**SSM College of Engineering and Technology**

*Parihaspora, Pattan*

# **Customer Segmentation Using**

# **K-Means Clustering**

## **A PROJECT REPORT**

***Submitted by***

## **BURHAN NABI**

***in partial fulfillment for the requirement of the award***

***of internship in***

## **DATA SCIENCE**

**at**

## **Exposys Data Labs**

**DECLARATION**

This is to declare that the work in the Major Project entitled “**Customer Segmentation Using K-Means Clustering”** submitted by **Burhan Nabi** in partial fulfillment of the requirements for the Internship is a bona fide work carried out by us under the supervision and guidance of **Exposys Data Labs.** It does not contain materials copied from other published work, and the work is not published anywhere.

**Name & Signature of Student:**

Burhan Nabi

**ABSTRACT**

We live in a world where large and vast amount of data is collected daily. Analysing such data is an important need. In the modern era of innovation, where there is a large competition to be better then everyone, the business strategy needs to be according to the modern conditions. n. It deals with the issues that are already discussed by the researchers and also identifies the research gap for the further researches. It focuses on the definition, basis of market segmentation and issues related to market segmentation in detail. This research paper will provide information about the knowledge gap and will show a path for future research in the area of market segmentation, which is the heart of marketing now a day. The business done today runs on the basis of innovative ideas as there are large number of potential customers who are confounded to what to buy and what not to buy. The companies doing the business are also not able to diagnose the target potential customers. This is where the machine learning comes into picture, the various algorithms are applied to identify the hidden patterns in the data for better decision making. The concept of which customer segment to target is done using the customer segmentation process using the clustering technique. In this paper, the clustering algorithm used is K-means algorithm which is the partitioning algorithm, to segment the customers according to the similar characteristics. To determine the optimal clusters, elbow method is used.

**CONTENTS**

**DECLARATION …………………………………………………………………………………....2**

**ABSTRACT ………………………………………………………………….………………………3**

**CONTENTS ………………………………………………………………………………………….4**

**INTRODUCTION……………………………………………………………………………………5**

**EXISTING METHOD ………………………………………………………………………………6**

**PROPOSED METHOD WITH ARCHITECTURE………………...……….……………….……7**

1. **K - MEANS**

**METHODOLOGY….. ………………………………………………………….………..…………9**

**IMPLEMENTATION ……………………………………………………….…………..…………12**

**CONCLUSION …………………………..………………………………….……………...………12**

**INTRODUCTION**

Over the years, the competition amongst businesses is increased and the large historical data that is available has resulted in the widespread use of data mining techniques in extracting the meaningful and strategic information from the database of the organisation. Data mining is the process where methods are applied to extract data patterns in order to present it in the human readable format which can be used for the purpose of decision support. Today where the world is being recognized as global village marketing has become vital ingredient for every business success. It is almost become difficult to every competitor to survive in market for a prolonged period because competition is cut to throat. Change or die is the core faith of marketing. That is why development of right marketing strategy over time is required.

Right marketing Strategy is something that helps companies achieves marketing objectives. Marketing objectives help achieve corporate objectives and corporate objectives aim to achieve a competitive advantage over rival organizations. Effective marketing strategies or marketing campaigns often consist of a combination of several marketing tactics that work together in a synergistic way to establish your brand, reduce sales resistance, and create interest and desire for your product or service. Today marketing is everywhere, formally or informally, people and organization engage in vast number of activities that we call as marketing. But still there is one constraint before all companies that they cannot connect to all customers in large, broad or diverse market Every company want to focus on customers within their capacity and with customers intimacy. For this market is to divide into groups of consumers or segments with distinct needs and wants. This strategy of dividing the market in homogenous group is known as segmentation. Even companies, who have mass marketing phenomena, are now adopting this new world’s strategy i.e. segmentation.

The purpose of segmentation is the concentration of marketing energy and force on subdividing to gain a competitive advantage within the segment. It’s analogous to the military principle of concentration of force to overwhelm energy. Concentration of marketing energy is the essence of all marketing strategies and market segmentation is the conceptual tool to help in achieving this focus. Market segmentation was first put forward in the middle of 1950s by Wendell.R.Smith, an American professor of marketing. “Market segmentation is to divide a market into smaller groups of buyers with distinct needs, characteristics, or behaviors who might require separate products or marketing mixes.” (Charles W. Lamb 2003). Segmentation is the process of dividing the market into groups of customers or consumers with similar needs

**Existing Method**

Depending on the products and brand you are marketing, one method of customer segmentation might prove more effective than another. There are various customer segmentation techniques are there. We’ll analyse few of them.

