Market Segmentation Case Study Report

By:

Darshan Sunil Shewale

Data science intern

Date: 09-Nov-2023



**Index**

1. Abstract

2. Market Segmentation

* + Introduction to market segmentation and its importance.
* 2.1 What is Market Segmentation?
  + Explanation of the concept of market segmentation.
* 2.2 Why is it important?
  + Discussion of the significance and benefits of market segmentation.
* 2.3 Types of Market Segmentation
  + Overview of various segmentation approaches.

3.How to do it

* + Explanation of the practical steps involved in market segmentation.
* 3.1 Data Exploration
  + Explanation of the initial steps in data analysis.
* 3.2 Data Cleaning
  + Discussion of data cleaning procedures.
* 3.3 Data Preprocessing
  + Explanation of data preprocessing steps.
* 3.4 Descriptive Analysis
  + Exploration of data through summary statistics and visualization.
* 3.5 Principal Component Analysis (PCA)
  + Introduction to PCA and its role in segmentation.
* 3.6 K-Means Clustering
  + Introduction to K-Means clustering and its application.
* 3.7 Elbow Plot
  + Explanation of the Elbow Method for determining the number of clusters.
* 3.8 Interpreting the Results
  + Discussion of how to interpret segmentation results.
* 3.9 Segment Profiling
  + Explanation of profiling customer segments.
* 3.10 Marketing Strategies
  + Discussion of tailoring marketing strategies based on segmentation.

4.why to use PCA with K-means for market segmentation.

5.Conclusion

6.References

Abstract

Market segmentation is a critical strategy for businesses to understand and target their customer base effectively. This report explores the process of market segmentation using various techniques, including data preprocessing, Principal Component Analysis (PCA), and K-Means clustering. We apply these techniques to a real-world dataset from McDonald's to demonstrate their practical application in identifying and understanding customer segments.

# 1. Market Segmentation

1.1 What is Market Segmentation?

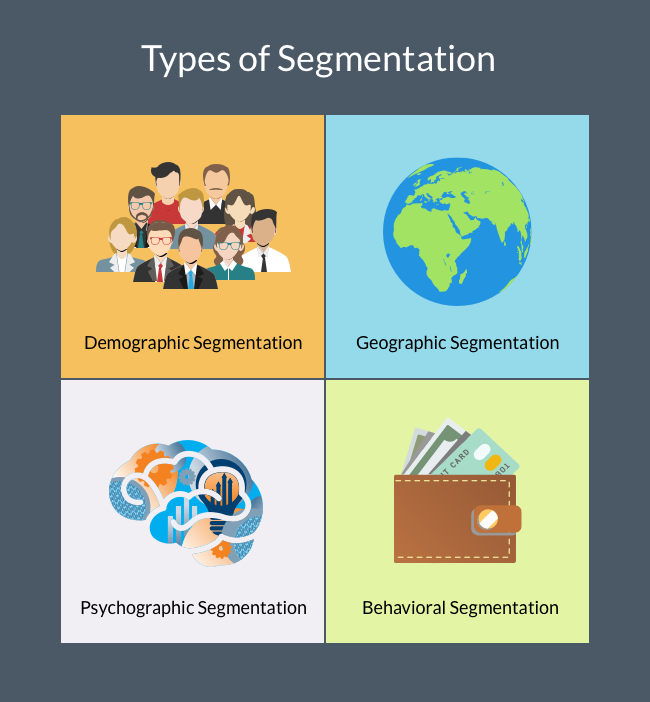
Market segmentation is a strategic approach that involves dividing a diverse market into smaller, homogeneous segments based on specific characteristics or criteria. These segments can then be targeted with tailored marketing strategies, products, and services.



