
Market Segmentation – A Case Study

ON McDonalds Dataset

Shubham Penti

Date: 18-02-2024

Contents

Abstract.....	1
1. Market Segmentation.....	1
1.1 What is Market Segmentation.....	1
1.2 Why is it important.....	2
1.3 Types of Market Segmentation.....	2
1.3.1 Geographic Segmentation.....	3
1.3.2 Demographic Segmentation.....	3
1.3.3 Psychographic Segmentation.....	4
1.3.4 Behavioral Segmentation.....	4
2. The How.....	5
2.1 Data Exploration.....	5
2.2 Data Cleaning.....	5
2.3 Data Preprocessing.....	6
2.3.1 Numerical Variables.....	6
2.3.2 Categorical Variables.....	6
2.4 Descriptive Analysis.....	6
2.5 Principle Component Analysis (PCA).....	6
2.6 K-Means Clustering.....	7
2.6.1 What is this algorithm.....	7
2.6.2 How does it work.....	7
2.6.3 The Elbow Method.....	8

2.6.4 Why to use this algorithm.....	8
3. Market Segmentation Case Study on McDonalds Dataset.....	9
4. Conclusion.....	10
5. References.....	10

Market Segmentation – A Case Study

Abstract

This report represents a brief introduction to a marketing strategy popularly known as Market Segmentation (MS) which is the process of grouping an audience into subgroups based on shared characteristics. It also highlights the key points related to MS along with their theory and practical implementation using Python.

1. Market Segmentation

1.1 What is Market Segmentation?

Market Segmentation is the process of identifying segments of the target market and then dividing that target market into subgroups based on different factors such as demographics, needs, priorities, common interests, and other psychographic and/or behavioral criteria used to understand the target audience.



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.
- More importantly, It can enhance profits [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.

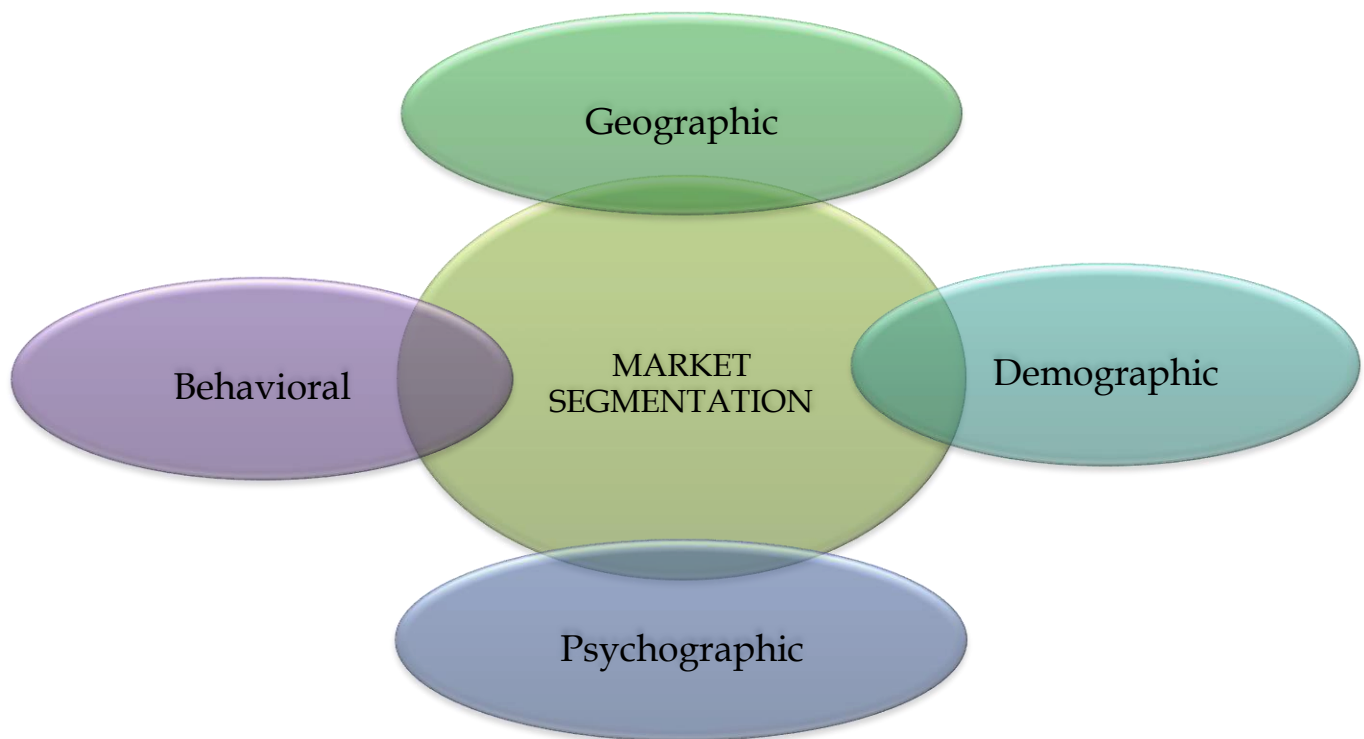


Fig: Methods of Market Segmentation

1.3.1 Geographic Segmentation

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.

