
CUSTOMER SEGMENTATION

Name

Swapnita Pandey

Table of Contents

Chapter No.	Description	Page No.
Chapter 1	Introduction	1
Chapter 2	Literature Survey	4
Chapter 2	Methodology	6
Chapter 3	Result and Discussion	9
Chapter 4	Conclusion and Future Work	11
	References	

Chapter 1

Introduction

1.1 Introduction

The most successful companies today are the ones that know their customers so well that they can anticipate their needs, data scientists and analyst play a key role in unlocking these in depth insights and segmenting the customers to better serve them.

What is Customer Segmentation?

Customer Segmentation is the practice of dividing a company customer into groups that reflect similarity among customers in each group.

The goal of segmenting customers is to decide how to relate to customers in each segment to maximize the values of each customer to the business. The process requires a thought, strategy, understanding of how to manage and group your customers and which data you will use to do this.

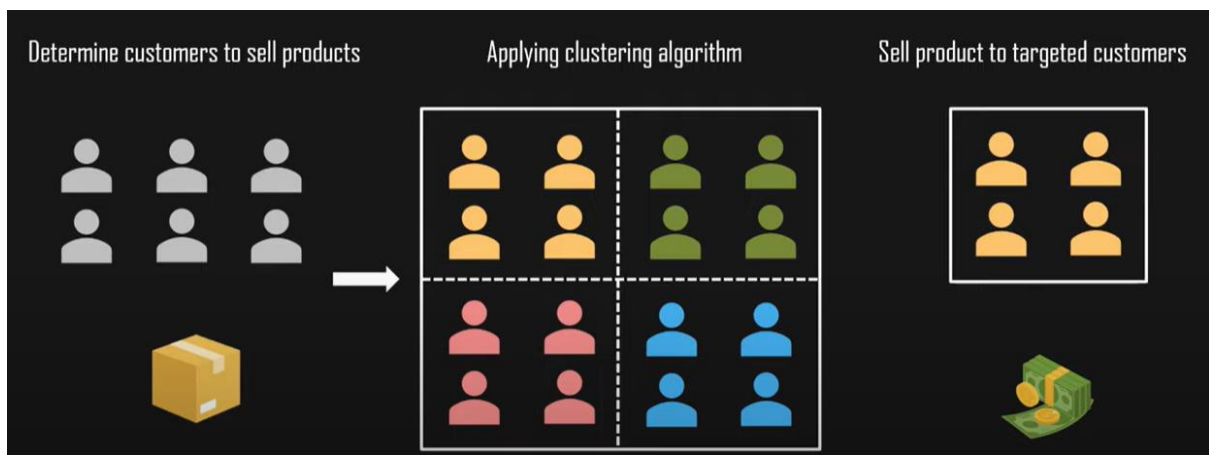


Fig 1.1An illustration of customer segmentation

Types of Customer Segmentation

1. **Demographic Segmentation:** Demographic Segmentation refers to statistical data about a group of people. It's examples include age, gender, income, location, education and so on.

E.g. Vehicle manufacturer that sells a luxury car brand. This company would likely target an audience that has higher income.
2. **Psychographic Segmentation:** Psychographic Segmentation categories audiences and segments customers by factors that relate to their personalities and characteristics. The examples include interests, lifestyles, psychological influences, motivation, priorities and so on. These are more difficult to identify and requires research.

E.g., The car luxury brand may choose to focus on customers who value quality and status.
3. **Behavioral Segmentation:** Behavioral Segmentation focuses on how the customer acts. E.g., Spending habits, purchasing habits, user status, brand interactions etc. Requires you to know about your customers actions.

E.g., The luxury car brand choosing to target customers who have purchased a high-end vehicle in the past years.
4. **Geographic Segmentation:** It categorizes customers based on geographic borders. E.g., ZIP code, city, country or climate. Can refer to a defined geographic boundary or a type of area.

E.g.: Luxury car company choosing to target customers who live in the warm climates where vehicles don't need to be equipped for snowy weather.



Fig 1.2 Examples of Customers Segmentation types

Why Machine Learning?