1.2 Why is it important?

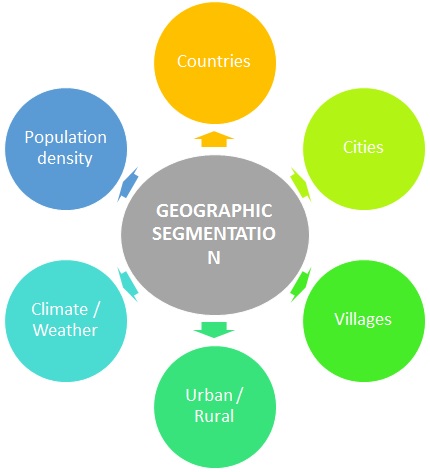
Market segmentation is essential for several reasons:

* **Improved Marketing Effectiveness:** Targeted marketing efforts are more efficient and cost-effective.
* **Enhanced Customer Satisfaction:** Tailoring products and services to specific segments increases customer satisfaction.
* **Competitive Advantage:** Understanding unique customer needs gives a competitive edge.
* **Maximized Profitability:** Focusing on high-value segments boosts profitability.

1.3 Types of Market Segmentation

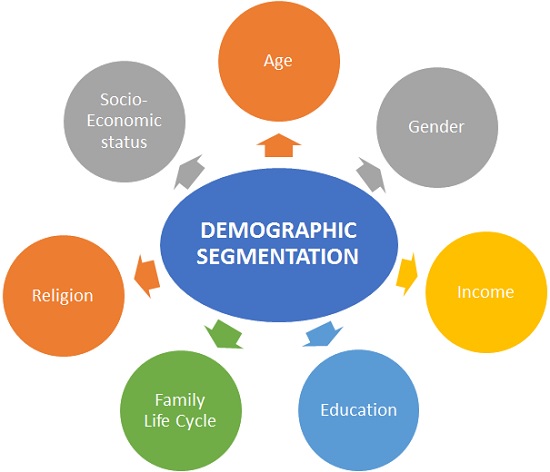
Market segmentation can take various forms:

* **Geographic Segmentation:** Divides the market based on geographical factors like location, region, climate, or population density.



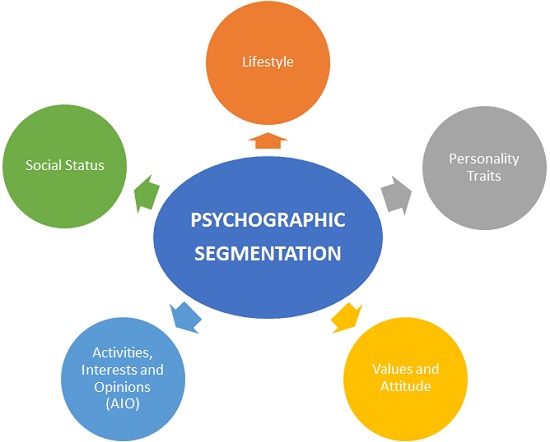
1. Location: Businesses can segment their market by specific locations, such as countries, regions, cities, or even neighborhoods.
2. Climate: Climate plays a crucial role in consumer behavior. For example, clothing retailers might tailor their products differently for customers in hot and cold climates.
3. Population Density: Urban, suburban, and rural areas may have distinct consumer preferences. Urban dwellers might seek convenience, while rural consumers may prioritize different product attributes.
4. Cultural Differences: Different regions often have unique cultures, traditions, and languages that influence consumer behavior. Adapting marketing strategies to local customs is vital.
5. Local Regulations: Legal and regulatory requirements can vary by location, impacting product offerings and marketing campaigns.
6. Competitive Landscape: Competitors' strengths and weaknesses can differ by region, influencing market entry strategies.

* **Demographic Segmentation:** Segmentation based on demographic variables such as age, gender, income, education, and family size.



