

MINI PROJECT (2020-21)

ON

Customer SEGMENTATION

**AUTHOR:** SANA AFREEN

**SECTION:** E

**ROLL NO:** 2014831\_(48)

**SEMESTER:** 4th

**BATCH:** 2019 - 23

ABSTRACT

**Background**

For service companies, it is essential to deliver good customer service to stay competitive out in the market. Customer needs are changing over time and there are different needs for different customer groups. Companies of today are in constant need of figuring out new ways to keep their customers loyal and at the same time keep up with their competitors. Because of different needs and different type of customers it is important for companies to be aware of their different customer groups. A customer segmentation can be done to differentiate customers from each other. By making a customer segmentation, customers can be divided into different groups based on different characteristics and they can be distinguished from each other. This enable companies to easier meet the customer needs and use different marketing strategies for different customers. this report focuses on segmenting customers using machine learning methods (K-Means and Agglomerative Hierarchical Clustering) to generate newfound ways to explore consumer base.Where this study focuses on the issue of customer segmentation .

INTRODUCTION

For a retailer, understanding the components of their consumer base is key to maximizing their potential in a market; the retailer that attracts the most customers will acquire the most market share. . In fact, the high costs of gaining a new customer or getting back an old customer force

retailers to seriously consider how to allocate resources to optimize not just volume of customers, but the retention of them as well6.

it is generally in the best interest of the retailer to devote efforts to retaining customers by understanding them on as deep of a level as necessary. However, examining the intricate, rich relationships between retailer and their consumer base involves understanding how difffferent components of the base behave. Namely, how difffferent segments of customers act similarly or differently from other segments10. One method of approaching customer understanding is through the lens of customer segmentation. In short, customer segmentation analysis is the process of grouping customers in such a way that customers within one particular group are similar to each other but difffferent from customers in other groups.In order to perform customer segmentation analysis at a high level, retailers have begun to incorporate aspects of machine learning into the analysis of their customers. More specififically, retailers are utilizing unsupervised machine learning tools such as clustering . Because of this massive potential, retailers across all industries are attempting to leverage clustering algorithms such as K-Means or hierarchical clustering to more accurately and quickly segment their customers..

PROBLEM DESCRIPTION

Customers are of different character and have different needs. Therefore, it is not optimal to have same strategy and marketing for every customer. For a company, it is important to segment their customers and identify the differences between the customer segments to easier meet the customer needs and take care of those customers who are of high importance for the Company.The purpose with the project was to examine how a company can segment their customers based on their different characteristics.To understand the customers like who can be easily converge [Target Customers] so that the sense can be given to marketing team and plan the strategy accordingly

TARget

By the end of this case study , you would be able to answer below questions.

* How to achieve customer segmentation using machine learning algorithm (KMeans Clustering) in Python in simplest way.
* Who are your target customers with whom you can start marketing strategy [easy to converse]
* How the marketing strategy works in real world.
* Difference between K-means clustering and Hierarchical Clustering.

REQUIREMENTS

**LANGUAGE USED:** Python

**SOFTWARE REQUIREMENTS:** Anaconda3 -> Jupiter

**HARDWARE REQUIREMENTS:** Windows Desktop

**REQUIRED INSTALLMENTS:**

* Python
* Anaconda3
* Jupiter

**Libraries:**

* Pandas
* Numpy
* Matplotlib
* Sklearn
* Seaborn

# **APPROACH TO SOLVE THE PROGRAM**

* Create a Business case
* Preparing the data
* Data Analysis and exploration
* Clustering Analysis
* Choosing Optimal Hyperparameters
* Visualization