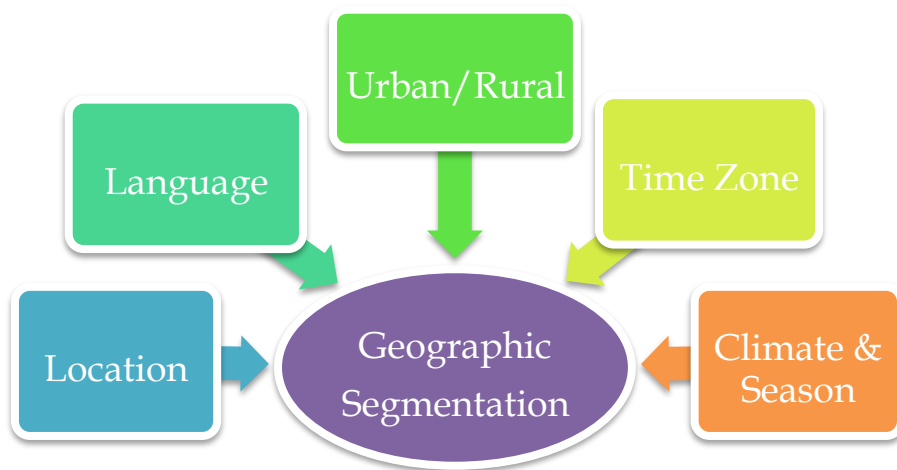


Fig: Geographic Segmentation

1.3.2 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.

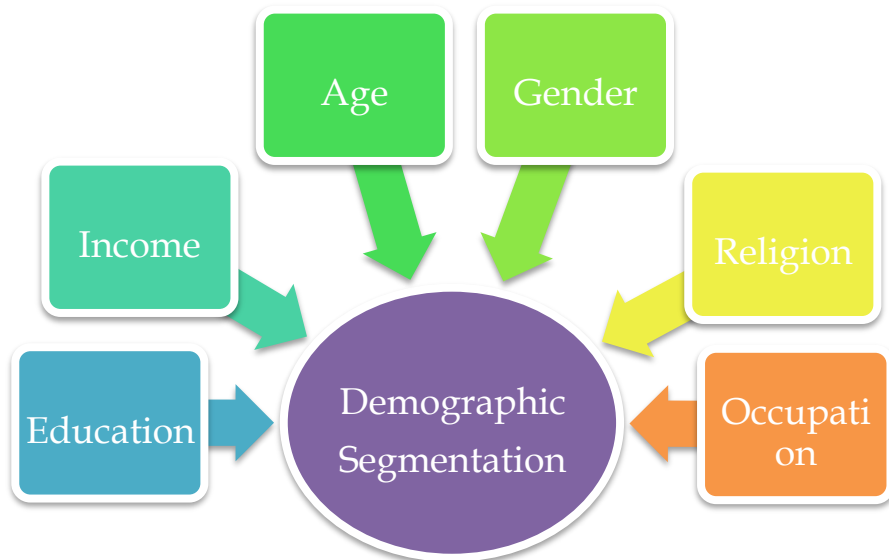


Fig: Demographic Segmentation

1.3.3 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.