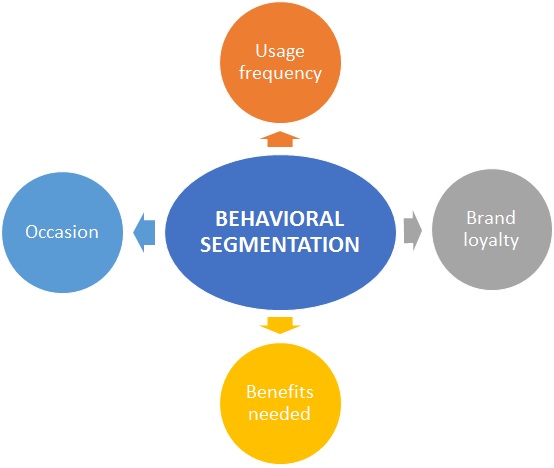
1. Age: Age groups often have distinct needs and preferences. For example, marketing strategies for teenagers differ significantly from those targeting retirees.
2. Gender: Gender-based segmentation considers whether products or services appeal more to males, females, or are gender-neutral.
3. Income: Income levels determine purchasing power. Luxury brands target higher income brackets, while value brands cater to budget-conscious consumers.
4. Education: Education levels can impact consumer sophistication, influencing their preferences for complex or technical products.
5. Family Size: Family size affects the type and quantity of products purchased. Families with children have different needs than single individuals.
6. Marital Status: Single, married, divorced, or widowed individuals may have varying lifestyle and consumption patterns.
7. Occupation: Occupation can reveal consumer interests, needs, and income levels.
8. Ethnicity and Race: Cultural backgrounds can influence preferences, such as food choices or clothing styles.

* **Psychographic Segmentation:** Groups consumers by lifestyle, values, interests, personality traits, and behaviors.



1. Lifestyle: Lifestyle segmentation looks at how individuals live their lives. Are they adventurous, health-conscious, or environmentally aware? Lifestyle choices can guide product development and marketing strategies.
2. Values: Values-based segmentation considers consumers' core beliefs and principles. Brands that align with these values can create strong emotional connections.
3. Interests and Hobbies: Understanding consumers' interests and hobbies helps in targeting products or experiences that match their passions.
4. Personality Traits: Personality traits, such as introversion/extroversion or risk-taking tendencies, can influence buying decisions and brand preferences.
5. Behaviors: Psychographic segmentation includes behavioral aspects, such as shopping habits, brand loyalty, and usage patterns.

* **Behavioral Segmentation:** Divides the market based on customer behavior, including purchase history, brand loyalty, and usage patterns.



1. Purchase History: Understanding what, when, and how often consumers purchase can help identify loyal customers and their buying patterns.
2. Brand Loyalty: Some consumers are loyal to specific brands, while others are open to trying new products. This helps in tailoring loyalty programs and incentives.
3. Usage Patterns: How consumers use a product or service can vary widely. Some may be heavy users, while others are occasional users.
4. Occasion-Based Buying: Some products are purchased on special occasions (e.g., gifts for holidays or birthdays). Identifying these occasions can guide marketing efforts.
5. Benefits Sought: Consumers may seek specific benefits from a product, such as convenience, cost savings, or status. Understanding these motivations helps in product positioning.

