# Customer Segmentation

A

Mini Project Report

Submitted

by

| Ms.Pratiksha Pawar | 2127048 |
|---|---|
| Ms. Samiksha Pawar | 2127050 |
| Ms. Vishakha Jadhav | 2127022 |
| Ms.Harshada Waghchoude | 2127068 |

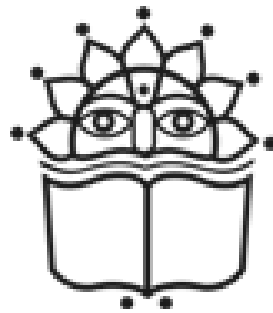In partial fulfillment for the requirement of Project Based Learning-II

of

## Bachelor of Artificial Intelligence and Data Science

Under the guidance of

### Prof.R.V.Panchal

(Assistant Professor)

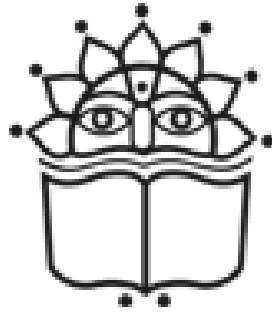# Department of Artificial Intelligence and Data Science

Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology

Bhigwan Road, Vidyanagari

Baramati-413133

2021-2022

Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

# Department of Artificial Intelligence and Data Science

# Certificate

THIS IS TO CERTIFY THAT FOLLOWING STUDENTS

| | |
|---|---|
| **Ms.Pratiksha Pawar** | **2127048** |
| **Ms.Samiksha Pawar** | **2127050** |
| **Ms.Vishakha Jadhav** | **2127022** |
| **Ms.Harshada Waghchoude** | **2127068** |

HAVE SUCCESSFULLY COMPLETED THEIR PROJECT WORK ON

## CUSTOMER SEGMENTATION

DURING THE ACADEMIC YEAR **2021-2022** IN THE PARTIAL FULFILLMENT TOWARDS THE COMPLETION OF **PROJECT BASED LEARNING-II** IN **ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

Project Guide                                                    Head, Deptt. of AI & DS
**(Prof.R.V.Panchal)**                                        **(Digambar Padulkar)**

Principal
**(Dr. R. S. Bichkar)**

**Internal Examiner**                                              **External Examiner**

# Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ' Customer Segmentation'.

# Abstract

Effective decisions are mandatory for any company to generate good revenue. In these days competition is huge and all companies are moving forward with their own different strategies. We should use data and take a proper decision. Every person is different from one another and we don't know what he/she buys or what their likes are. But, with the help of machine learning technique one can sort out the data and can find the target group by applying several algorithms to the dataset. Without this, It will be very difficult and no better techniques are available to find the group of people with similar character and interests in a large dataset. Here, The customer segmentation using K-Meansclustering helps to group the data with same attributes which exactly helps to business the best. We are going to use elbow method to find the number of clusters and at last we visualize the data    ł

# Contents

# SYNOPISIS

## 1.1 Project Title

CUSTOMER SEGMENTATION

## 1.2 Techinal Keywords

Clustering, Elbow Method, K-Means Algorithm, Customer Segmentation.

## 1.3 Problem Statement

Segmentation is the best application of unsupervised learning. Using clustering, identify segments of customers in the dataset to target the potential user base. They divide customers into various groups according to common characteristics like gender, age, interest, and spending habits so they can market to each group effectively. Use K-Means Clustering and also visualize the gender and age distributions. Then analyze their annual income and spending scores. As it describes about how we can divide the customers based on their similar characteristics according to their needs by using k-means clustering which is a classification of unsupervised machine learning.

## 1.4 Abstract

Effective decisions are mandatory for any company to generate good revenue. In these days competition is huge and all companies are moving forward with their own different strategies. We should use data and take a proper decision. Every person is different from one another and we don't know what he/she buys or what their likes are. But, with the help of machine learning technique one can sort out the data and can find the target group by applying several algorithms to the dataset. Without this, It will be very difficult and no better techniques are available to find the group of people with similar character and interests in a large dataset. Here, The customer segmentation using K-Means clustering

helps to group the data with same attributes which exactly helps to business the best. We are going to use elbow method to find the number of clusters and at last we visualize the data.