Machine Learning models can process customer data and discover recurring patterns across various features. In many cases machine learning algorithms can help marketing analysts find customer segments that would be very difficult to spot through intuitions and manual examination of data.

Why Unsupervised Learning?

Unsupervised learning is training of a machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here, the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Unsupervised learning can be classified into 2 categories of algorithms: Association and Clustering

An Association rule learning problem is where you want to discover rules that describe large portion of your data such as people that buy eggs and also tend to buy wine. A clustering problem is where you want to discover the inherent grouping in the data such as grouping customers by purchasing behavior.

Chapter 2

Literature Survey

Approaches to Customer Segmentation

January 2007: Journal of Relationship Marketing

Customer segmentation has virtually unlimited potential as a tool that can guide firms toward more effective ways to market products and develop new ones. As a conceptual introduction to this topic, we study how an innovative multi-national firm (Migros Turk) has developed an effective set of segmentation strategies. This illustrates how firms can construct novel and inventive approaches that provide great value. A-priori, and custom designed post-hoc methods are among the most important approaches that a firm should consider. We then review general approaches to customer segmentation, with an emphasis on the most powerful and flexible analytical approaches and statistical models. This begins with a discussion of logistic regression for supervised classification, and general types of cluster analysis, both descriptive and predictive. Predictive clustering methods include cluster regression and CHAID (Chi-squared automatic interaction detection, which is also viewed as a tree classifier). Finally, we consider general latent class models that can handle multiple dependent measures of mixed type. These models can also accommodate samples that are drawn from a pre-defined group structure (e.g., multiple observations per household). To illustrate an application of these models, we study a large data set provided by an international specialty-goods retail chain.

A practical yet meaningful approach to customer segmentation

1 October 1998: Journal of Consumer Marketing

This paper introduces the concept of the Customer Value Matrix (A matrix that divides customers into four groups (uncertain, spender, frequent, best) based on their frequency and monetary values in RFM analysis.), a customer segmentation approach that is especially well-suited for small retail and service businesses.

The discussion offers insights into the reasons for the development of this practical approach, a concrete methodology for its implementation, and strategic and tactical applications of the concept. The material is supported with strong evidence from “real-world” examples featuring a variety of small retail and service businesses. The paper concludes with a discussion of the managerial implications for companies that manage chains of small retail or service businesses as to how they can take advantage of local relationship marketing.