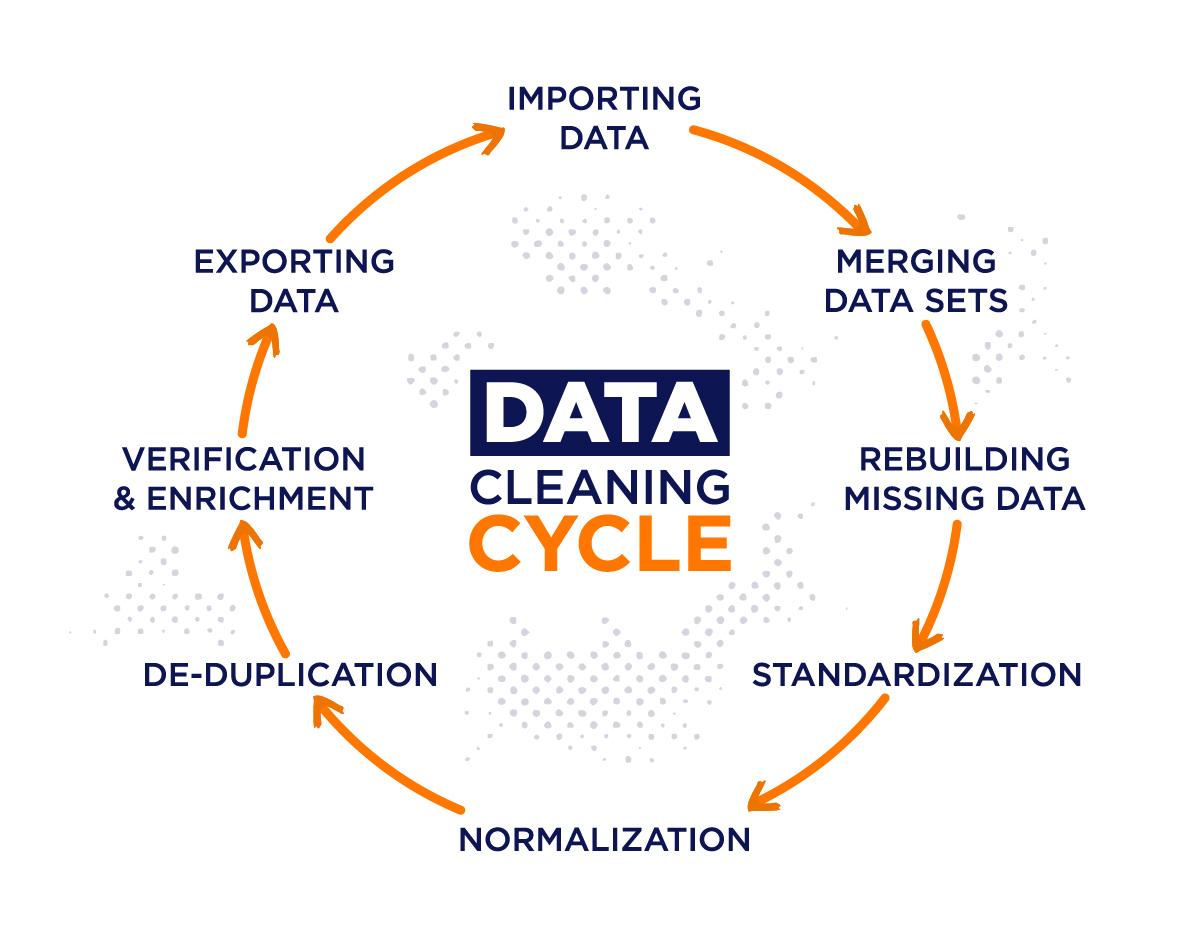
# 2. How to do it

2.1 Data Exploration

Before diving into segmentation, we explore the dataset's structure, variables, and potential data quality issues.

* Data Collection: Gather your dataset from various sources.
* Data Inspection: Check the dataset's basic properties like size and data types.
* Data Summary: Calculate statistics for numerical data and count categories for categorical data.

2.2 Data Cleaning

To ensure reliable results, we identify and address missing values, outliers, and inconsistencies in the dataset.

1. Importing Data: Retrieve data from various sources like databases, spreadsheets, or APIs.
2. Merging Data Sets: If your data comes from multiple sources, combine them into a single dataset.
3. Handling Missing Data: Address missing values by either removing, imputing, or interpolating them.
4. Standardization: Ensure consistent units and formats for data, making it easier to work with. For example, converting all dates to a uniform format.
5. Normalization: Scale numerical data to a common range, often between 0 and 1, to prevent variables with larger ranges from dominating the analysis.
6. De-duplication: Identify and remove duplicate records or entries to maintain data integrity.
7. Verification & Enrichment: Cross-check data for accuracy and completeness. You may need to validate data against external sources or enrich it with additional information.
8. Exporting Data: After cleaning, save the cleaned dataset for further analysis.

2.3 Data Preprocessing

Data Preprocessing involves several steps to prepare your data for analysis, and it can vary based on the type of data you have. Here's how you typically handle numerical and categorical data:

1. Numerical Values:

* Handling Missing Values: Check for missing numerical data and decide whether to remove or impute these values. Common techniques for imputation include mean, median, or regression-based imputation.
* Scaling/Normalization: Depending on the algorithm you plan to use, you might need to scale or normalize numerical features. Scaling ensures that all numerical features have the same influence on the model. Common methods include Min-Max scaling (scaling values between 0 and 1) or Z-score normalization (scaling with mean=0 and standard deviation=1).