## 1.5 Goals and Objectives

Goals : The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business Objectives: 1) To reduce risk in deciding where, when, how, and to whom a product, service, or brand will be marketed; 2) To increase marketing efficiency by directing effort specifically toward the designated segment in a manner consistent with that segment's characteristics.

## 1.6 Plan of Project Execution

## 1.7 Project Plan

### Project Plan

| TASK | START DATE | Duration |
|---|---|---|
| 1)Identifying Problem statement | 2/10/2022 | 10 |
| 2) Gathering Resources about Project | 2/21/2022 | 7 |
| 3) Planning | 3/14/2022 | 7 |
| 4) Designing | 3/22/2022 | 6 |
| 5) PPT of Project | 3/29/2022 | 3 |
| 6) Coding | 3/31/2022 | 4 |
| 7) Project Report | 4/12/2022 | 8 |

**2**

# Technical Keywords

## 2.1 Area of Project

Project consists of Machine Learning.Different Machine Learning tools and techniques used to analyse and extract segments.

## 2.2 Technical Keywords

Clustering algorithms generates clusters such that within the clusters are similar based on some characteristics. Similarity is defined in terms of how close the objects are in space. The elbow method is based on the observation that increasing the number of clusters can help to reduce the sum of within-cluster variance of each cluster. This is because having more clusters allows one to capture finer groups of data objects that are more similar to each other. K-means algorithm in one of the most popular centroid based algorithm.The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster. Customer Segmentation is one of most important applications of unsupervised learning.

# 3

# Introduction

## 3.1  Motivation of the Project

It is finding the most profitable customer groups within the entire pool of customers. With this we can better understand customer interests, choices and purchasing patterns which will give us an idea of which attributes are more tightly related to the customers and the business as well.

## 3.2  Literature Survey

Customer Classification Over the years, the commercial world has become more competitive, as organizations such as these have to meet the needs and wants of their customers, attract new customers, and thus improve their businesses. The task of identifying and meeting the needs and requirements of each customer in the business is a very difficult task. This is because customers may vary according to their needs, wants, demographics, shapes, taste and taste, features and so on. As it is, it is a bad practice to treat all customers equally in business. This challenge has led to the adoption of the concept of customer segmentation or market segmentation, where consumers are divided into sub-groups or segments where members of each subcategory exhibit similar market behaviors or features. Accordingly, customer segmentation is the process of dividing the market into indigenous groups. Clustering data Clustering is the process of grouping the information in the dataset based on some similarities. There are a number of algorithms which can be chosen to be applied on a dataset based on the situation provided. However, no universal clustering algorithm exists that's why it becomes important to opt for appropriate clustering techniques. In this paper, we have implemented three clustering algorithms using python sklearn library. E. K-Mean K- means that an algorithm is one of the most popular classification algorithm. This clustering algorithm depends on the centroid where each data point is placed in one of the overlapping K clusters pre-programmed into the algorithm, The clusters are created that correspond to the hidden pattern in the data that provides the information needed to help decide the execution process

# Problem Definition And Scope

## 4.1 Problem Statement

Segmentation is the best application of unsupervised learning. Using clustering, identify segments of customers in the dataset to target the potential user base. They divide customers into various groups according to common characteristics like gender, age, interest, and spending habits so they can market to each group effectively. Use K-Means Clustering and also visualize the gender and age distributions. Then analyze their annual income and spending scores. As it describes about how we can divide the customers based on their similar characteristics according to their needs by using k-means clustering which is a classification of unsupervised machine learning.

### 4.1.1 Goals and Objetives

Goals : The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business Objectives: 1) To reduce risk in deciding where, when, how, and to whom a product, service, or brand will be marketed; 2) To increase marketing efficiency by directing effort specifically toward the designated segment in a manner consistent with that segment's characteristics.

