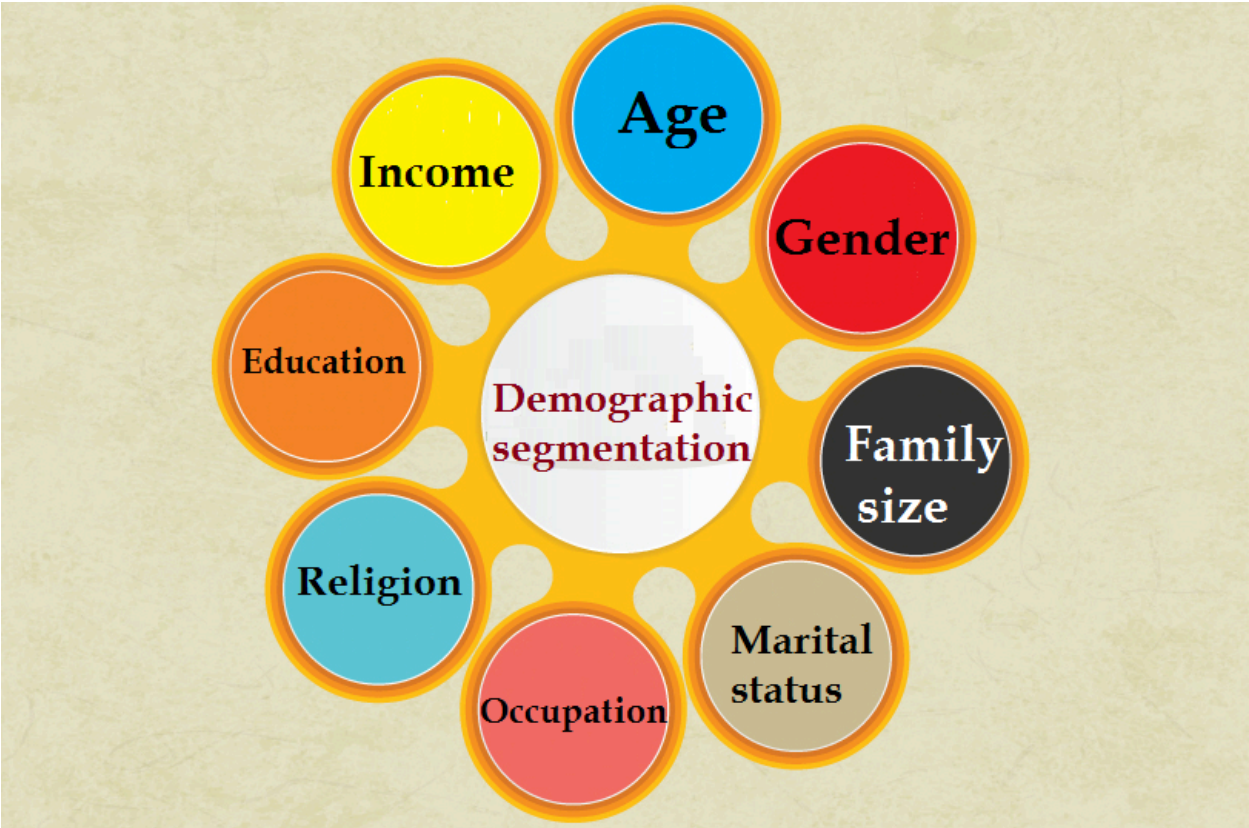
1.3 Types of Market Segmentation

There are different types of market segments that you can create. Below are the four major types of Market

1 . Demographic segmentation

Demographic segmentation is a market segmentation strategy that involves dividing a target audience based on demographic factors such as age, gender, income, education level, occupation, marital status, and family size. This approach helps businesses understand the characteristics and preferences of different demographic groups within their market, allowing them to tailor products, services, and marketing campaigns to better meet the specific needs and interests of each segment. Demographic segmentation is widely used due to its simplicity and effectiveness in reaching distinct

customer groups with targeted messaging and offerings.