1. **Segmenting Clusters based on their needs**

This model segments customers[based on their needs](http://www.targetmarketingmag.com/article/divide-conquer-a-primer-needs-based-segmentation-27857/). Of all the methods of segmentation, this one offers the marketer the most accurate way to target customer segments. It is highly scalable because the marketer can designate as many needs-based segments as preferred.

Needs-based segmentation often derives from what drives potential leads toward your business in the first place. If you are a cloud service provider and a site visitor comes to you through a blog post on file sharing, then you may determine that prospect needs a solution that simplifies file sharing. As a result, you might target that prospect with additional content surrounding that concept to help usher the user further down your sales funnel. The drawback is that the needs of customers can be difficult to define. It is also hard for marketing to express to the sales reps who will be working with each customer segment.

1. **Segmenting Clusters based on Customer Sophistication**

As the name suggests, segmenting based on customer sophistication means dividing your target audience based on their product or industry acumen. Like other methods included in this post, segmenting by customer sophistication offers the opportunity to tailor your campaigns toward a lead’s specific needs. However, instead of splitting customers by their needs, firmographic information, or potential value to your business, customer sophistication looks exclusively at a target company’s awareness of the problem your product solves.

To use the cloud service provider example again, an unsophisticated lead may be one that still keeps hard copies of every document in filing cabinets. Or, it may be a brand new startup that has not encountered the massive storage complexity associated with running a tech company yet.

1. **Segmenting Clusters based on Behavior**

For demand generation marketers, behavioural segmentation acts as a strong complement to tiered segmentation for maximizing the value of your existing customer base. Behavioural segmentation looks at the ways your current customers interact with your product or solution to determine two critical things:

* Could this customer benefit from expanding their current solution?
* Is this customer at risk of churning our solution?

In other words: is there upsell potential or do we risk losing this customer completely? In either case, a good demand generation marketer can recognize those trends and plug the customer into an appropriate segment designed to either nudge them toward buying more, or in the case of churn, mitigate risk by improving the health of their product usage.

1. **Segmenting Clusters based on Firmographics**

Marketers love firmographics because the cost to collect the data and use it for segmentation is fairly inexpensive. Additionally, firmographics are easily translatable to the sales team; marketers can convey the firmographic description of any particular customer segment to the sales department with little trouble.

On the downside, firmographics have the same disadvantages as demographics for B2C marketing. Just because a customer is a 40-year-old female does not mean she wants to buy dish detergent, and just because a company has 1,500 employees and an annual revenue of over $5 million does not mean they need a cloud service provider. The conclusions you can draw from segments based on firmographics are limited.

It is dangerous to make assumptions as a B2B marketer. You cannot assume all companies that opened up less than three years ago need a specific type of software, or that startups are not ready to purchase large equipment. Firmographics can lead you to mistaken conclusions about your customer segments.

**5. Segmenting Clusters based on Tiering**

Customer tiering is a method of segmentation based on[how well the customer matches the goals of your business](http://www.wired.com/insights/2013/04/tiered-customer-support-is-dead-and-why-thats-good-for-business/). For instance, you can use customer tiering to segment customers based on how much revenue you can expect them to bring to your business during the duration of your relationship, or by how closely that customer matches your own sales and marketing strategies.

This is a forward-thinking approach to segmentation because it ranks the importance of a customer or lead based on how much that customer can potentially bring in terms of value. Many businesses have taken tiered segmentation to a whole new level in the last few years in the form of [account-based marketing](https://www.leadspace.com/10-tips-abm-digital-advertising/), a strategy that focuses sales and marketing activities on a limited number of accounts believed to yield the highest potential value for your business. Rather than leveraging the power of big data and marketing automation to scale campaigns across a broad range of potential leads, account-based marketing turns the sights of both the sales and marketing teams toward a common goal of maximizing the potential return from a shortlist of accounts.