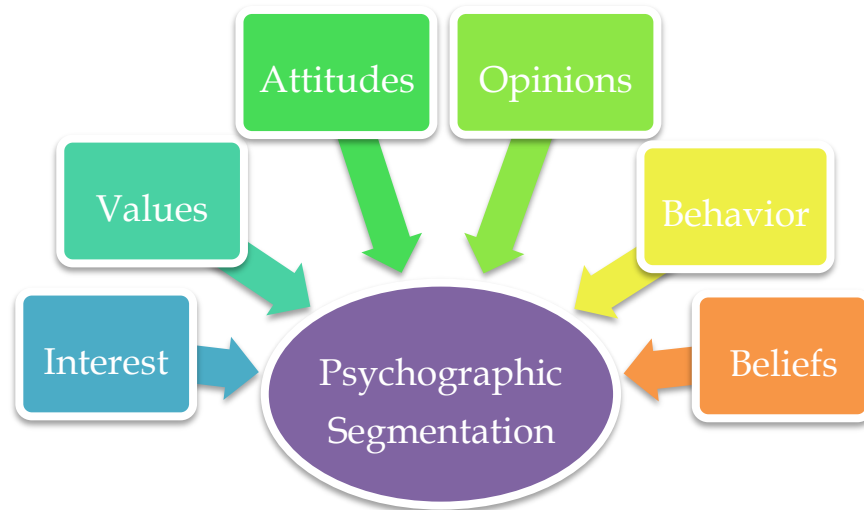


Fig: Psychographic Segmentation

1.3.4 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.

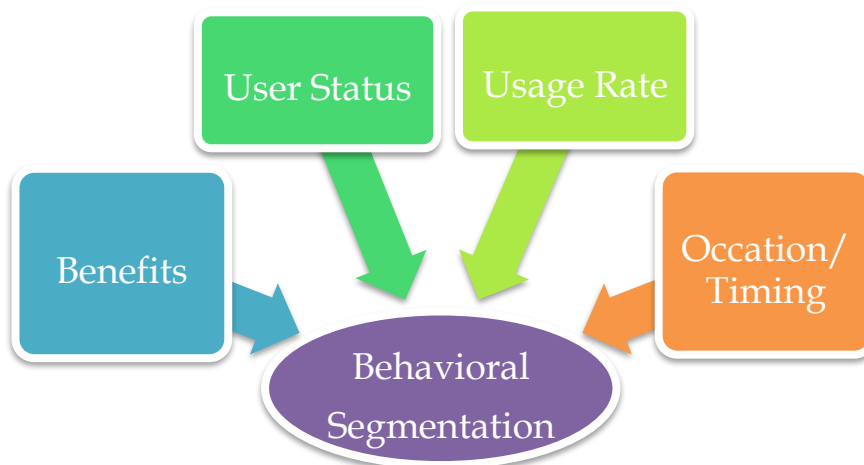


Fig: Behavioral Segmentation

2. The How

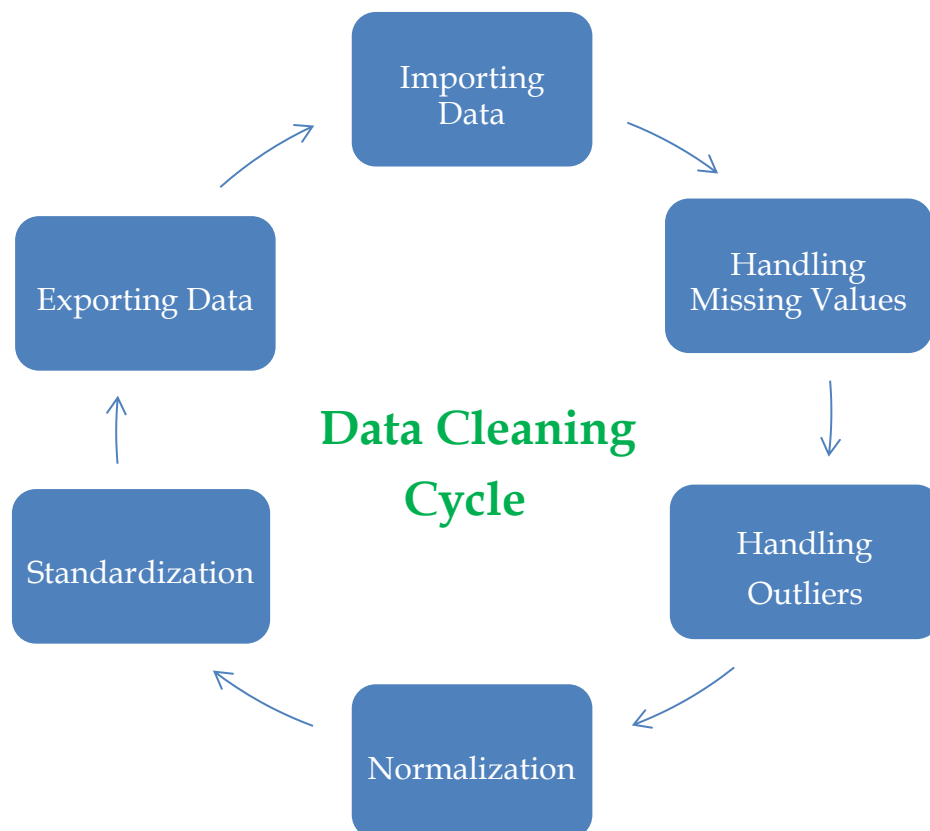
Following are the key points involved in Market Segmentation.

