# A Case Study on Market Segmentation

by

**Akhilesh Kumar Shah**

**Project: 2.0** @Feynn Labs ML Intern
**Date of Submission:** 27-May-2024

# Contents

# Market Segmentation – A Case Study

**Abstract**
This report provides a concise introduction to a marketing strategy widely known as Market Segmentation (MS), which involves grouping an audience into subgroups based on shared characteristics. It outlines the fundamental concepts of MS, detailing both theoretical aspects and practical implementation using Python. Additionally, the report discusses the importance of MS in modern marketing practices, illustrating how it helps businesses tailor their strategies to better meet the needs of specific customer segments. By leveraging MS, companies can enhance their marketing effectiveness, improve customer satisfaction, and drive business growth. This report includes a case study on McDonald's, demonstrating real-world application of MS techniques.

## 1. Market Segmentation

### 1.1 What is Market Segmentation?

Market Segmentation involves identifying distinct segments within a target market and dividing this market into subgroups based on various factors such as demographics, needs, priorities, common interests, and other psychographic and behavioral criteria. This process helps businesses gain a deeper understanding of their target audience by recognizing the unique characteristics and preferences of each segment. By doing so, companies can develop tailored marketing strategies that resonate more effectively with each subgroup, leading to better engagement and conversion rates.

**Fig:** Market Segmentation

### 1.2 Why is it important?

- Market segmentation is essential because: It enables you to focus on the individuals who are most likely to use your content or purchase your product.
- You can use this targeting in product, sales, and marketing initiatives by being aware of your market segments.
- It can also improve your cycles of research and development by helping you design products for distinct markets, such as a given gender or economic bracket (high-income vs. low-income).
- Above all, it can increase revenue [1, 2, 3].

