Customer Segmentation

-by Yatharth Anand

Abstract

In this project I have implemented an unsupervised learning algorithm i.e. k means clustering for the purpose of customer segmentation.
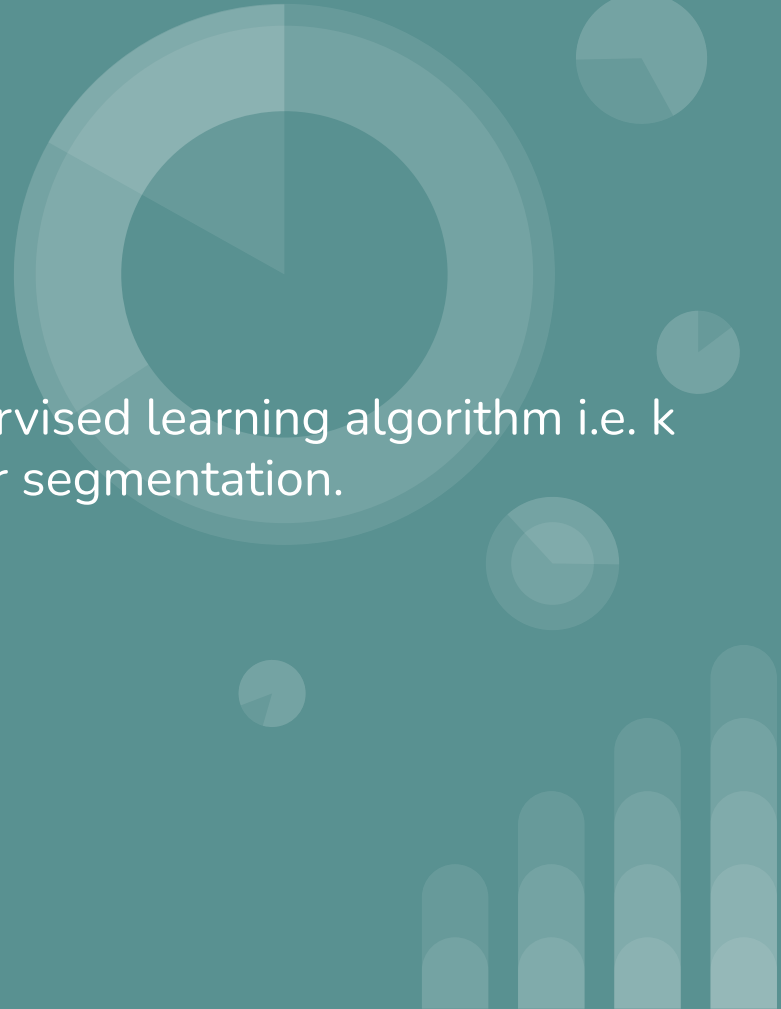
Table of Contents

Introduction
What is Customer Segmentation? Why is it used?

Customer segmentation is the process by which one can divide the customers up based on common characteristics – such as demographics or behaviors, so one can market to those customers more effectively.

Customer Segmentation allows marketers to better tailor their marketing efforts to various audience subsets.Customer segmentation requires a company to gather specific information /data about customers and analyze it to identify patterns that can be used to create segments.

Existing Methods

There are 4 basic methods of customer segmentation-

1.Demographic Segmentation -Based on things that vary with demography like age, sex, marital status, family size, occupation, education level, income, race, nationality and religion.

2.Behavioral Segmentation-Dividing your audience based on behaviors they display allows you to create messaging that caters to those behaviors.

3. Geographic Segmentation- Based on location they are from.

4.Psychographic Segmentation-Psychographic segmentation is similar to demographic segmentation, but it deals with characteristics that are more mental and emotional.

Used Method

In this project I have done a demographic customer segmentation. Given dataset consisted information about customers like gender, age, spending score and annual income.Using these data I formed different clusters of customers thus segmenting based of their demography.

Methodology
What is K-Means clustering?

A clustering algorithm called K-Means clustering was used for customer segmentation.