2.1 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.

2.2 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.



2.3 Data Preprocessing

2.3.1 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 modelling techniques. centring 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 centre data is to subtract the mean value from each data point. Subtracting the mean centres, the data around zero and sets the new mean to zero.

2.3.2 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 centre 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.

2.5 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 first variable (principal component) contains most of the variability, the second principal component contains the second most 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, which groups the unlabelled dataset into different clusters. K defines the number of pre-defined 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.

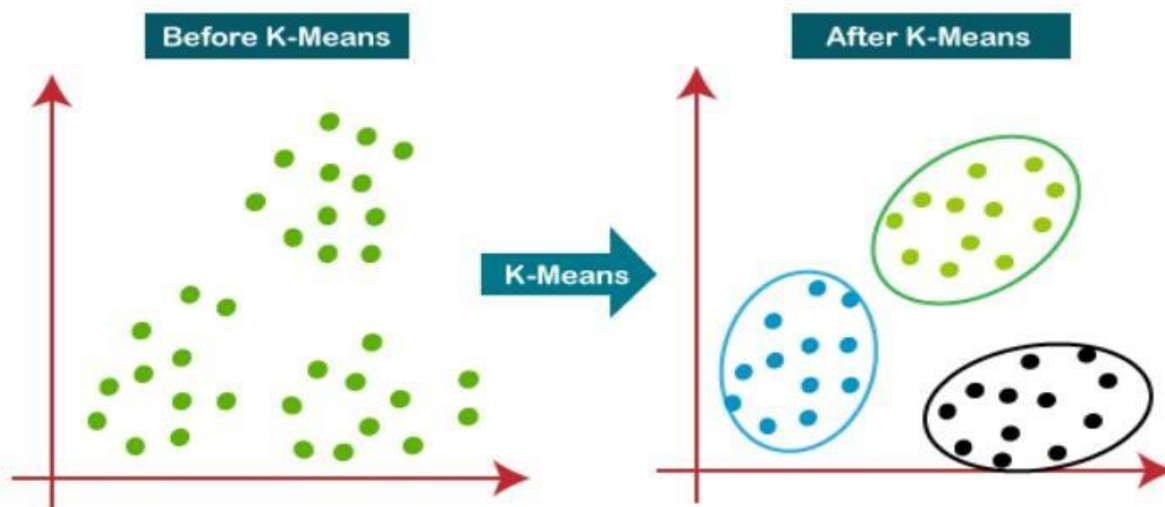


Fig: K-Means Clustering

2.6.2 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 centres 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 step 6.
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.

2.6.3 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.