### 4.1.2 Statement of Scope

To cluster customers based on common purchasing behaviors for future operations/marketing projects To incorporate best mathematical, visual, programming, and business practices into a thoughtful analysis that is understood across a variety of contexts and disciplines To investigate how similar data and algorithms could be used in future data projects To create an understanding and inspiration of how data science can be used to solve real-world problems

## 4.2   Methodology of Problem Solving

The data set used to implement clustering and K-means algorithm was collected from a store of data.The attributes in the data set has CustomerId, gender, age, annual income, spending score.K-means algorithm in one of the most popular centroid based algorithm. .The elbow method is based on the observation that increasing the number of clusters can help to reduce the sum of within-cluster variance of each cluster. This is because having more clusters allows one to capture finer groups of data objects that are more similar to each other. To define the optimal clusters, Firstly, we use the clustering algorithm for various values of k. This is done by ranging k from 1 to 10 clusters. Then we calculate the total intra-cluster sum of square. Then, we proceed to plot intra-cluster sum of square based on the number of clusters.Average Silhouette Method:With the help of the average silhouette method, we can measure the quality of our clustering operation. With this, we can determine how well within the cluster is the data object. If we obtain a high average silhouette width, it means that we have good clustering. The average silhouette method calculates the mean of silhouette observations for different k values. With the optimal number of k clusters, one can maximize the average silhouette over significant values for k clusters.Using the silhouette function in the cluster package, we can compute the average silhouette width using the kmean function. Here, the optimal cluster will possess highest average.Gap Statistics method: For computing the gap statistics method we can utilize the clusGap function for providing gap statistic as well as standard error for a given output.

## 4.3   Outcomes

Finding an optimal number of unique customer groups will help you understand how your customers differ, and help you give them exactly what they want. Customer segmentation improves customer experience and boosts company revenue. Discovering all of the different groups that build up a more meaningful customer base permits you to get into customers' brains and give them precisely what they crave, enhancing their participation and expanding profits.

## 4.4   Application

Data Analysis Clustering can also helps to discover distinct groups in their customer base

## 4.5   Constraints

Dataset of customer should be in form of Zip format

## 4.6  Software and Hardware resources required

Hardware required Processor:-intel i3

Storage:-2KB

Software required

Platform:Windows

IDE:Google Colab

Programming Language:Python

# 5

# Project Plan

## 5.1 Project Resources

1.People: There are total four peoples in a group.We have divided project work among us.

2.Hardware:
Storage:2kB
Processor:intel i3

3.Software:
Platform:Operating System:Windows 7 or above
IDE:Jupyter Notebook or Google Colab
Programming Language:Python
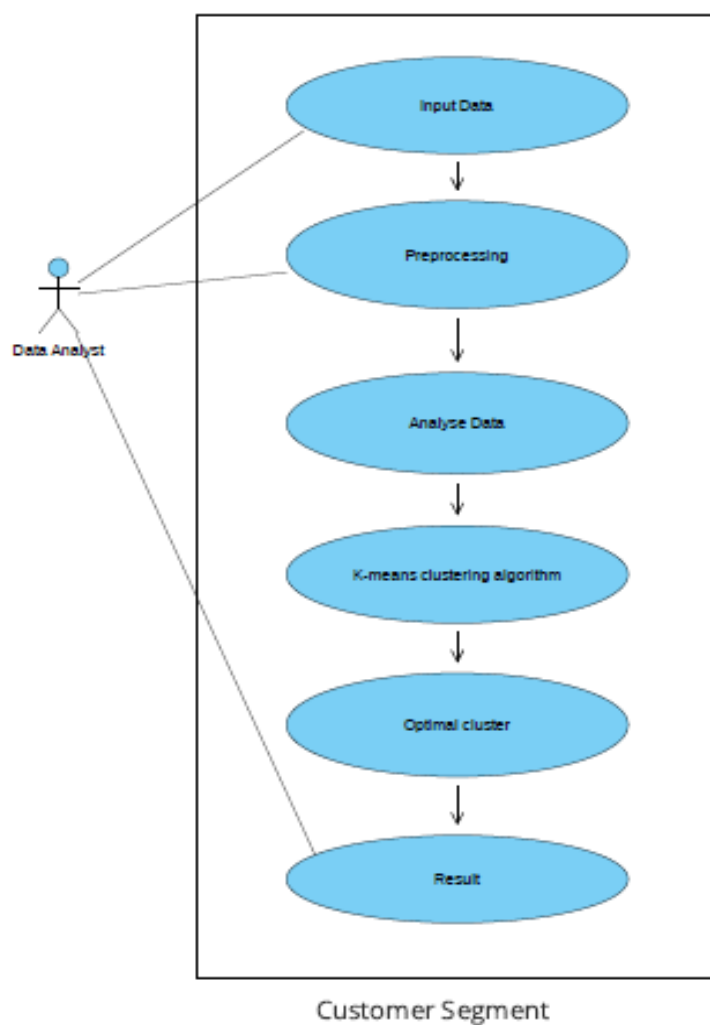
## 5.2    Plan of Project execution