### 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 Segmentation.

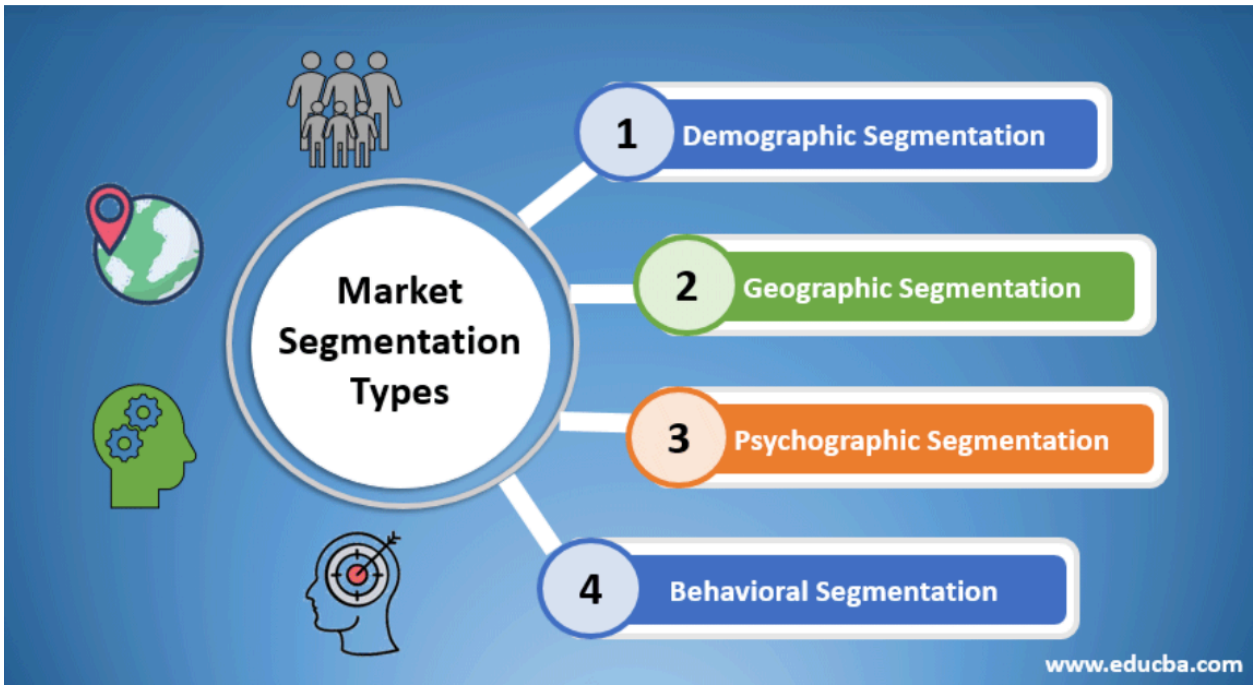


**Fig:** Methods of Market Segmentation

#### 1.3.1 Geographic Segmentation

Your target segment is divided up via segmentation according to geographical factors like nation, state, etc. Customers are also identifiable by considering the features of their locality, such as language, urban/suburban, rural, etc.

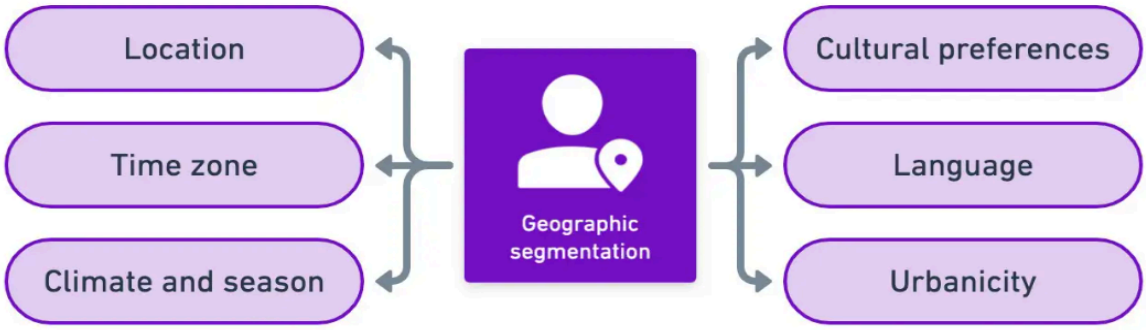


**Fig:** Geographic Segmentation

### 1.3.2 Demographic Segmentation

The target audience is divided using this kind of segmentation strategy according to individual differences. Age, sex, marital status, number of children, employment, income, education, and nationality are a few examples of these variables.

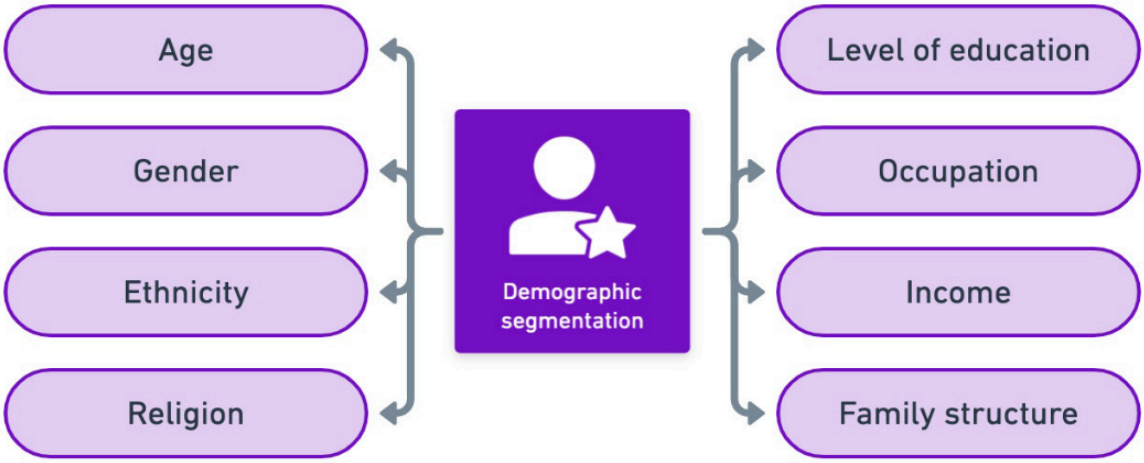


**Fig:** Demography Segmentation

### 1.3.3 Psychographic Segmentation

Psychographic Segmentation divides the target market according to psychological and emotional traits. Psychographic qualities encompass several aspects such as personality traits, interests, beliefs, values, attitudes, and lifestyles.