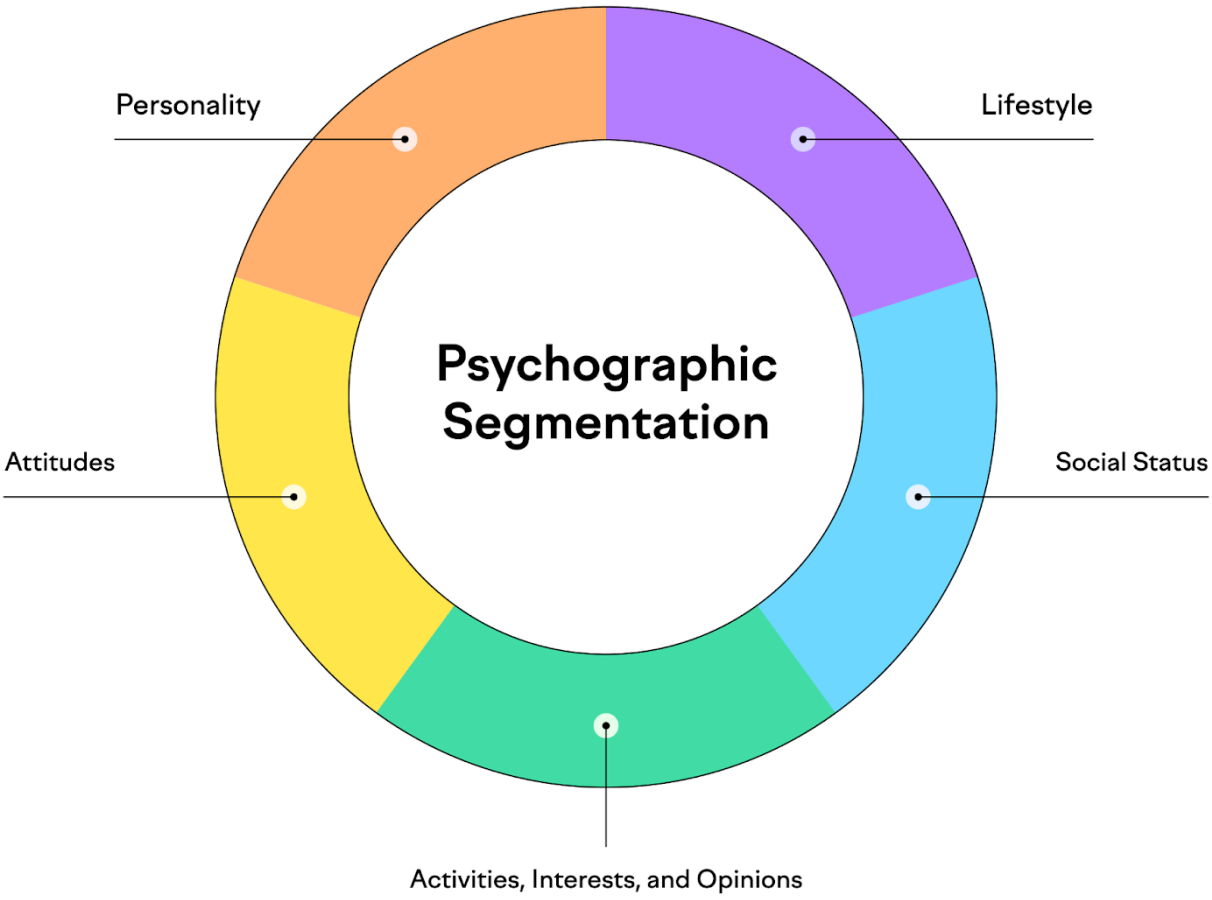


2. Geographic Segmentation

Geographic Segmentation splits up your target segment based on locations such as country, state etc. Customers can also be identified by taking into account the characteristics of the area they live in for example language, urban, suburban, rural etc.