**PROPOSED METHOD WITH ARCHITECTURE**

Clustering algorithms try to find natural clusters in data, the various aspects of how the algorithms to cluster data can be tuned and modified. Clustering is based on the principle that items within the same cluster must be similar to each other. The data is grouped in such a way that related elements are close to each other. In this project we’ll use K – Means clustering algorithm to segment the cluster.

**K-Means Clustering**

K-Means clustering is an unsupervised machine learning algorithm that divides the given data into the given number of clusters. Here, the “K” is the given number of predefined clusters, that need to be created.

It is a centroid based algorithm in which each cluster is associated with a centroid. The main idea is to reduce the distance between the data points and their respective cluster centroid.

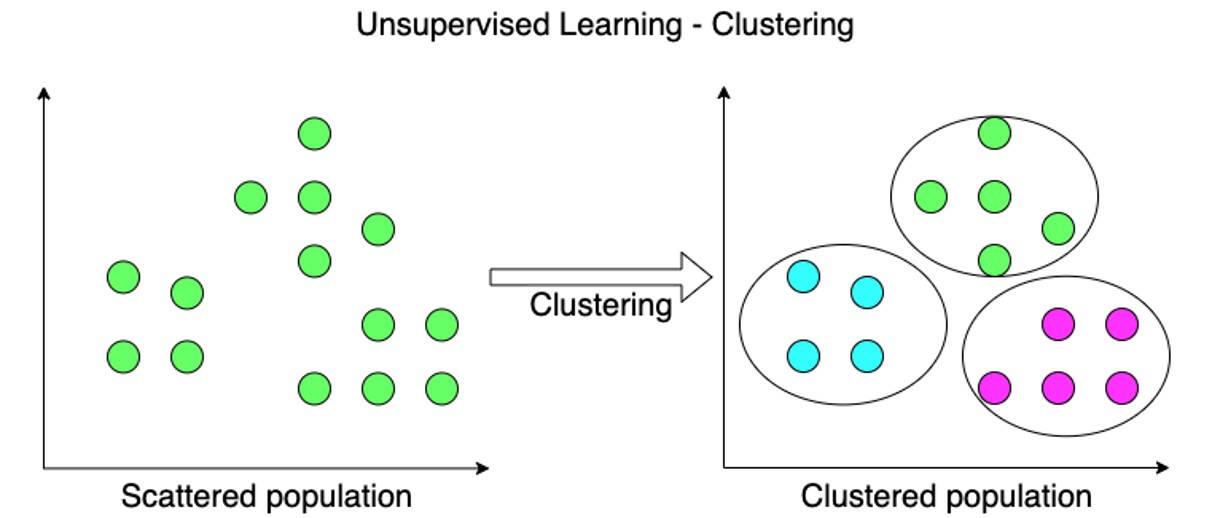
The algorithm takes raw unlabelled data as an input and divides the dataset into clusters and the process is repeated until the best clusters are found.

K-Means is very easy and simple to implement. It is highly scalable, can be applied to both small and large datasets. There is, however, a problem with choosing the number of clusters or K. Also, with the increase in dimensions, stability decreases. But, overall K Means is a simple and robust algorithm that makes clustering very easy.