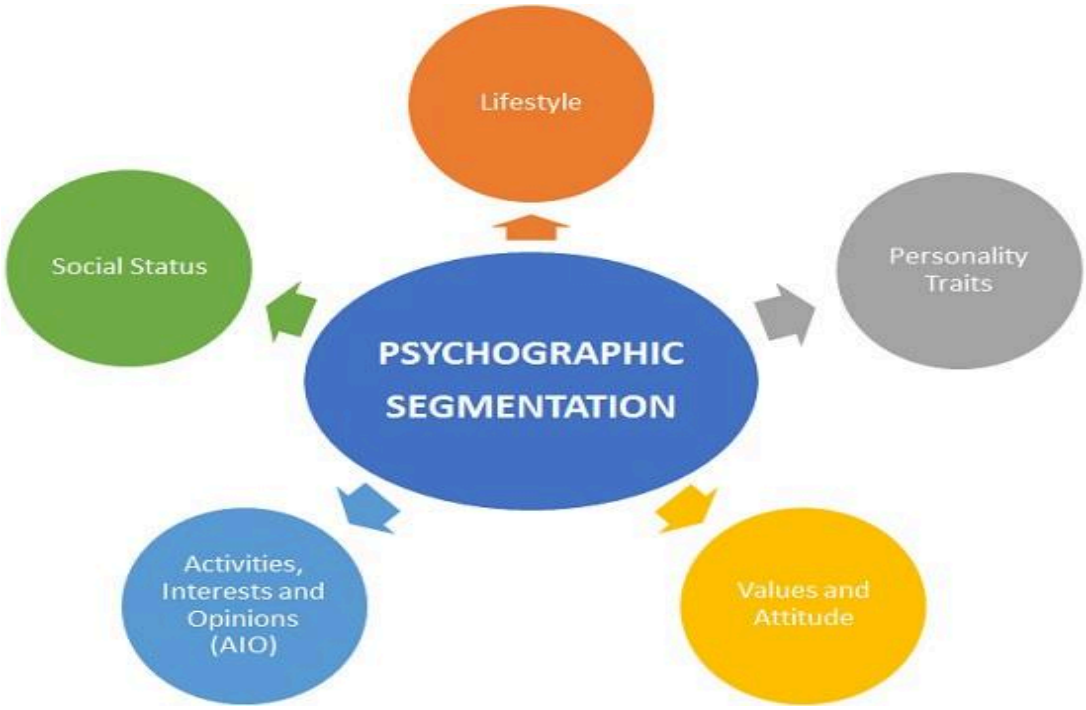


**Fig:** Psychographic Segmentation

### 1.3.4 Behavioral Segmentation

A type of marketing segmentation known as "behavioral segmentation" splits the target market according to demonstrated behavioral patterns. This sort of segmentation looks at the ways that customers behave, including how they respond to a product, service, promotion, or brand, and how they know about it, feel about using it, and what they like and dislike about it.