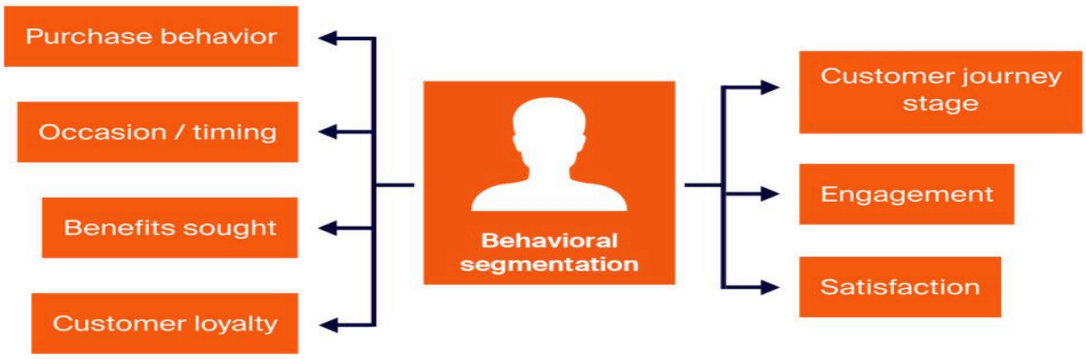
3. Psychographic Segmentation

Psychographic Segmentation splits the target market based on characteristics that are mental and emotional. Some examples of psychographic characteristics include personality traits, interests, beliefs, values, attitudes and lifestyles.



4 Behavioral Segmentation

Behavioral segmentation is a form of marketing segmentation that divides the target market based on behavioral patterns exhibited. This segmentation type studies the behavioral traits of consumers — their knowledge of, attitude towards, use of, likes/dislikes of, or response to a product, service, promotion, or brand



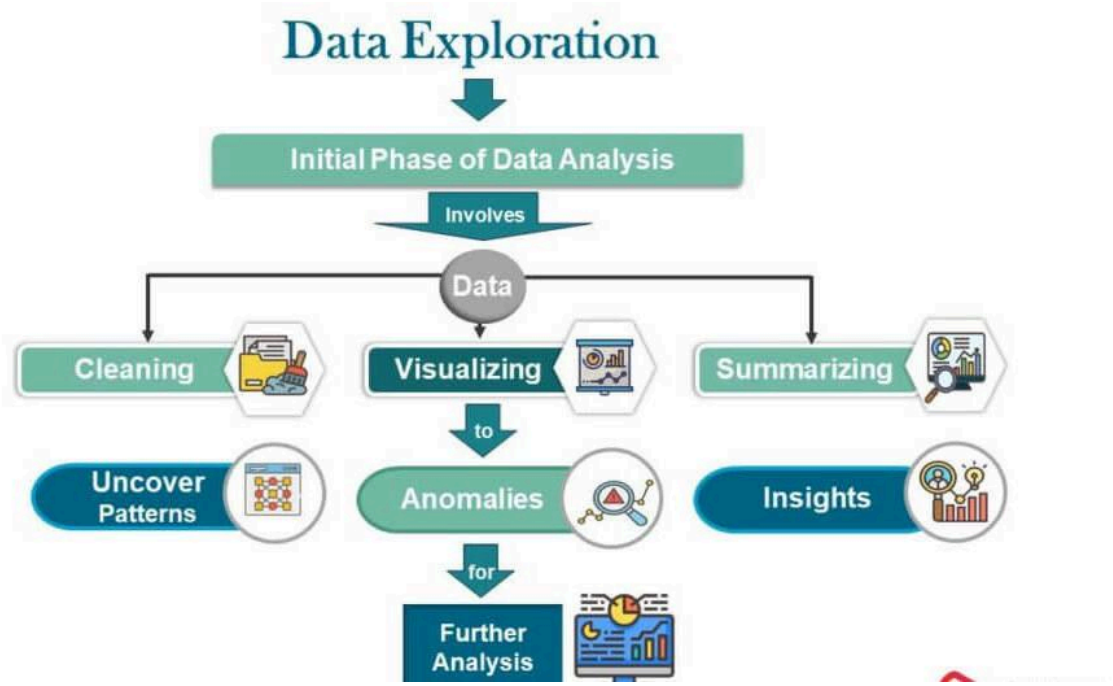
HOW

Following are the key points involved in Market Segmentation.