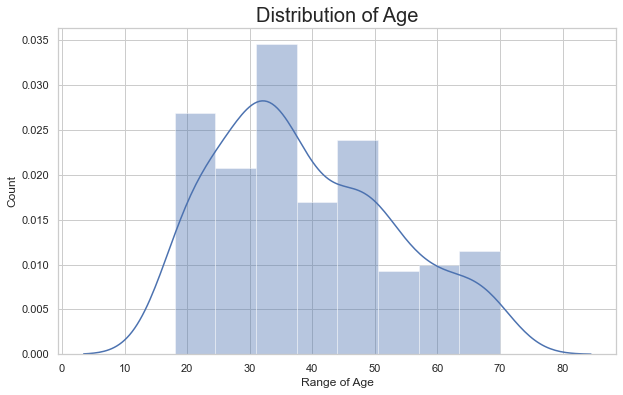
**IMPLEMENTATION**

**How K-Means works**

1. Specify number of clusters K.
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn’t changing.

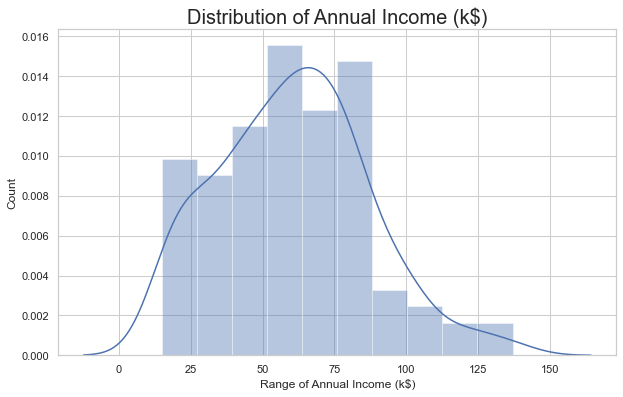
****

**Distribution of Age**

****

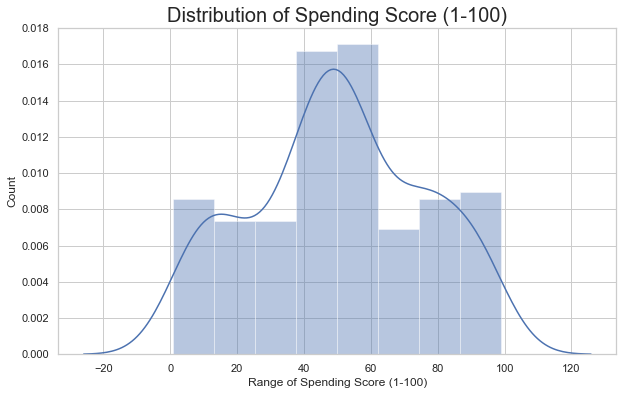
There are customers of a wide variety of ages.

**Distribution of Annual Income**

****

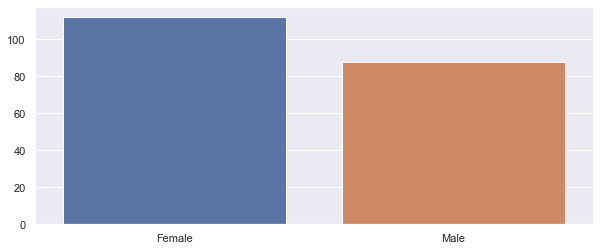
Most of the annual income falls between 50K to 85K.

**Distribution of Spending Score**

****

The maximum spending score is in the range of 40 to 60.

**Gender Analysis**

****

More female customers than male.

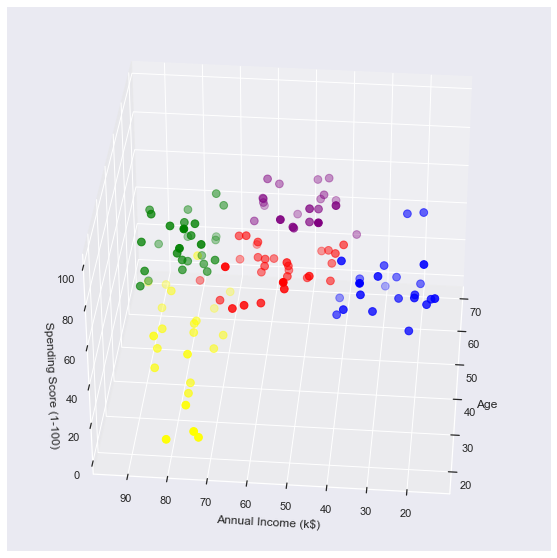
**Clustering Based on Two Features**

****

We can clearly see that 5 different clusters have been formed from the data. The **Blue** cluster is the customers with the least income and least spending score, similarly, the **Brown** cluster is the customers with the most income and most spending score.

**RESULT & OUTCOMES:**

Know we will work on 3D data. Apart from the spending score and annual income of customers, we shall also take in the age of the customers.



**METHODOLOGY**

The data set used to implement clustering and K-Means algorithm was collected from a store of shopping mall. The data set contains 5 attributes and has 200 tuples, representing the data of 200 customers. The attributes in the data set has Customer Id, gender, age, annual income(k$), spending score on the scale of (1-100).

**CONCLUSION**

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major application of K-means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.