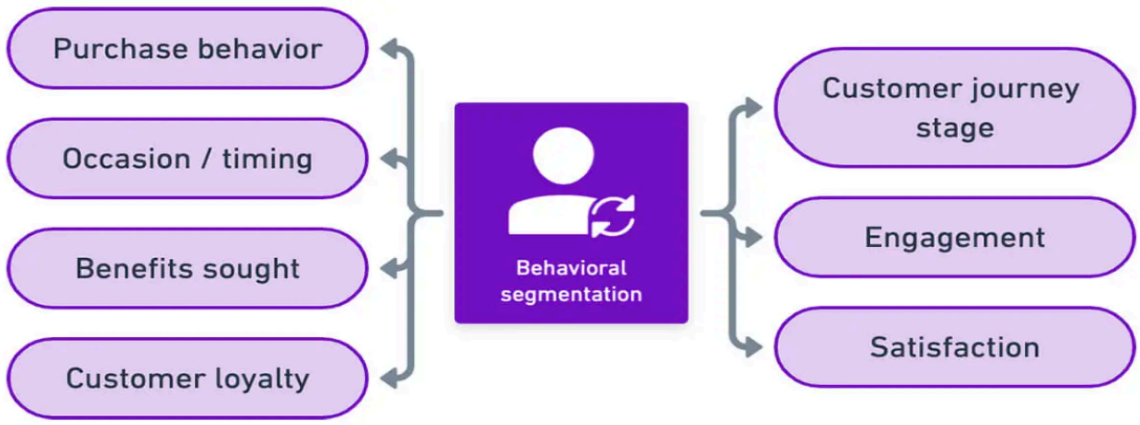


**Fig:** Behavioral Segmentation

## 2. The Method

These are the main ideas behind market segmentation.

### 2.1 Data Exploration

To find insights right away or to pinpoint regions or patterns that require more investigation, this initial stage of data analysis involves exploring and visualizing the data. Finding the measurement levels of the variables, looking into the univariate distributions of each variable, and evaluating the dependency structures between variables are all made easier with the use of data exploration. Furthermore, preprocessing and preparation of the data may be necessary to make it suitable for use as input in various segmentation techniques. The outcomes of the data exploration phase shed light on which segmentation techniques are most appropriate for identifying distinct market segments.

### 2.2 Data Cleaning

The practice of repairing or eliminating inaccurate, corrupted, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. It is common for data to be mislabeled or duplicated when integrating different data sources. Cleaning the data is the first step before starting data analysis. This entails verifying that all values have been accurately recorded and that consistent labels have been applied to the categorical variable levels. The range of possible values for many metric variables is known ahead of time.

### 2.3 Data Preprocessing

#### 2.3.1 Numerical Variables

Numerical variables are difficult to compare because they frequently have various ranges and scales. Furthermore, when employing specific modeling methodologies, variables with larger values may take precedence over those with smaller values. A frequent pre-processing step is centering and scaling, which places numerical variables on a common scale to prevent any one variable from dominating the others. Subtracting the mean value from each data point is the simplest method for centering data. The new mean is set to zero by subtracting the mean centers and the data surrounding zero.

#### 2.3.2 Categorical Variables

Typically, two pre-processing steps are used for categorical variables. Before additional analysis, one is combining levels of categorical variables; the other is, if appropriate, converting category variables to numeric ones. Examining Information It can be helpful to combine categorical variable levels if the initial categories were overly distinct (i.e., there were too many).

### 2.4 Descriptive Analysis

The kind of data analysis known as descriptive analysis aids in the constructive description, display, or summarization of data points so that patterns that satisfy all of the data's requirements may show up. It is among the most crucial procedures for analyzing statistical data. The variability of a data set, central tendency, and frequency distribution are the three primary categories of descriptive statistics. The data's frequency distribution indicates how frequently it happens, its central tendency indicates where its distribution center is, and the variability of a data set indicates how dispersed it is. For numerical data, scatter plots, box plots, and histograms are useful graphical techniques. Frequency count bar graphs help visualize categorical variables.

### 2.5 Principal Components Analysis

Principal component analysis (PCA) is a method used for reducing the dimensionality of large datasets. It achieves this by transforming a large set of variables into a smaller set that retains most of the original information. While this reduction in variables can lead to a slight loss in accuracy, the goal of dimensionality reduction is to balance this trade-off by enhancing simplicity. Smaller datasets are easier to explore, visualize, and analyze, making them more efficient for machine learning algorithms to process without unnecessary variables. In PCA, the first principal component captures the most variability, followed by the second principal component, which captures the next highest amount of variability, and so on.