Data exploration

It is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. Data exploration helps to identify the measurement levels of the variables; investigate the univariate distributions of each of the variables; and assess dependency structures between variables. In addition, data may need to be pre-processed and prepared so it can be used as input for different segmentation algorithms. Results from the data exploration stage provide insights into the suitability of different

segmentation methods for extracting market segments.



Data cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. The first step before commencing data analysis is to clean the data. This includes checking if all values have been recorded correctly, and if consistent labels for the levels of categorical variables have been used. For many metric variables, the range of plausible values is known in advance.

Data Preprocessing

Numerical Variables

Numeric variables are often on different scales and cover different ranges, so they can't be easily compared. What's more, variables with large values can dominate those with smaller values when using certain modeling techniques. centering and scaling is a common pre-processing task that puts numeric variables on a common scale so no single variable will dominate the others. The

simplest way to center data is to subtract the mean value from each data point. Subtracting the mean centers, the data around zero and sets the new mean to zero.

Categorical Variables

Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones, if it makes sense to do so Exploring Data Merging levels of categorical variables is useful if the original categories are too differentiated (too many).

2.4 Descriptive Analysis

Descriptive Analysis is the type of analysis of data that helps describe, show, or summarize data points in a constructive way such that patterns might emerge that fulfills every condition of the data. It is one of the most important steps for conducting statistical data analysis. The three main types of descriptive statistics are frequency distribution, central tendency, and variability of a data set. The frequency distribution records how often data occurs, central tendency records the data's center point of distribution, and variability of a data set records its degree of dispersion. Helpful graphical methods for numeric data are histograms, box-plots, and scatter plots. Bar plots of frequency counts are useful for the visualization of categorical variables.

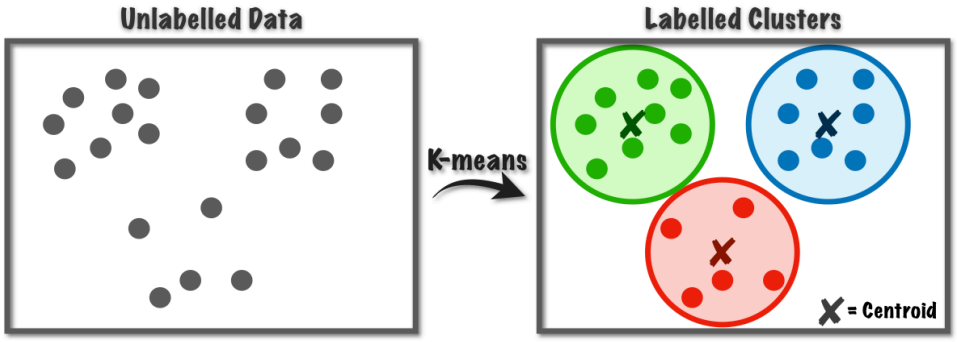
Principal Components Analysis

Principal component analysis, or simply PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

The K-Means Clustering Algorithm

What is this algorithm?

K-Means Clustering is an unsupervised learning algorithm, which groups the unlabelled dataset into different clusters. K defines the number of predefined clusters that need to be created in the process, so if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on. The figure below shows what K Means Clustering does.



How does it work?

The K Means Algorithm is implemented in the following steps:

1. Decide the number of clusters, i.e. K

2. Select K random points in the dataset. These points will be the centers of each of the K clusters and shall be called Centroids.

3. Assign each data point in the dataset to one of the K centroids, based on the point's distance from each of the centroids.