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of predefined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

The k-means clustering algorithm mainly performs two tasks:

1.Determines the best value for K center points or centroids by an iterative process.

2.Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.
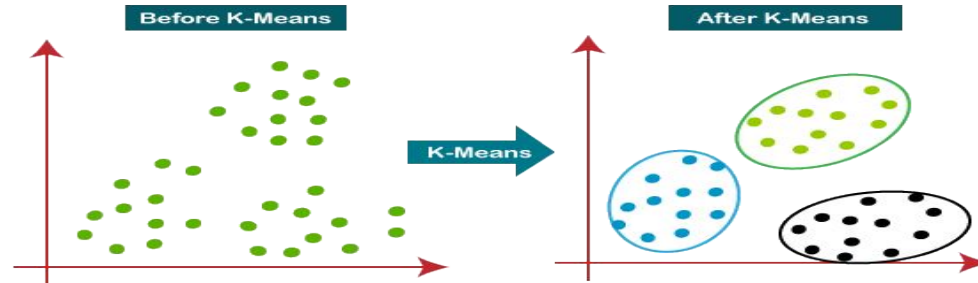
Implementation

In this project I have used Scikit-Learn Library for the implementation of K-Means clustering algorithm on the dataset.

First I used the Elbow Method to find the minimum no of clusters that is forming inside the dataset between annual income and spending scores of the customers which was found to be 5.

Then I plotted those 5 clusters on the graph to for visualization purpose of clusters
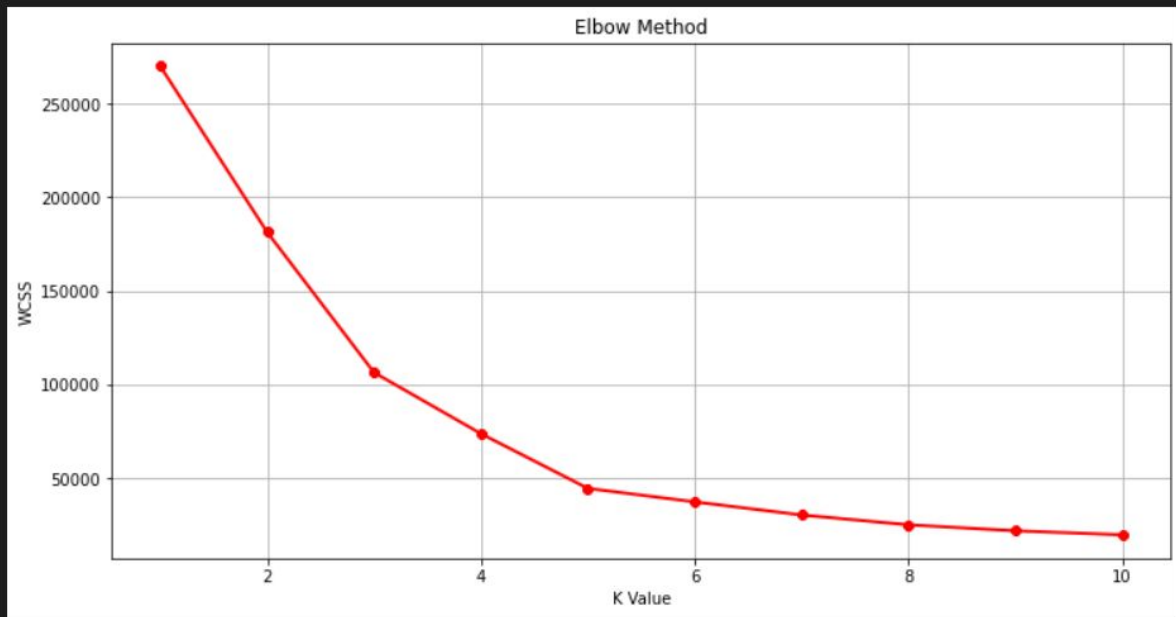
Elbow Method

Calculate the Within Cluster Sum of Squared Errors (WSS) for different values of k, and choose the k for which WSS first starts to diminish. In the plot of WSS-versus k, this is visible as an elbow.

