Market Segmentation - A Case Study

Summary Of Step 1, 2 And 3.

Submitted by:

Iman Chandra

Table Of Contents

Market Segmentation – A Case Study

Abstract:

This report provides a detailed overview of **Market Segmentation (MS)**—a widely used marketing strategy that involves dividing a broad consumer or business market into smaller, more manageable segments based on shared characteristics such as demographics, behaviors, or preferences. It explains the core concepts behind market segmentation and explores its practical applications. In particular, the report demonstrates how segmentation can be performed using **Python**, highlighting data-driven techniques for identifying and analyzing market segments efficiently.

1. Market Segmentation

1.1 What is Market Segmentation?

Market segmentation is the process of dividing a target market into distinct subgroups or segments based on shared characteristics such as demographics, needs, priorities, interests, behaviours, or psychographic traits. These segments help businesses better understand their audience and tailor their marketing efforts to meet the specific needs of each group.

1.2 Why is Market Segmentation Important?

- Market segmentation plays a key role in helping businesses grow and succeed. Instead of trying to appeal to everyone, it allows you to focus on the people who are most likely to be interested in your product or service. This means you can spend your time and resources more efficiently.

- By understanding your market segments, you can tailor your marketing, sales, and product strategies to better meet the specific needs of each group. For example, your approach for high-income customers might be very different from your strategy for budget-conscious buyers.

- Segmentation also helps guide product development. Knowing what different groups want makes it easier to design and improve products that truly meet their expectations—like creating different versions for men and women, or for younger and older users.

- Most importantly, it helps boost profits. When you give people exactly what they're looking for, they're more likely to buy, come back, and recommend your brand to others.

1.3 Types of Market Segmentation

There are several ways to segment a market, depending on the kind of product or service and the audience you're trying to reach. Below are the four major types of market segmentation:

1. Demographic Segmentation

This type of segmentation technique splits the target audience based on people-based differences. These factors include things like age, sex, marital status, family size, occupation, education level, income, race, nationality and religion.

Example: A luxury car brand may target high-income professionals aged 35–55.

2. Geographic Segmentation

This groups people based on their location—such as country, state, city, climate, or even neighbourhood. It's useful for businesses that serve specific areas or tailor their offerings based on regional preferences.

Example: A clothing brand offering winter wear in colder regions and light wear in tropical areas.

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.

Example: A brand selling eco-friendly products targeting environmentally conscious consumers.

4. Behavioural Segmentation

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

Example: A streaming service offering special deals to binge-watchers or frequent users.

These types can also be combined for even more precise targeting. For instance, a company might focus on young (demographic), urban (geographic), fitness enthusiasts (psychographic) who buy activewear frequently (behavioural).

2. The How

Following are the key points involved in Market Segmentation:

2.1 Data Exploration

Data exploration is the first and one of the most crucial steps in data analysis. It involves examining and visualizing the dataset to uncover initial patterns, trends, and potential insights. This step helps analysts understand the structure and quality of the data before applying any segmentation techniques.

During data exploration, you:

- Identify the types and measurement levels of variables (e.g., categorical, numerical).
- Analyz the univariate distributions (how each variable behaves on its own).
- Explore relationships and dependencies between variables (e.g., correlations or groupings).

In many cases, the data also needs pre-processing such as cleaning, handling missing values, or transforming variables so that it can be properly used with segmentation algorithms.

The insights gained from this stage guide the choice of suitable segmentation methods and help ensure that the analysis is both meaningful and effective.

2.2 Data Cleaning

Data cleaning is a vital step in preparing your dataset for analysis. It involves identifying and correcting issues such as missing values, duplicate records, formatting errors, incorrect entries, or inconsistencies in labelling.

When data is collected from multiple sources, there's a high chance of duplication, mismatched labels, or corrupted entries. That's why cleaning the data is often the first thing you do before beginning any meaningful analysis.