Chapter 3

Methodology

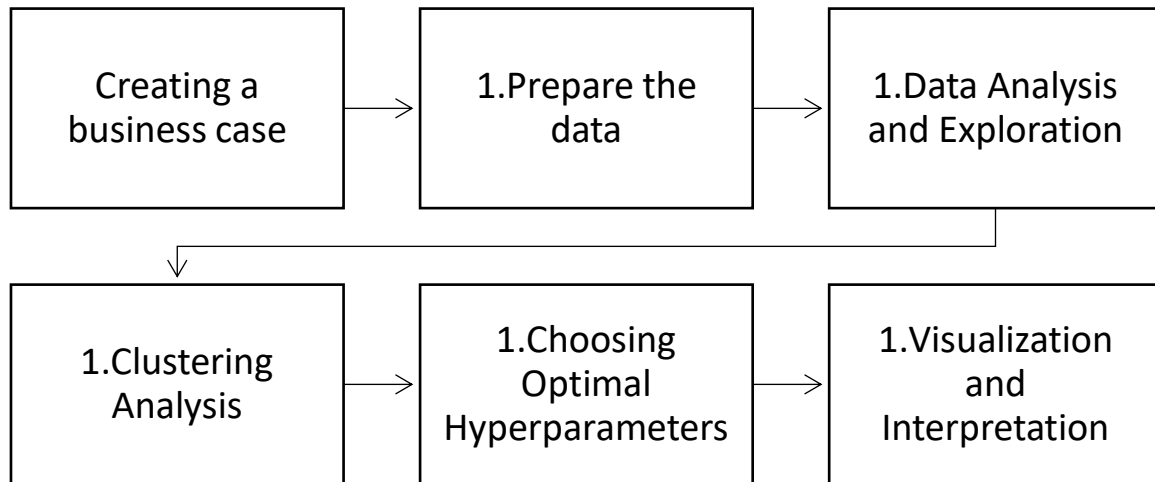


Fig 2.1 A chart representing the methodology followed

1. **Creating a business case:** Finding the most profitable customer groups within the entire pool of customers.
2. **Prepare the data:** The more the data the merrier as you will be able to find more patterns and trends within the dataset. We will also need a set of features depending on the most important matrix for the business followed by pre-processing the data to remove inconsistencies in data as they will help in better data analysis.
3. **Data Analysis and Exploration:** Finding relations and patterns in data to understand customer interest, choices and purchasing patterns so that we know which attributes are more closely related to customer and to the business as well.
4. **Clustering Analysis:** In the context of customer segmentation, clustering analysis is the use of mathematical model to discover groups of similar customers based on finding the smallest variations among customers within each group. The goal of cluster analysis is to accurately segment customers in order to achieve more effective customer marketing via personalization.

K-Means algorithm is an iterative algorithm that divides the unlabeled dataset into k predefined distinct non overlapping subgroups where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeps the clusters as different as possible.

There are 2 possible ways of choosing the number of clusters:

Elbow Method: Here you draw a curve between WCSS (Within Cluster Sum of Squares) and the number of clusters (K-Value). The curve looks like a human arm hence known as the elbow method. The elbow point gives us the optimum number of clusters. After the elbow point there are very slow changes in the value of WSS. So, the elbow point value is taken as the number of clusters. Elbow is the point of inflection on the curve, underlying model fits best at that point.

Purpose Based: We can apply K-means algo to get different clusters based on variety of purposes. You can partition the data on different matrices and see how well it performs for a particular case. You can partition the dataset into different number of clusters depending on the purpose that you want to meet.

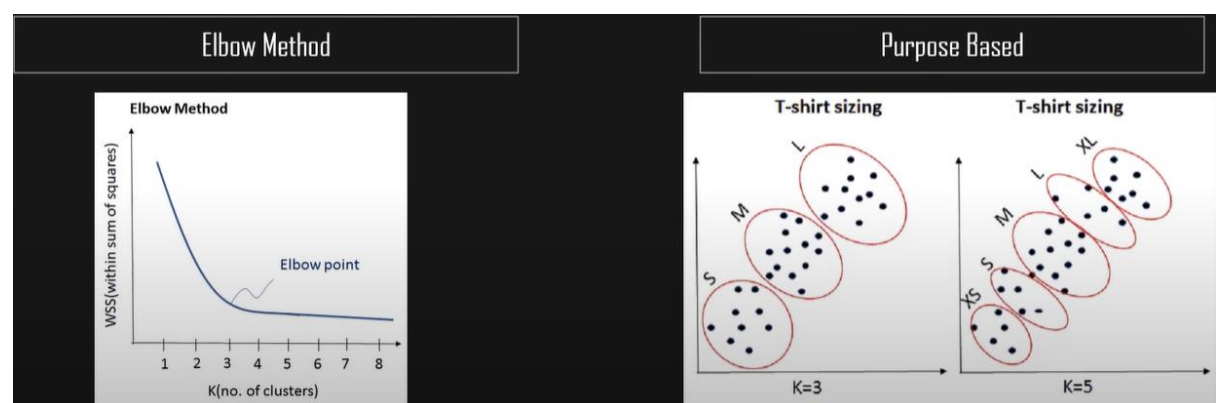


Fig 2.2 Figure representing differences between Elbow method and Purpose Based

Execution of algo

Initialization: Randomly select centroid for each cluster

Cluster Assignment: Calculating the distance between cluster centroid and data points, Now, depending upon the distance between cluster centroid and data point, the data points are grouped.

Move Centroid: Move the cluster centroid to the average of data points of a group.

Optimization: Repeat the above 2 steps till the cluster centroid stops changing it's position.

Once, the centroid becomes static the clusters are said to be **converged**.

Tools Used

Python is currently the most popular language for machine learning. Python code is easier to understand.