### 2.6 The K-Means Clustering Algorithm

#### 2.6.1 What is this algorithm?

K-Means Clustering is an unsupervised learning algorithm used to group unlabeled datasets into distinct clusters. The parameter K specifies the number of clusters to form during the process. For instance, if K is set to 2, the algorithm will create two clusters; if K is set to 3, it will create three clusters, and so forth. The illustration below demonstrates the K-Means Clustering process.

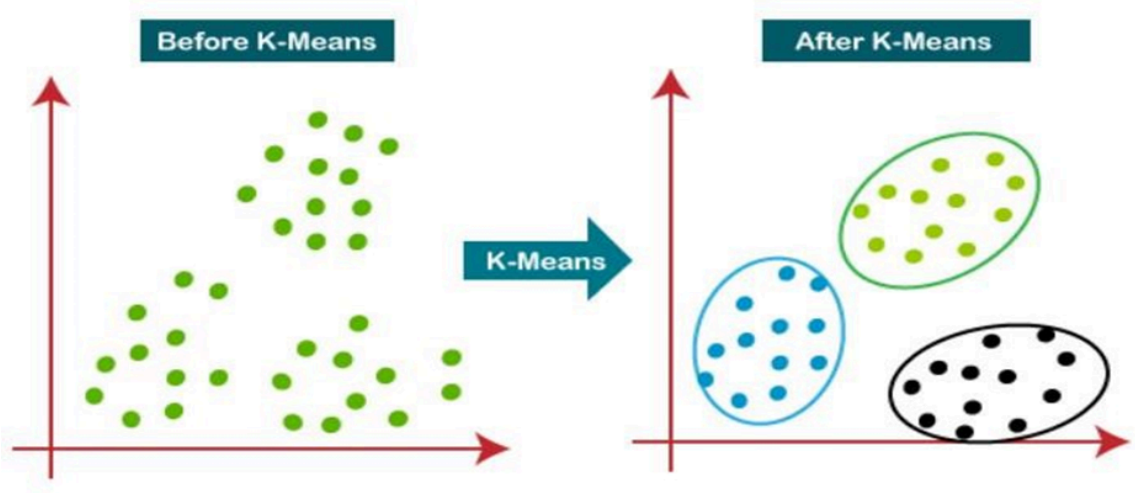


**Fig:** K-Means Clustering

### 2.6.2 How does it work?

The K-Means Algorithm operates through the following steps:

1. Determine the number of clusters, denoted as K.
2. Randomly select K points from the dataset to serve as the initial centers of the K clusters, known as Centroids.
3. Assign each data point to the nearest centroid based on the distance between the point and each centroid.
4. Once all points are assigned, update the position of each centroid to the mean of the points in its cluster.
5. Repeat Step 3 with the updated centroids. If any data points change their assigned cluster, return to Step 4. Otherwise, proceed to Step 6.
6. Calculate the variance for each cluster.
7. Repeat the clustering process a predetermined number of times, or until the sum of the variances of all clusters is minimized.

### 2.6.3 The Elbow Method

Determining the optimal number of clusters for dividing data is a crucial step in any unsupervised learning technique. One of the most popular methods for identifying this optimal value of K is the Elbow Method. This approach involves calculating the sum of squared errors (SSE) for different cluster sizes, plotting these values on a graph, and identifying the ideal number of clusters by locating the "elbow" point where the slope of the graph changes from steep to shallow.