Some key tasks in this step include:

- Verifying data entry: Ensuring all values have been recorded accurately.
- Standardizing categories: Making sure categorical variables use consistent labels (e.g., "Male" vs. "M").
- Handling outliers or invalid values: For many numeric variables, expected value ranges are known, so values outside those ranges can be flagged and reviewed.

A clean and consistent dataset not only ensures the accuracy of your segmentation results but also makes the modelling process much smoother and more reliable.

2.3 Data Preprocessing

Before applying segmentation algorithms, the data must be pre-processed to ensure it's in a suitable format. This involves transforming both numerical and categorical variables to make them compatible with machine learning techniques and to avoid bias in the results.

2.3.1 Numerical Variables

Numerical variables often exist on different scales—some might range from 0 to 1,000, while others range between 0 and 10. When using models that rely on distance calculations (like k-means), variables with larger values can dominate those with smaller values, leading to skewed results.

To avoid this, cantering and scaling is commonly applied:

- Cantering: Subtracting the mean value from each data point shifts the data to have a mean of zero.
- Scaling: Dividing by the standard deviation (or another value) ensures all variables are on a common scale.

This standardization process allows fair comparison across variables and ensures no single variable overpowers the rest in the segmentation model.

2.3.2 Categorical Variables

Categorical variables require different preprocessing steps. Two common approaches include:

- Merging similar levels: If a categorical variable has too many categories, some of them may be combined (or grouped) to reduce noise and improve model performance. This is especially useful when some categories have very few observations.
- Encoding categories: Categorical variables often need to be converted into numeric format to be used in machine learning models. This can be done using techniques like:
    - Label Encoding (for ordinal data)
    - One-Hot Encoding (for nominal data)

These transformations help machine learning algorithms interpret the categorical data in a way they can process effectively.

2.4 Descriptive Analysis

Descriptive analysis is a fundamental step in understanding the basic features of a dataset. It helps summarize and present data in a meaningful way, making it easier to identify patterns, trends, and anomalies before applying more advanced analytical techniques.

This type of analysis doesn't make predictions—it simply describes what's in the data. It is especially useful in preparing the dataset for deeper exploration or modelling.

There are three main types of descriptive statistics:

1. Frequency Distribution: Shows how often each value or category occurs in the dataset.

2. Measures of Central Tendency: Indicates the central point of the data, typically using the mean, median, or mode.
3. Measures of Variability: Describes how spread out the data is, using statistics like range, variance, and standard deviation.

To visualize and better understand the data, several graphical tools are commonly used:

- Histograms, box plots, and scatter plots for numeric variables.
- Bar charts for showing the frequency of categorical variables.

These tools help in spotting outliers, trends, and groupings that may influence segmentation decisions later in the analysis process.

2.5 Principal Component Analysis (PCA)

- PCA is a dimensionality reduction technique used to simplify large datasets.
- It transforms a large set of variables into a smaller set of uncorrelated variables called *principal components*.
- The goal is to retain as much of the original information (variance) as possible while reducing complexity.
- Helps in removing redundant or less informative variables, which may slow down or mislead analysis.
- First principal component captures the maximum possible variance in the data; each subsequent component captures the remaining variance.
- Useful for:
  - Easier data visualization (especially in 2D or 3D).
  - Reducing noise in the data.
  - Improving performance of machine learning algorithms by eliminating irrelevant features.
- There is a small trade-off in accuracy, but it's often acceptable for the benefit of speed and simplicity.

2.6 The K-Means Clustering Algorithm

2.6.1 What is this algorithm?

- K-Means Clustering is an unsupervised machine learning algorithm used to group data into clusters based on similarity.
- It works on unlabelled data, meaning the data has no predefined categories or outputs.
- The number K refers to the number of clusters you want to form.
  - If K = 2, the algorithm forms 2 clusters.

    - If K = 3, it forms 3 clusters, and so on.
- Each cluster contains data points that are more similar to each other than to those in other clusters.
- The algorithm tries to minimize the distance between data points and their respective cluster centre (called the centroid).