Pandas: Python stands for python data analysis library. Can be used to import data from various file formats such as csv files, json, sql, Microsoft excel and create a python object with rows and columns and can also be used to write data into a file. Pandas can be used for various data manipulation operations merging, reshaping, viewing, selecting, filtering, and analyzing data as well.

NumPy: Basic level external library in python. Consists of multi-dimensional array as well as matrix data structure. It can be used to perform Fourier transformation and complex mathematical operations on arrays such as statistical and algebraic operations.

Matplotlib: Visualization library. It can be used to create interactive graphs, charts and maps.

Scikit learn: Helps to make use of supervised and unsupervised algorithm just by importing the algorithms using library.

Seaborn: Python data visualization library based on matplotlib. It provides high level interface for drawing attractive and informative statistical graphs.

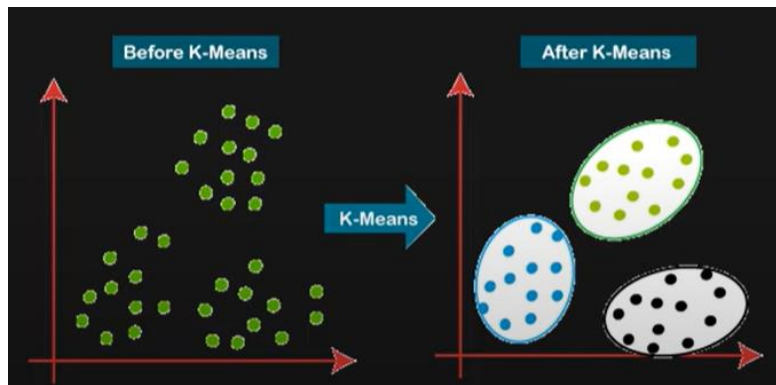


Fig 2.3 Illustration of working of K-Means algorithm

5. **Choosing Optimal Hyperparameters:** Choosing the best set of hyper-parameters for an algorithm is called hyper-parameter optimization or tuning. This is the next step because it helps us find the most accurate and rewarding customer groups based on our previous work.
6. **Visualization and Interpretation:** Visualizing the finding and interpreting them. Having profitable customer profiles at fingertips helps business improve marketing campaign, targeting future launches and product roadmaps. This gives the organization much more clear concepts about which client have the highest retention rate.