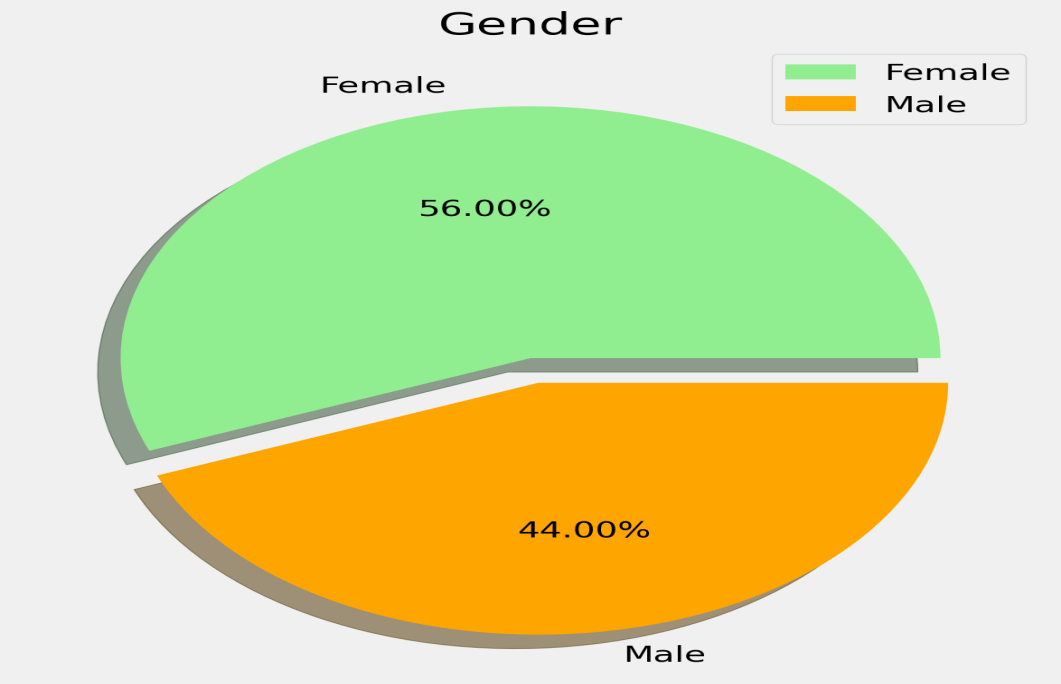
ABOUT DATA

Data contains the basic information (ID, age, gender, income, spending score) about the customers.

DATA ANALYSIS

Data is one of the important features of every organization because it helps business leaders to make decisions based on facts, statistical numbers and trends. Due to this growing scope of data, data science came into picture which is a multidisciplinary field. It uses scientific approaches, procedure, algorithms, and framework to extract the knowledge and insight from a huge amount of data.

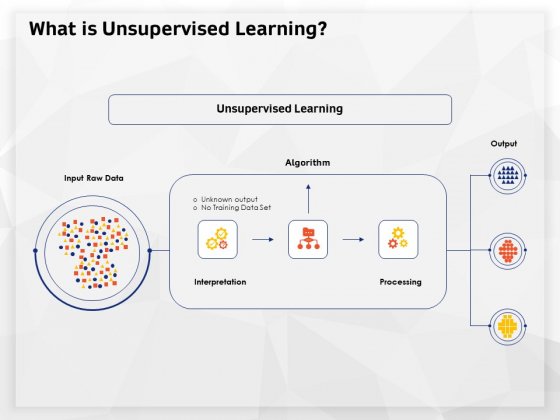
Like:

* Age countplot , Income countplot, Spending Score countplot
* Pairplot ,Scatterplot, and so on.
* 

CUSTOMER SEGMENTATION USING MACHINE LEARNING

While many applications of machine learning, such as regression and classifification, focus on predicting the outcome or value of an instance,these applications do not attempt to understand similarities between instances, just the relationship between instances and their respective outputs. Thus, when it comes to searching for algorithms or methods that look for similarities between features of instances, the focus must turn from supervised machine learning to unsupervised machine learning. Determining whether an algorithm is a part of supervised and unsupervised machine learning is contingent upon whether the instances used to train the model in the training data contain their target value. In all cases of supervised machine learning training, instances are paired with a target value, which could be a scalar or a vector depending on the context. In contrast, unsupervised machine learning deals with data that is not paired with a target value.

In technical terms, clustering is an unsupervised machine learning technique that groups instances into clusters based on the similarities between instances. This just states that clustering is one way of viewing or evaluating data by looking at the natural groupings or segments that separate instances in the data.



K-MEANS MODEL

K-means clustering algorithm computes the centroids and iterates until we it finds optimal centroid. It assumes that the number of clusters are already known. It is also called flat clustering algorithm. The number of clusters identified from data by algorithm is represented by ‘K’ in K-means.