1. Categorical Values:

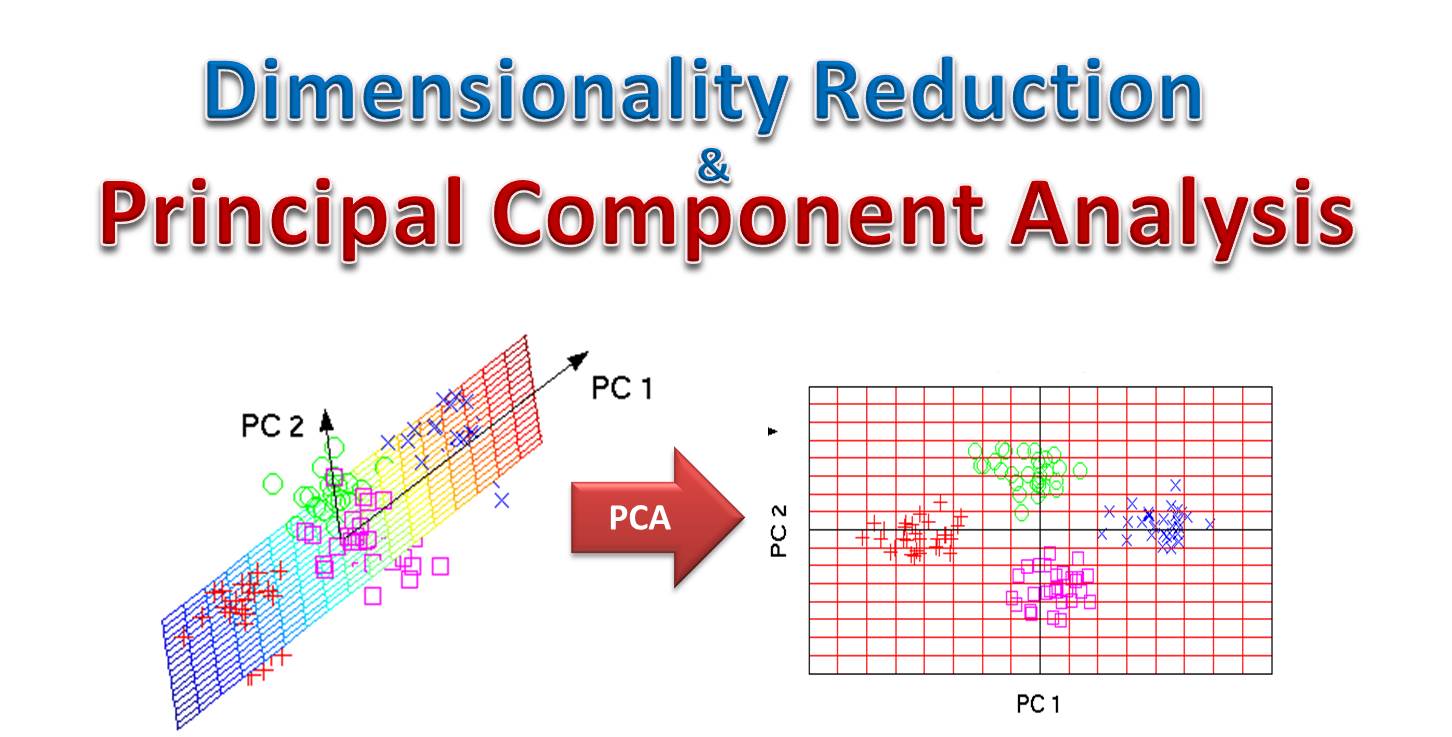
* Handling Missing Values: Similar to numerical data, deal with missing categorical values by removing or imputing them. For categorical data, imputation might involve using the mode (most frequent category) or a special category for missing values.
* Encoding Categorical Data: Machine learning models require numerical input, so you need to encode categorical data into numerical format. Common encoding methods include:
  + One-Hot Encoding: Creates binary columns for each category, indicating its presence (1) or absence (0).
  + Label Encoding: Assigns a unique numerical label to each category. Be cautious with this method, as it can imply ordinal relationships that might not exist.
* Feature Engineering: Sometimes, you can create new features from categorical data. For example, you might extract information from date columns (day of the week, month, etc.) or create interaction features between two or more categorical variables.