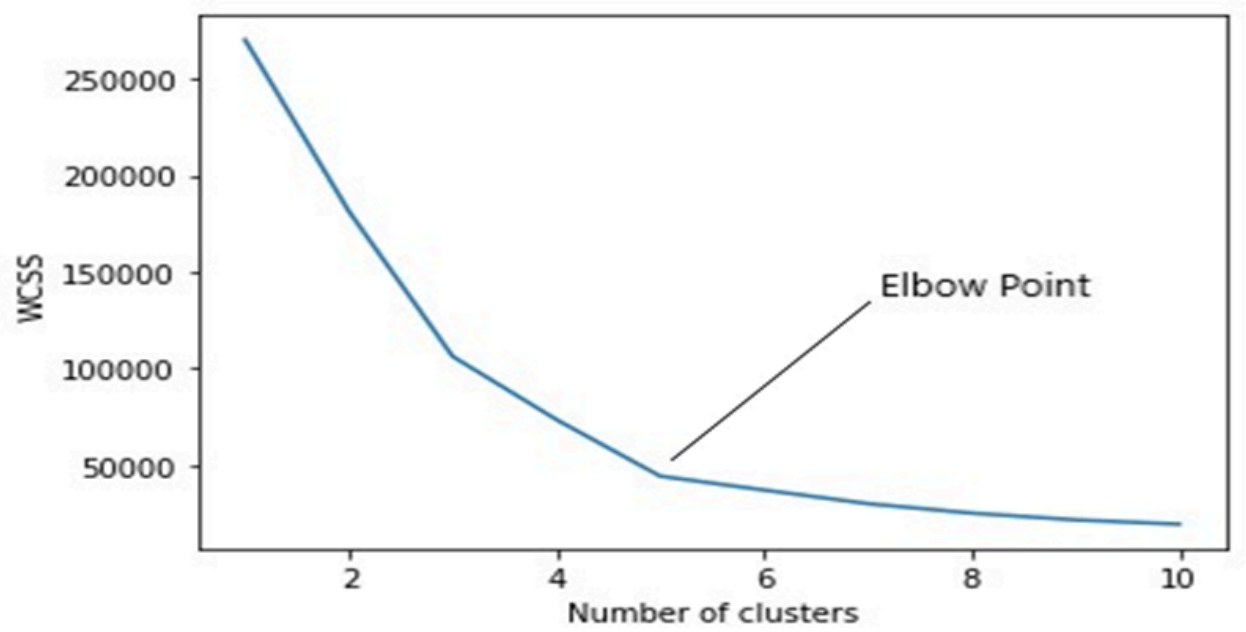


**Fig:** The Elbow Method

**2.6.4 Why use this algorithm?**

Data features can have varied values, leading to high overall variance. The main goal of clustering is to segment the data based on similar feature values. Clustering algorithms, such as K-Means, group data into clusters based on these values, so that similar values are assigned to the same cluster.

**Advantages of K-Means Clustering:**

- Relatively simple to implement
- Scales well to large datasets
- Guarantees convergence
- Adapts to clusters of different shapes and sizes, such as elliptical clusters

The image below illustrates the K-Means Clustering Algorithm repeated over several iterations until the minimum sum of variances for each cluster is achieved.