The optimal K value is found to be 5 using the elbow method.

Plot of Age vs Annual Income vs Spending Scores
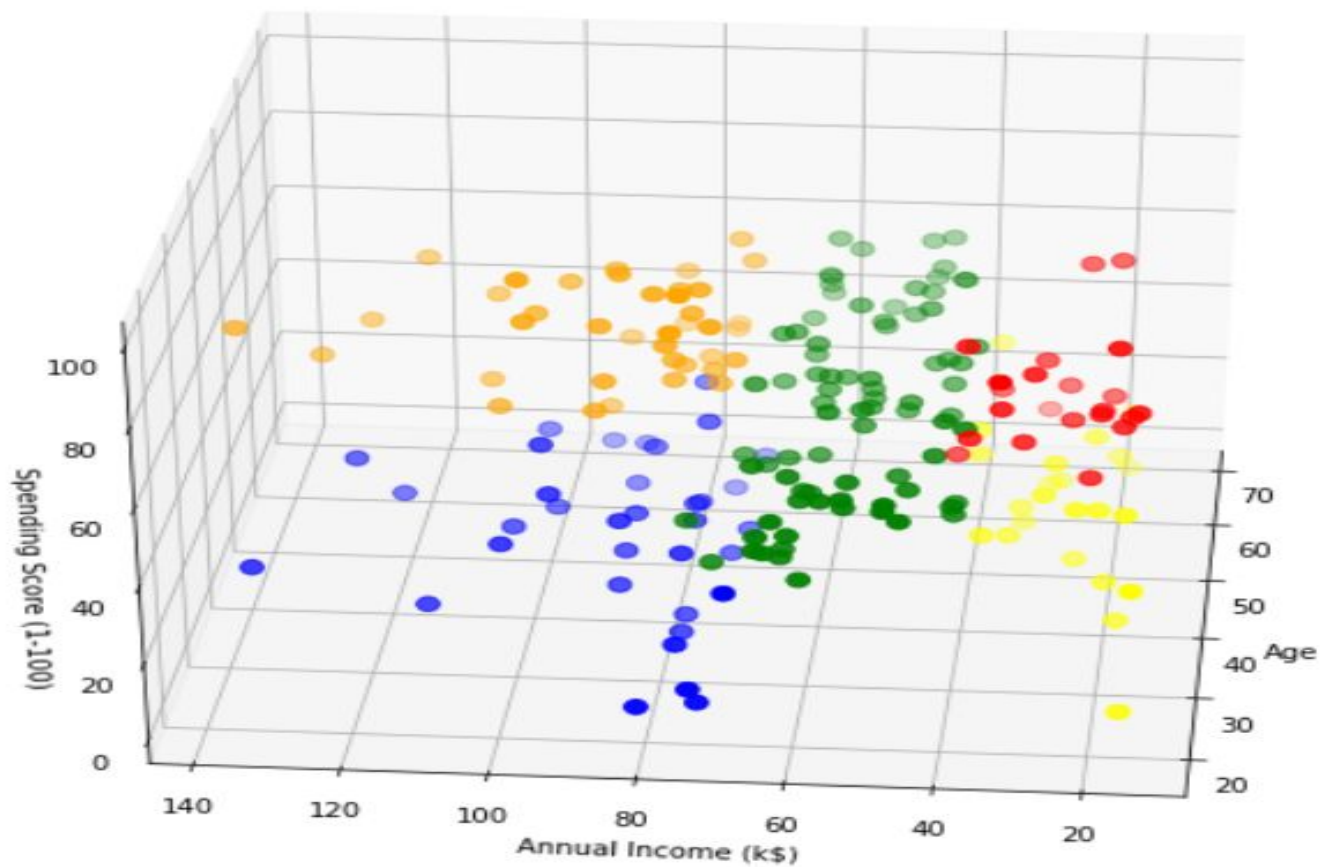
Conclusion

K means clustering is one of the most popular clustering algorithms and is great for visualizing and understanding 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.