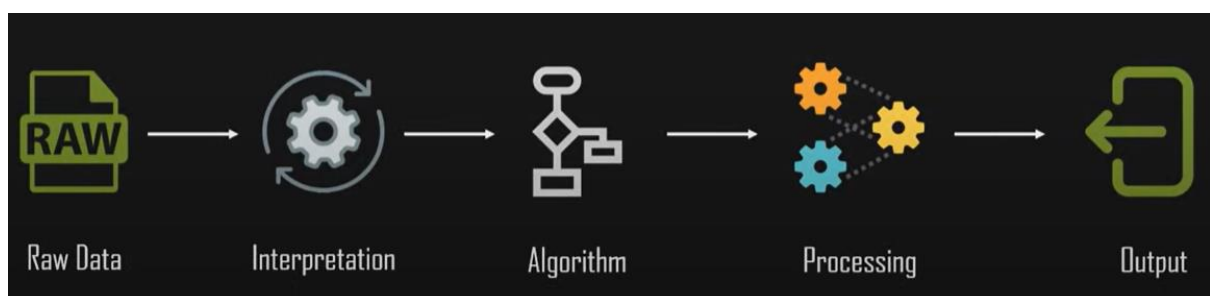


Fig 2.4 Steps followed in the project

Chapter 4

Result and Discussion

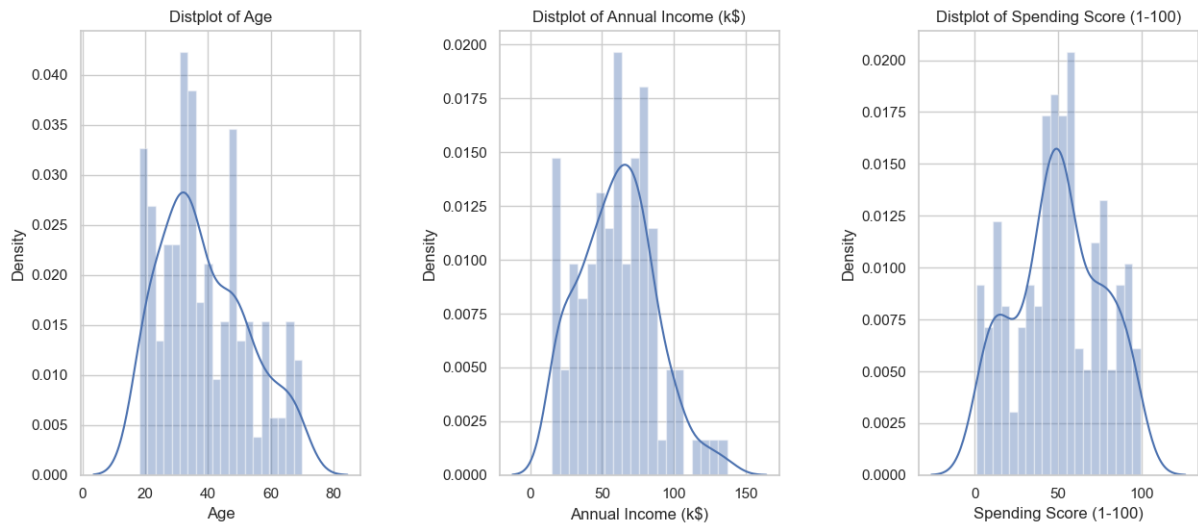


Fig 4.1 Graph representing analysis of customers according to their age, Annual Income and Spending Scores

It can be observed that the age density is highest between 20-40 years of age, annual income density is highest around 50-100, and spending score is highest around 50.

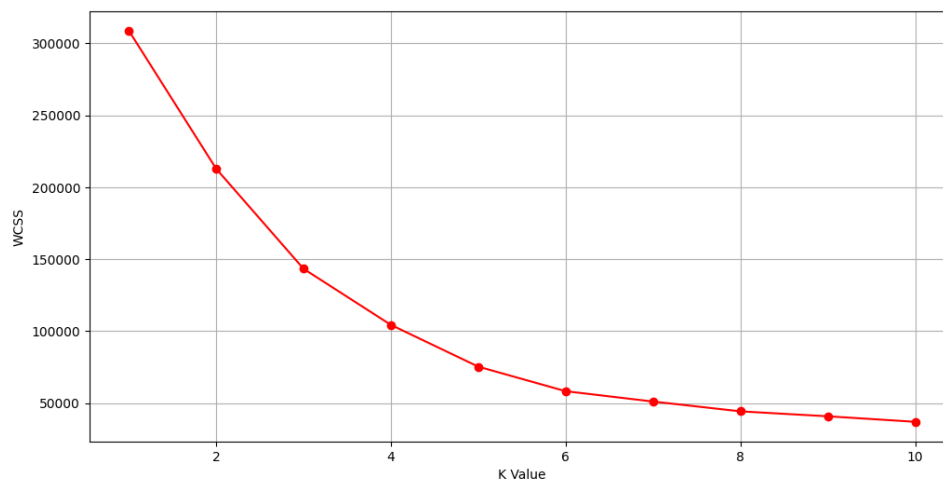


Fig 4.2 Elbow Plot

- From this Elbow plot we can observe that the point of inflection is at K-Value = 5.

- After K-Value = 5, there is a very slow change in the value of WCSS (Within Cluster Sum of Squares)
- So, the elbow point value, i.e., K-Value = 5 is taken as the Elbow point which indicates the optimal number of clusters to be formed.