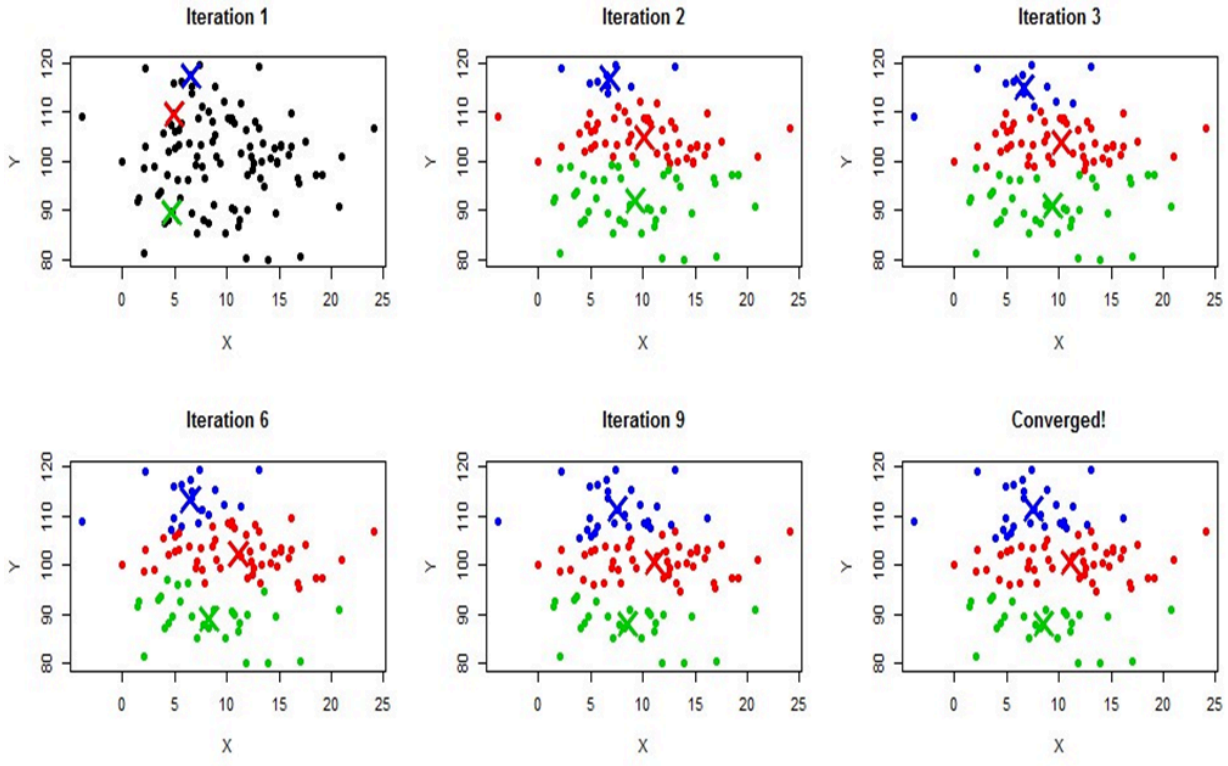


**Fig:** K-Means Clustering in action

## 3. Market Segmentation Case Study on McDonald's Dataset

For the complete code implementation, please refer to any of the following GitHub links.

**GitHub Link:** (Akhilesh Kumar Shah)
https://github.com/Akhushah/Feynn_Labs_Internship_2024/blob/main/Akhilesh_Market_Segmentation_Mcdonalds.ipynb

**Dataset Link:** (McDonald's Dataset)
https://github.com/Akhushah/Feynn_Labs_Internship_2024/blob/main/mcdonalds_dataset.csv

## 4. Conclusion

This study highlights that market segmentation is considered essential by marketing practitioners for various reasons, including targeting, proposition development, price formulation, and mass communication strategies. Although the concept of segmentation is straightforward, the process is complex and requires careful consideration of multiple factors. Literature indicates that many marketers are concerned about the practical implementation and integration of segmentation into marketing strategies. Future research priorities in segmentation include selecting and incorporating new variables into segmentation models and developing innovative segmentation strategies.

Market segmentation enables companies to identify their target audiences and personalize marketing campaigns more effectively. This approach is crucial for maintaining a competitive edge, as it allows businesses to understand their customers, anticipate their needs, and seize growth opportunities. By utilizing this powerful technique, companies can enhance decision-making, optimize marketing efforts, and improve their bottom line.

The success of market segmentation heavily depends on data quality. Therefore, it is vital to choose your data provider carefully, ensuring you have access to the latest industry information presented in accessible and easy-to-understand formats.

## References

[1] https://link.springer.com/book/10.1007/978-981-10-8818-6
[2] https://rb.gy/gn7kuw
[3] https://rb.gy/ehf02b
[4] https://rb.gy/1azcm9