ALGORITHM

1. Clusters the data into k groups where k is predefined.

2. Select k points at random as cluster centres.

3. Assign objects to their closest cluster centre according to the Euclidean distance function.

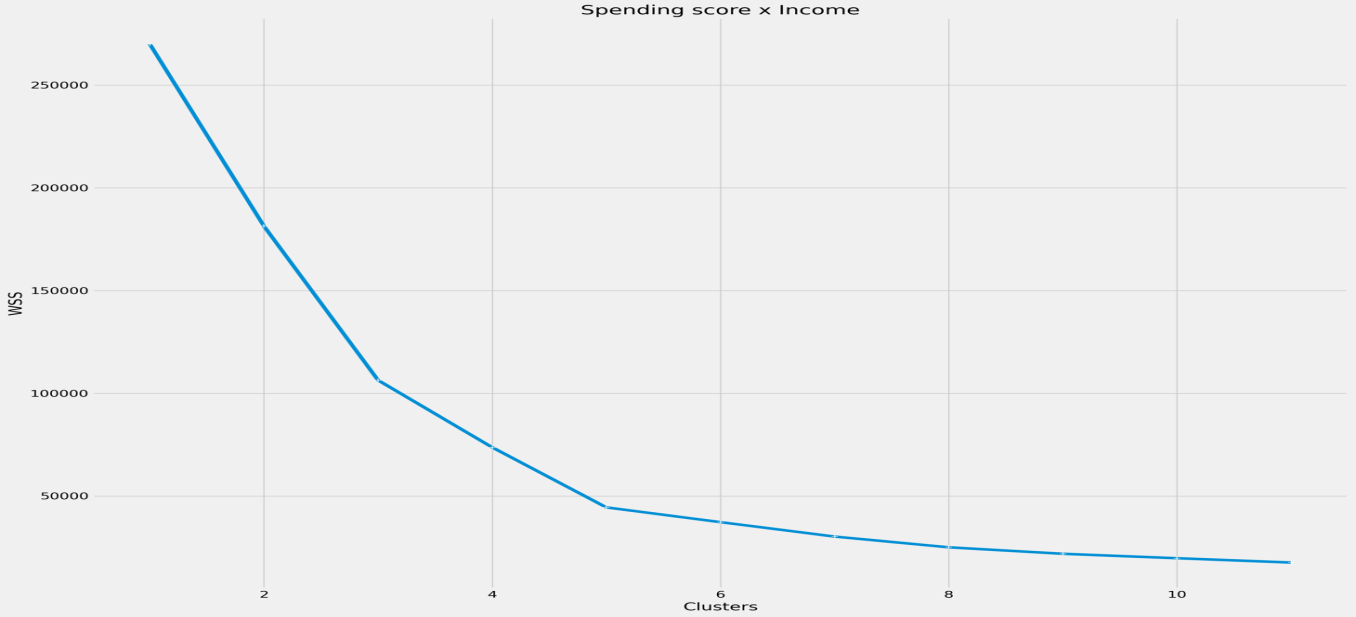
4. Calculate the centroid or mean of all objects in each cluster.

5. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

****

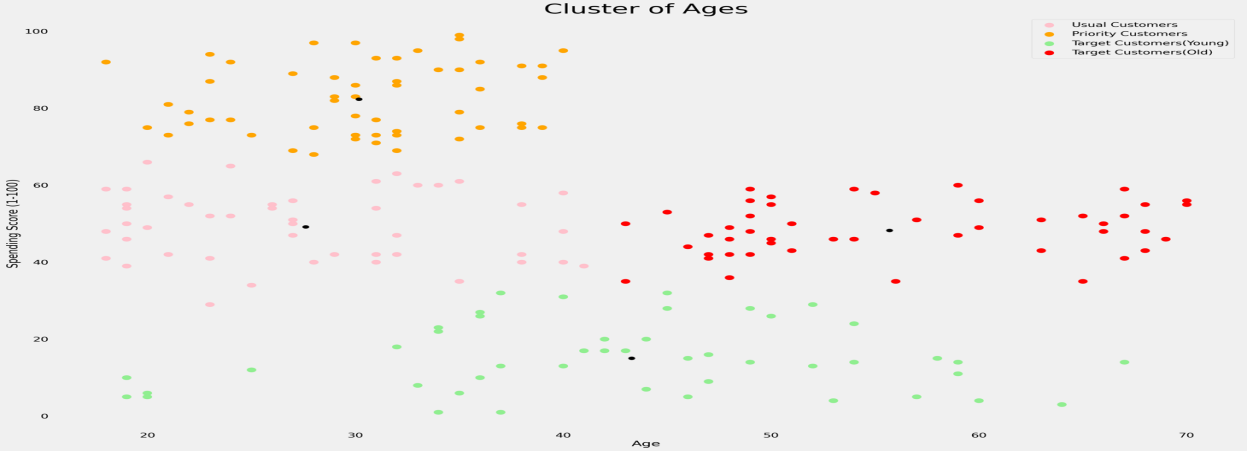
ELBOW METHOD

In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a dataset. The method consists od plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use.