2.4 Descriptive Analysis

Descriptive Analysis is a crucial step in data analysis and market segmentation. Here's what it involves:

1. Data Summary: Begin by summarizing key statistics and characteristics of your dataset. This includes measures like the mean, median, mode, range, variance, and standard deviation for numerical variables. For categorical variables, summarize the frequency distribution of categories.
2. Data Visualization: Create visual representations of your data. Histograms, box plots, scatter plots, and bar charts are valuable tools for understanding the distribution and relationships within your data. Visualization helps identify patterns, outliers, and potential segmentation criteria.
3. Segmentation Exploration: At this stage, you might not have identified segments yet, but you can explore your data for potential patterns. For example, if you're analyzing customer data, you can visualize customer clusters based on certain variables.
4. Correlation Analysis: Examine the correlations between variables. Correlation matrices and scatterplots can reveal relationships between numerical variables. Strong correlations might suggest variables that are important for segmentation.
5. Segmentation Variables: Identify which variables are suitable for segmentation. Descriptive analysis can help you pinpoint the most relevant features or variables for segmenting your market effectively.

2.5 Principal Component Analysis (PCA)

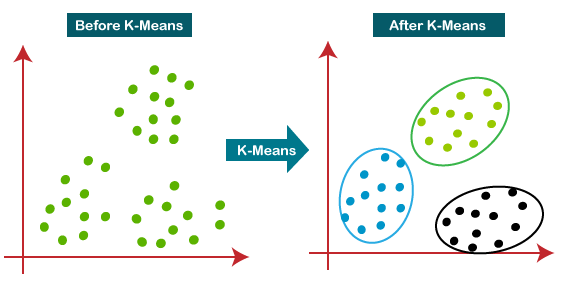


PCA reduces dimensionality while preserving information, helping to identify key variables influencing segmentation.

**Dimensionality Reduction with PCA:**

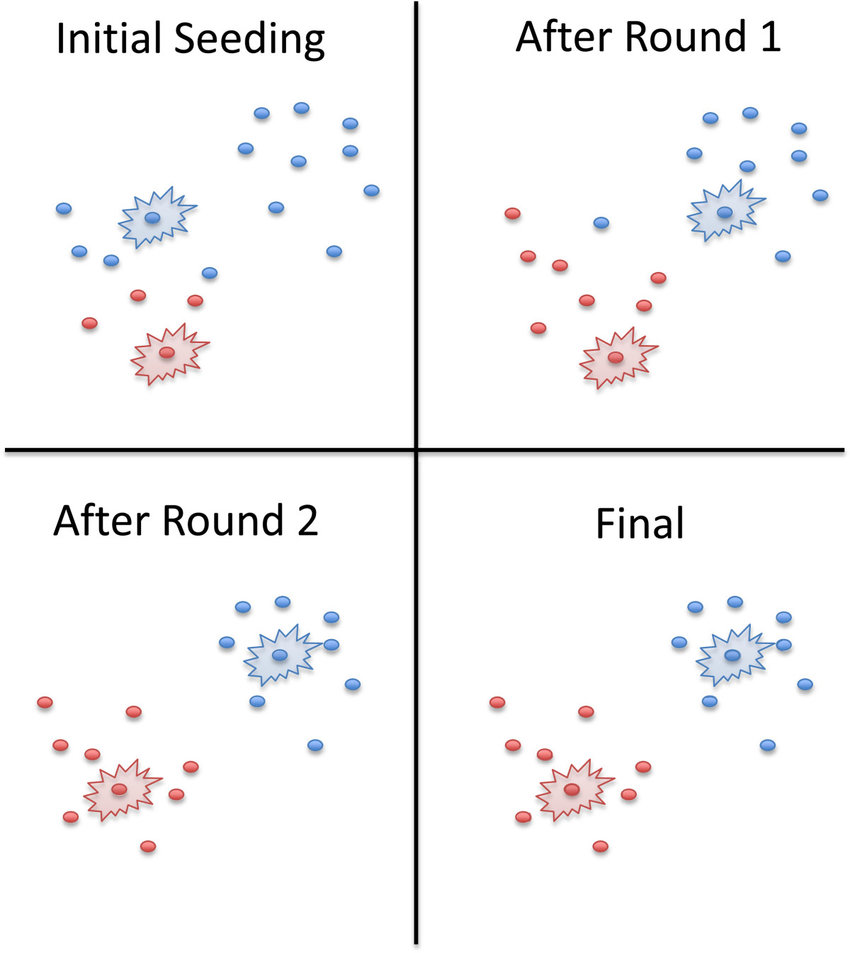
* **Initial Data:** Start with your original market dataset, which may include numerous variables (features) related to customer demographics, behavior, and preferences.
* **Standardization:** Before applying PCA, it's essential to standardize your data. Standardization scales each feature to have a mean of 0 and a standard deviation of 1. This ensures that variables with different scales contribute equally to the PCA.
* **PCA Transformation:** Apply PCA to your standardized dataset. PCA will generate principal components, which are linear combinations of the original variables. These components are ordered by the amount of variance they explain, with the first component explaining the most variance, the second component explaining the second most, and so on.
* **Choosing the Number of Components:** You'll need to decide how many principal components to keep. A common approach is to retain enough components to explain a certain percentage of the total variance in the data. For instance, you might choose to keep components that collectively explain 95% or 99% of the variance.