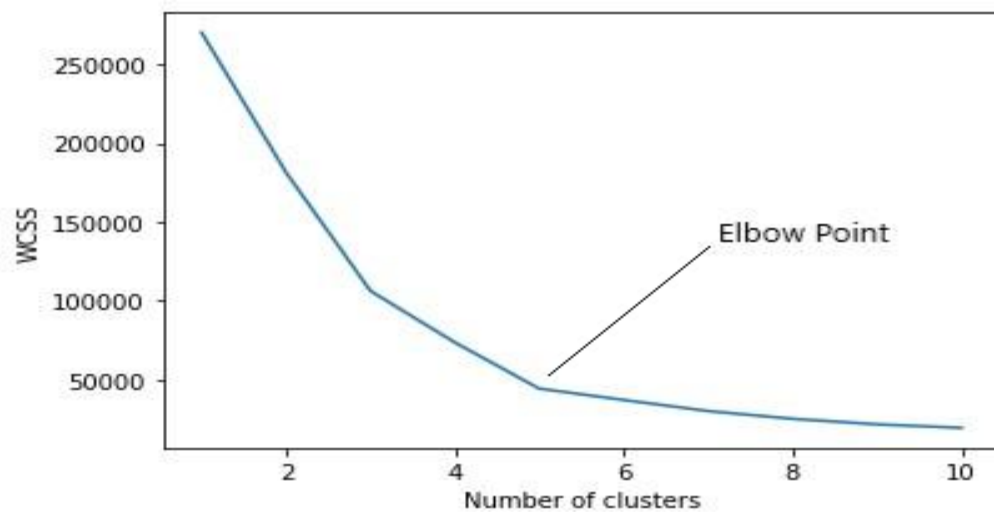


Fig: The Elbow Method

2.6.4 Why to 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 minimum sum of variance of each cluster is achieved.

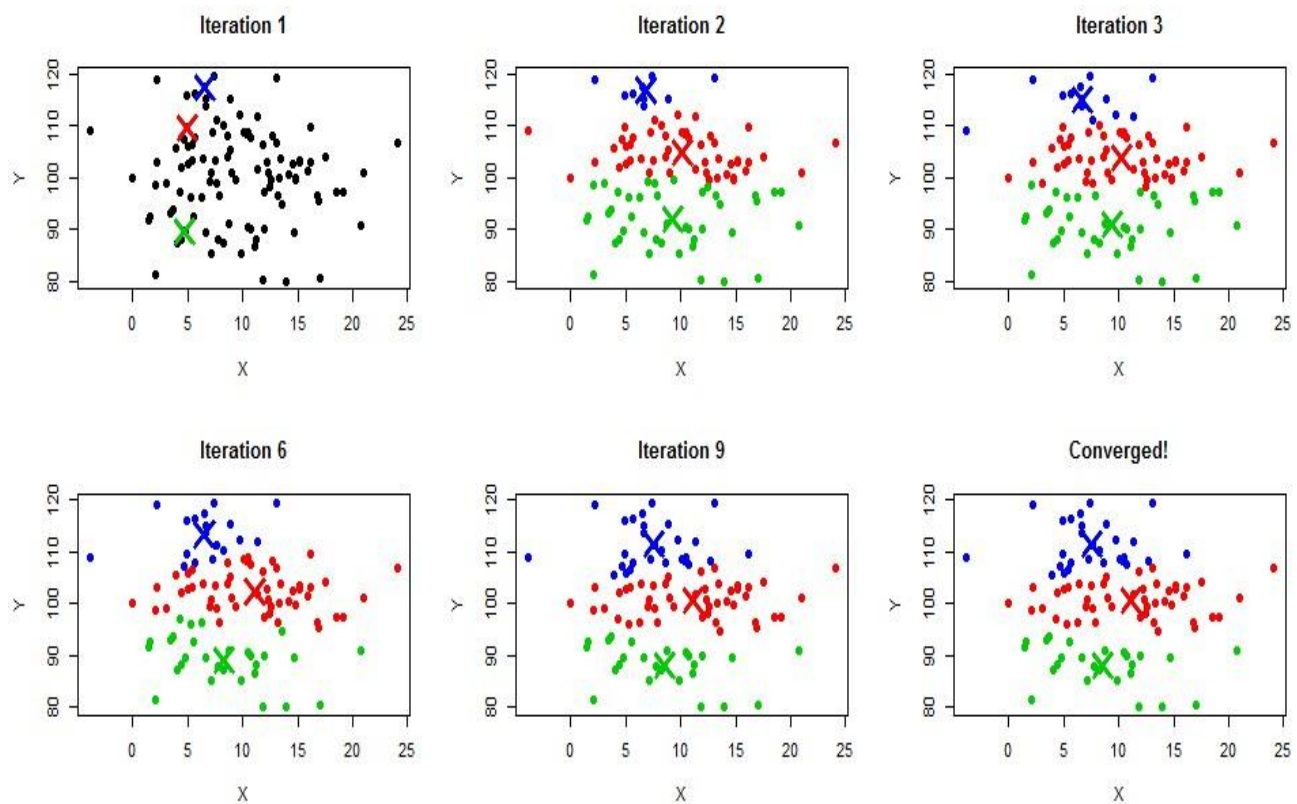


Fig: K-Means Clustering in action

3. Market Segmentation Case Study on McDonalds Dataset

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

<https://github.com/Abhishekdasari19/Feynn-Labs/blob/main/Market%20Segmentation%20Mcdonalds.ipynb>

4. Conclusion

It was noted in this study that market segmentation is regarded essential by marketing practitioners for various reasons, including targeting, proposition development, price formulation and developing of mass communication. Though being conceptualised as simple in its rationale, the process of segmentation is not necessarily easy and it requires various considerations that should be taken into account. From the literature it is evident that many marketers are expressing concern about implementation and the integration of segmentation into marketing strategy. To address this, priorities in the area of future segmentation research include the selection and incorporation of new variables into segmentation models, as well as developing new and innovative segmentation strategies.

Using market segmentation, companies are able to identify their target audiences and personalize marketing campaigns more effectively. This is why market segmentation is key to staying competitive. It allows you to understand your customers, anticipate their needs, and seize growth opportunities. This powerful technique allows you to improve your decision-making, marketing efforts, and improve your company's bottom line.

The key to successful market segmentation remains data quality; therefore, you need to pick your data provider after doing your due diligence, ensuring that you have access to the latest industry information in accessible and easy-to-understand formats.