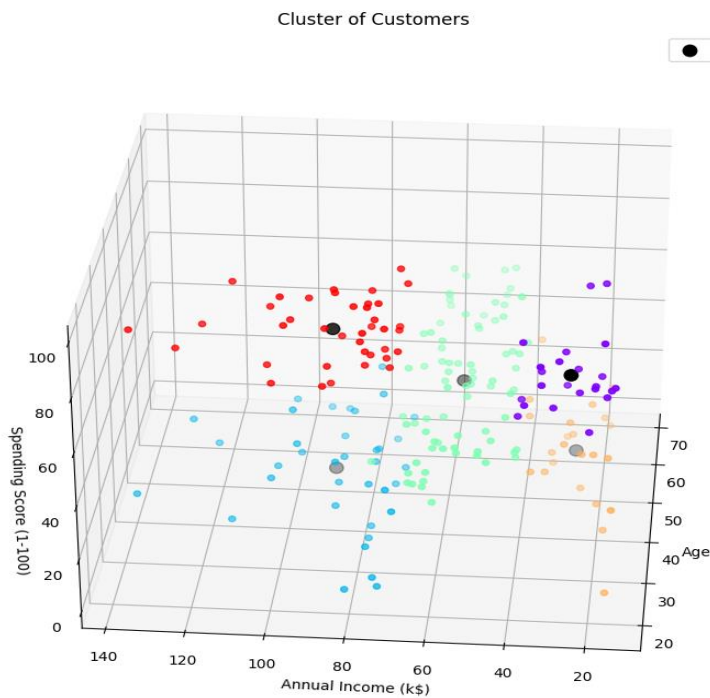


Fig 4.3 Analysis of spending score, age and Annual Income of customers

From the scatter plot, we can observe that some of the customers have less salary but more spending score. While some are less in their age but have more spending score. Such customers are not reliable for selling products.

Target audience cluster should be the one that includes maximum number of customers with decent salary, medium age group, and spending is also well. This is the most important intuitions which is extracted by a lot of companies to target their customers based on the available data. So, the center cluster includes the target audience to which the products should be sold first.

Chapter 5

Conclusion and Future Work

- K-Means algorithm can be used to divide unlabeled datasets into k predefined distinct groups.
- One of the ways of applying K-Means is Elbow method.

Advantages of Customer Segmentation

1. **Price Optimization:** Understanding the customer and financial status of your customers will assist you to pace up with the price optimization accordingly. This data helps in better allocation of resources which in return helps us to gain economy upscale.
2. **Enhances Competitiveness:** More the customer retention, more will be the revenue generated and all of this will enhance the competitiveness in the market. If you segment the market, you are well known to your customers, and you can even come up with new products or variations according to the change in preferences of your customers as the channel of distribution broadens by increasing competitiveness.
3. **Brand Awareness:** By segmenting customers, you can make them well aware of your plan. Identifying your brand will help your customers to directly engage with your products. This will increase your goodwill in the market and the brand value which is established among other competitors.
4. **Acquisition and Retention:** A personalized connection with your customers will help you to satisfy customers. About 75% of satisfied customers are more likely to retain with the optimization who regularly meet up with their needs. Better Customer Segmentation will lead to developing a better relationship with your prospective

customers. Customer Segmentation allows you to learn a great deal about your customers so you can cater to their needs effectively.

5. **Increased Revenue and Return on Interest (ROI):** By fine tuning your marketing message, you will see increase in your revenue because users will be more likely to make a purchase when they are delivered exactly what they need. In fact segmented and targeted email generates 58% of all revenue for company which isn't hard to believe. When you segment your emails you will also have a subject line that is personalized to recipients need which can increase the open rate by 26% and it's obvious that the more emails are opened the more sales you will make. Successful marketing not only requires knowledge about who your customers are but also where exactly they are in the buying process and customer segmentation based on such information can ensure your marketing campaigns are truly effective.

Due to the above advantages that Customer Segmentation provides to the user. It is a very commonly required field.

Future Work

The project is an analysis of various types of customer segmentation techniques. To make the project more user-friendly, in the future one may add a front end to this project.

References

- [1] Bruce Cooil, Lerzan Aksoy and Timothy L. Keiningham, January 2007: Journal of Relationship Marketing, Approaches to Customer Segmentation
<https://www.emerald.com/insight/content/doi/10.1108/07363769810235974/full/html>
- [2] Marcus, 1 October 1998: Journal of Consumer Marketing, A practical yet meaningful approach to Customer Segmentation
https://www.researchgate.net/publication/230557972_Approaches_to_Customer_Segmentati
on
- [3] Open Source Computer Vision (OpenCV) [Online]. <http://opencv.willowgarage.com/wiki/>