2.6.2 How Does It Work?

The K-Means algorithm follows these steps:

1. **Choose the number of clusters (K):**
   Decide how many clusters you want to divide the data into.
2. **Initialize centroids:**
   Randomly select K data points from the dataset as the initial centroids (cluster centers).
3. **Assign data points to the nearest centroid:**
   Measure the distance of each data point to each centroid and assign it to the closest one.
4. **Update centroids:**
   Recalculate the centroid of each cluster by taking the **mean** of all data points assigned to that cluster.
5. **Repeat assignment:**
   Reassign each data point to the new nearest centroid.
    a. If any data point has changed clusters, go back to Step 4.
    b. If no points change clusters, move to Step 6.
6. **Calculate cluster variance:**
   Measure the variance (spread) within each cluster.
7. **Repeat the full process:**
   Optionally repeat the entire process multiple times (with different random starting centroids) to minimize the **total within-cluster variance** and find the best clustering result.

2.6.3 The Elbow Method

- Choosing the right number of clusters (K) is a key step in K-Means clustering.
- The Elbow Method is a widely used technique to help determine the optimal value of K.

How it works:

1. Run K-Means for a range of cluster numbers (e.g., K = 1 to 10).
2. For each value of K, calculate the Within-Cluster Sum of Squares (WCSS):
    a. WCSS measures how compact the clusters are (lower is better).
3. Plot a graph with:
    a. X-axis: Number of clusters (K)
    b. Y-axis: WCSS (sum of squared distances)
4. Look for the "elbow point" — the value of K where the WCSS begins to decrease more slowly.
    a. This point represents a good balance between model simplicity and accuracy.

2.6.4 Why Use This Algorithm?

K-Means clustering is effective for grouping data points based on feature similarity. When data features have varying values, this algorithm helps segment the dataset into meaningful clusters by minimizing within-cluster variance.

Purpose:

- To group data points with similar feature values into the same cluster.
- Helps uncover natural groupings or structures within the data.
- Reduces complexity for analysis and modelling.

Advantages of K-Means Clustering:

- **Simple and easy to implement:**
  The algorithm follows a straightforward, iterative approach.
- **Efficient for large datasets:**
  It scales well and performs quickly even on high-volume data.
- **Guaranteed convergence:**
  The algorithm will always converge to a result (though it may be a local optimum).
- **Adaptable to different cluster shapes:**
  Can approximate various cluster geometries (e.g., spherical or elliptical) depending on the data distribution.

3. Market Segmentation Case Study on McDonalds Dataset

Kindly refer to any of the following GitHub links for the complete code implementation.

GitHub: -

5. Conclusion

Market segmentation is a core element of modern marketing, playing a vital role in helping businesses define their target audiences, shape compelling value propositions, set effective pricing strategies, and plan impactful communication campaigns. While the idea of dividing a market into smaller groups may seem straightforward, putting it into practice can be quite complex. It often requires careful analysis, the right tools, and a deep understanding of various customer factors.

Research shows that many marketers struggle to seamlessly integrate segmentation into their broader marketing strategies. This highlights the ongoing need to explore new segmentation variables and develop fresh, data-driven approaches that reflect evolving consumer behavior.

When done well, market segmentation allows businesses to create more personalized, targeted campaigns that resonate with specific audiences. This not only improves customer engagement but also strengthens a company's position in an increasingly competitive marketplace.

However, the success of segmentation efforts largely depends on the quality of the data used. Reliable, accurate, and up-to-date data—presented in a clear and usable format—is essential for making informed marketing decisions.

In short, market segmentation helps businesses better understand their customers, predict their needs, and make smarter strategic choices. In today’s fast-moving, data-focused world, it’s a powerful tool for driving growth and staying ahead of the curve.

References:

Dolnicar, S., Grün, B., & Leisch, F. (2018). *Market segmentation analysis: Understanding it, doing it, and making it useful*. Springer. https://www.springer.com/series/10101

Varini, K. (2011, October). *Market segmentation: Does it work?* Paper presented at the 29th EuroCHRIE Conference, Dubrovnik, Croatia.