2.6 K-Means Clustering



**K-Means Clustering:**

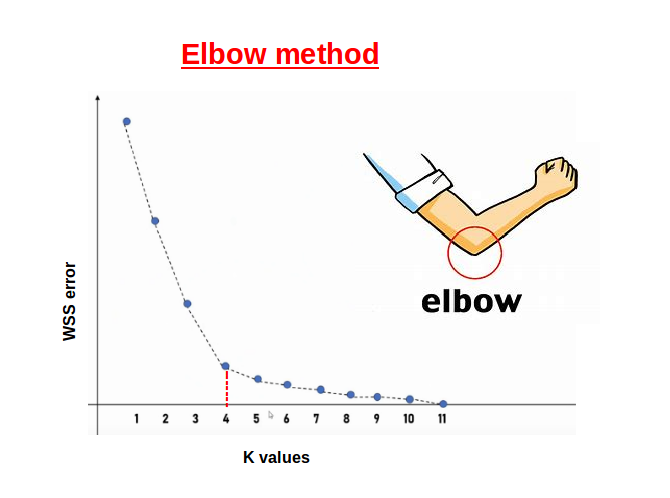
* **Reduced-Dimension Data:** Once you've selected the desired number of principal components, your data is now represented in a lower-dimensional space.
* **K-Means Algorithm:** Apply the K-Means clustering algorithm to the reduced-dimension dataset. K-Means aims to group data points into clusters based on their similarity. It does this by iteratively assigning data points to the nearest cluster centroid and updating the centroids until convergence.



**K-means clustering with action**

* **Choosing the Number of Clusters (K):** Determine the optimal number of clusters (K) for your dataset. This can be done using techniques like the Elbow Method, Silhouette Score, or Gap Statistics.
* **Cluster Interpretation:** Analyze the resulting clusters to understand the characteristics of each segment. You can look at the mean values of the original variables within each cluster to interpret what distinguishes one segment from another.

2.7 Elbow Plot



The Elbow Method is a visual tool used to determine the optimal number of clusters in K-Means clustering:

**1. Plotting WCSS:** Calculate the WCSS for a range of 'k' values (e.g., 1 to 10) by running the K-Means algorithm for each 'k.' WCSS is the sum of squared distances between data points and their assigned cluster centroids.

**2. Identifying the Elbow:** Plot the WCSS values against the number of clusters ('k'). The graph typically exhibits a downward trend. Look for the "elbow point" on the plot, which represents a point where the rate of decrease in WCSS slows down significantly.