4. Consider this clustering to be correct and reassign the Centroids to the mean of these clusters

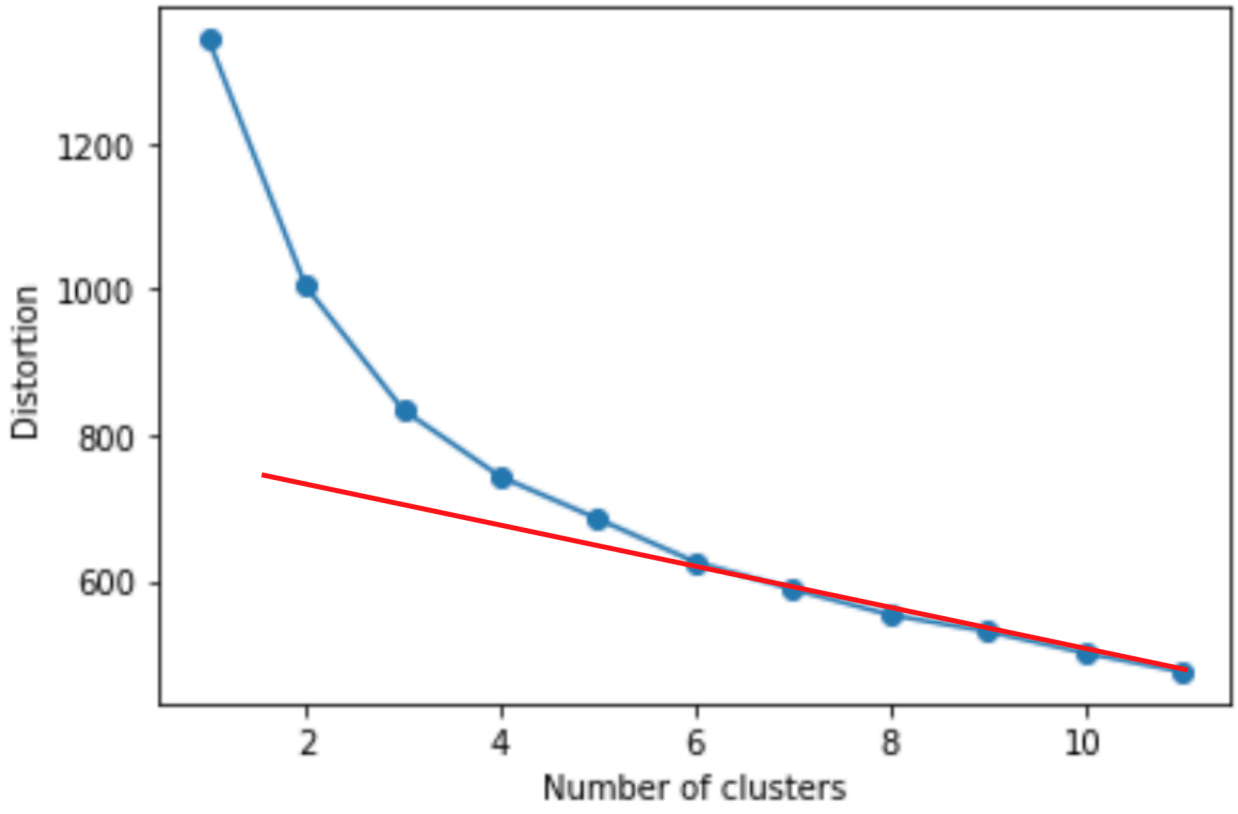
. 5. Repeat Step 3. If any of the points change clusters, Go to step 4. Else Go to step6

6. Calculate the variance of each of the clusters.

7. Repeat this clustering a specific number of times until the sum of variance of each cluster is minimum

The Elbow Method

Finding the ideal number of clusters to divide the data into is a critical stage in any unsupervised technique. One of the most prominent techniques for figuring out this ideal value of k is the elbow approach. It is probably the most well-known approach which involves calculating the sum of squares for each cluster size, graphing the results, and identifying the ideal cluster size by looking for an elbow where the slope changes from steep to shallow.



4 Why use this algorithm?

Each feature of the data could have varied values, increasing the overall variance of the feature. The main goal is to segment our data based on like-values features. Clustering algorithms separate the data into clusters based on their values, i.e. values belonging to a similar range will be assigned to the same cluster.

Advantages of K-Means Clustering

- Relatively simple to implement
- Scales to large data sets
- Guarantees convergence
- Generalises clusters to different shapes and sizes, such as elliptical clusters

The image below shows the Clustering Algorithm repeated for several iterations until the minimum sum of variance of each cluster is achieved.