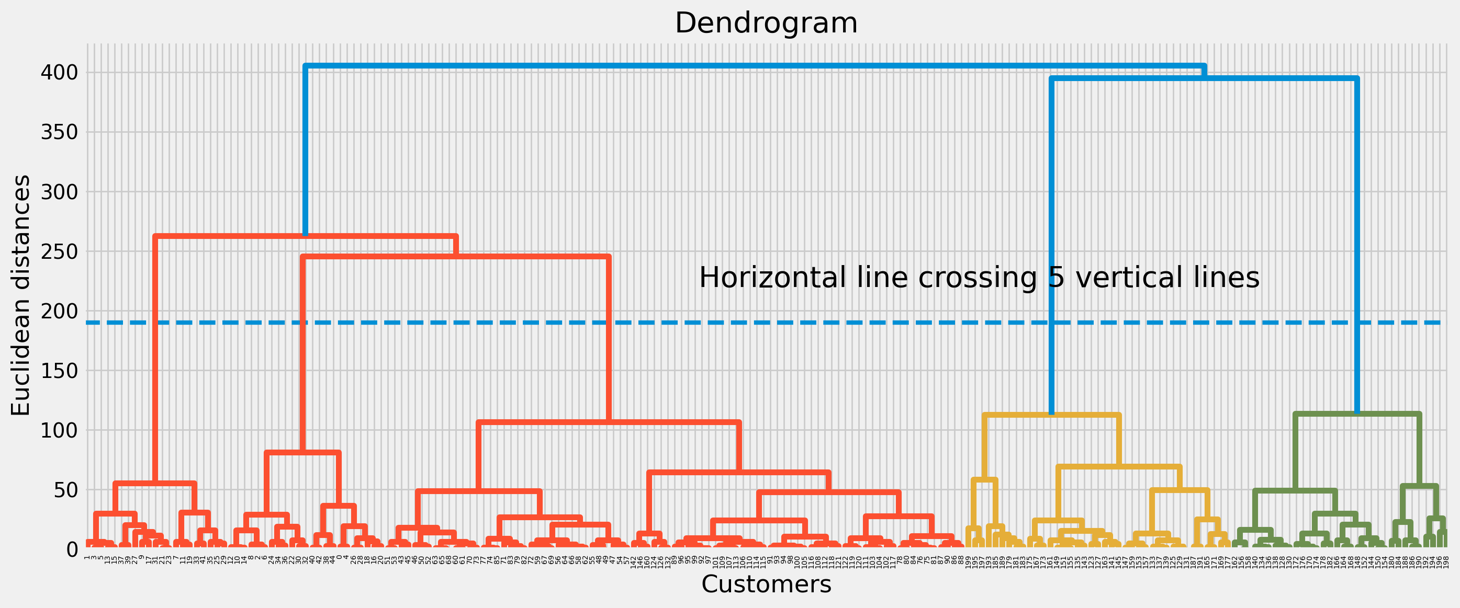
K-MEANS Cluster plot

Number of cluster that we find using elbow method is 5



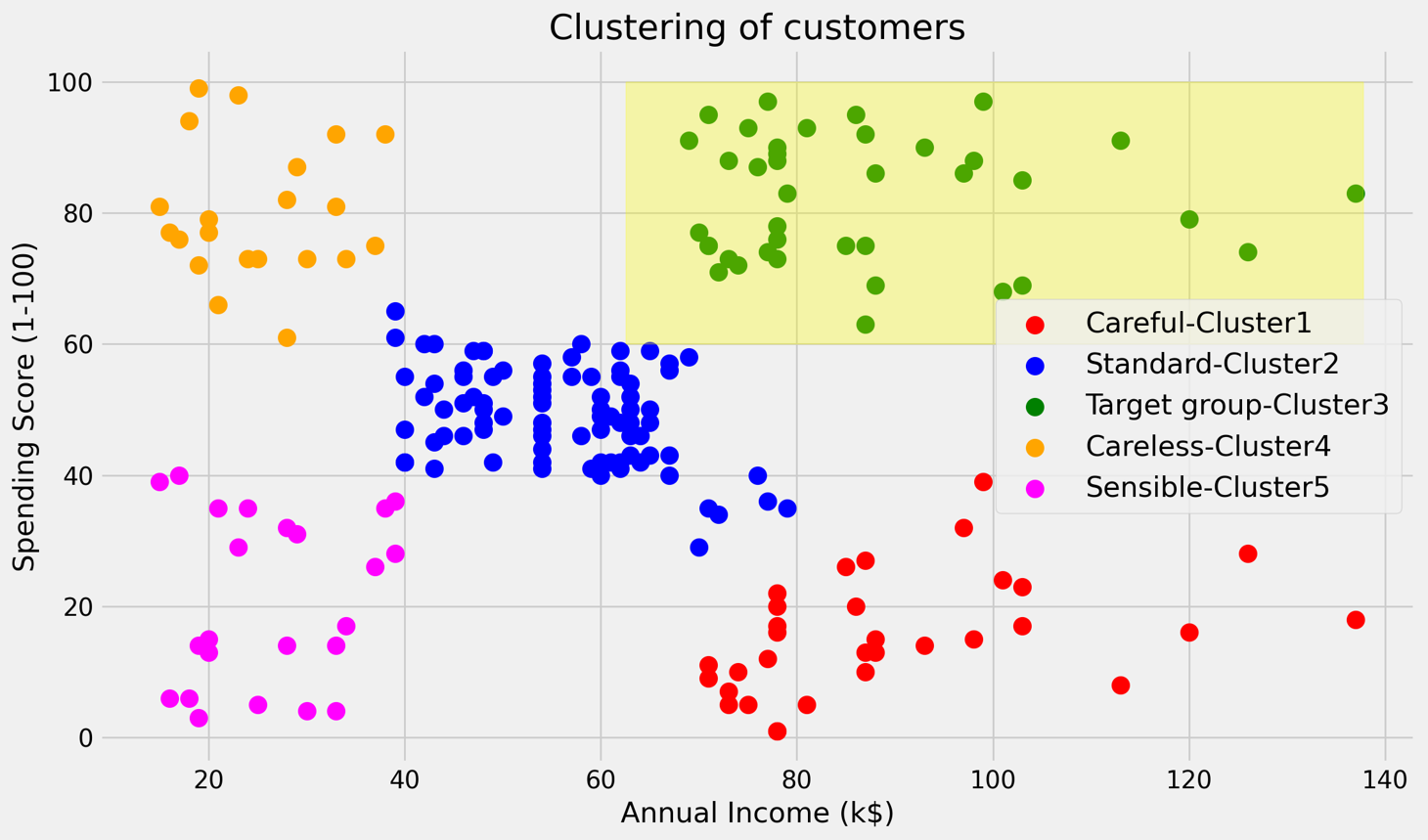
**DENDROGRAM**

It show the hierarchical relationship between objects. It is most commonly created as an output from hierarchical clustering.



Hierarchical Cluster

Hierarchical Agglomerative(bottom -up approach)



CONCLUSION

This project that we have projected is a system which will help the retailers to access the targeted customer easily. This system will help in overcoming some drawbacks that were ealier faced by the companies in graouping the customer according to their purchasing habit (factors like spending score and income of customer).

Also, retailers can track a customer activity within various groups, which Is helpul for determing who’s most receptive to outreach programs, rewards and incentives and remainders.

Link of code:[2286-SANA/Mini-Project-4sem-: Customer Segmentation (github.com)](https://github.com/2286-SANA/Mini-Project-4sem-/tree/main)

REFERENCES

Bhatnagar, Amit; Ghose, S. (2004), ‘A latent class segmentation analysis of

e-shoppers’, *Journal of Business Research* **57**, 758–767.

**URL:** *https://doi.org/10.1007/s00357-018-9259-9*