**3. Choosing 'k':** The 'k' value corresponding to the elbow point is considered the optimal number of clusters. However, the choice of 'k' should also align with the specific objectives of the segmentation analysis.

2.8 Interpreting the Results:

* + **Variable Loadings:** PCA provides information about how each original variable contributes to each principal component (variable loadings). You can use these loadings to interpret which customer attributes are essential in defining each cluster.
  + **Visualizations:** Visualize the clusters in the reduced-dimensional space to gain insights into their spatial distribution. You can create scatter plots or 2D/3D projections of the data based on the principal components.

2.9 Segment Profiling:

* + **Customer Segmentation:** The resulting clusters represent different customer segments based on their shared characteristics. You can profile these segments in terms of demographics, behaviors, and preferences.

2.10 Marketing Strategies:

* + **Tailored Marketing:** Tailor marketing strategies for each customer segment based on their unique characteristics and needs.
  + **Personalization:** Use the insights from segmentation to personalize product recommendations, advertising, and communication for better customer engagement.

# 3. why to use PCA with K-means for market segmentation

1. **Dimensionality Reduction:** Market segmentation datasets can be high-dimensional, containing numerous variables or features related to customer demographics, behavior, and preferences. High dimensionality can make clustering challenging and computationally expensive. PCA reduces the dimensionality by transforming the original variables into a set of uncorrelated variables called principal components. These components retain most of the variance in the data while reducing the number of dimensions, making the subsequent clustering more efficient and interpretable.
2. **Reducing Noise and Redundancy:** In real-world datasets, some variables may contain noise or be highly correlated with others. Noise can introduce instability in clustering results, while correlated variables can lead to redundancy. PCA identifies the most significant patterns and removes noise by emphasizing the variance explained by each principal component. It also provides an orthogonal set of components, reducing redundancy.
3. **Enhancing Interpretability:** PCA simplifies the dataset by expressing data variations in a lower-dimensional space. This simplification makes it easier to interpret the results of K-Means clustering. Instead of clustering in the original feature space, which might be challenging to visualize and understand, clustering in the reduced PCA space allows for clearer interpretation of segment characteristics.
4. **Focus on Essential Information:** PCA prioritizes variables that contribute the most to the dataset's variance. When applying K-Means clustering, focusing on these essential variables can lead to more meaningful and actionable segments. This can be particularly helpful when making business decisions based on the segments' characteristics.
5. **Computational Efficiency:** The reduced dimensionality achieved through PCA can significantly speed up the K-Means clustering process. K-Means calculates distances between data points and cluster centroids, which can be computationally intensive for high-dimensional data. PCA's dimensionality reduction helps in reducing the number of calculations, making the process more efficient.
6. **Overcoming Multicollinearity:** In cases where multicollinearity (high correlation between variables) exists, K-Means can struggle to differentiate between the contributions of correlated variables. PCA transforms these variables into uncorrelated components, ensuring that each component captures unique information about the data. This can lead to better cluster separation.
7. **Improved Stability:** Using PCA can enhance the stability of the clustering results. Without dimensionality reduction, small variations in the data or initial conditions can lead to different clustering outcomes. PCA helps in reducing the impact of noise and small variations, making the results more stable and consistent.

# 4. Conclusion

Based on our analysis, we draw insights into customer segments within the McDonald's dataset. We discuss how businesses can benefit from these findings by tailoring marketing strategies, products, and services to specific segments. Market segmentation is a powerful tool for enhancing customer satisfaction and maximizing profitability.

# 5. References

We acknowledge and cite the sources and references used in this report.

<https://link.springer.com/book/10.1007/978-981-10-8818-6>

<https://www.kaggle.com/code/priyashravs/mcdonalds-market-segmentation>

<https://365datascience.com/tutorials/python-tutorials/pca-k-means/>

<https://www.kaggle.com/code/rohitshirudkar/complete-customer-segmentation-pca-k-means>

<https://www.investopedia.com/terms/m/marketsegmentation.asp>