# Project Plan

| TASK | START DATE | Duration |
|------|-----------|----------|
| 1)Identifying Problem statement | 2/10/2022 | 10 |
| 2) Gathering Resources about Project | 2/21/2022 | 7 |
| 3) Planning | 3/14/2022 | 7 |
| 4) Designing | 3/22/2022 | 6 |
| 5) PPT of Project | 3/29/2022 | 3 |
| 6) Coding | 3/31/2022 | 4 |
| 7) Project Report | 4/12/2022 | 8 |

## Customer Segmentation

| Task | Date | Duration |
|------|------|----------|
| 1)Identifying Problem statement | February 10, 2022 | 10 |
| 2) Gathering Resources about... | February 21, 2022 | 7 |
| 3) Planning | March 14, 2022 | 7 |
| 4) Designing | March 22, 2022 | 6 |
| 5) PPT of Project | March 29, 2022 | 3 |
| 6) Coding | March 31, 2022 | 4 |
| 7) Project Report | April 12, 2022 | 8 |

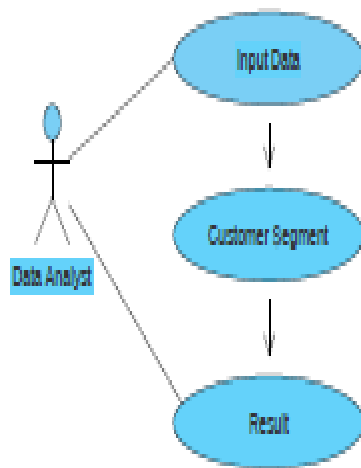January 14, 2022     February 8, 2022     March 5, 2022     March 30, 2022

Duration in  ■ Days

# Software Requirement Specification

## 6.1 Use case Diagram



Customer Segment

# Result Analysis

## 7.1 Graphs

Spending Scores



Annual Incomes



Clusters of Customers
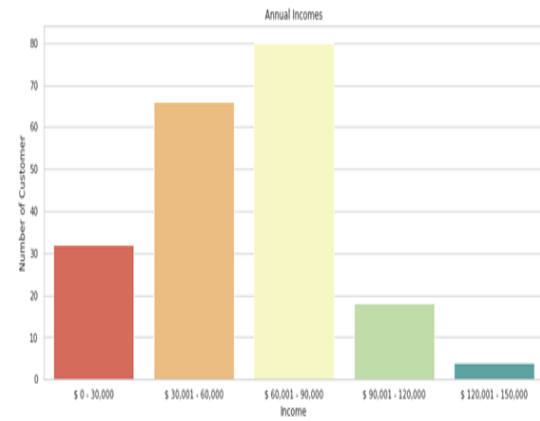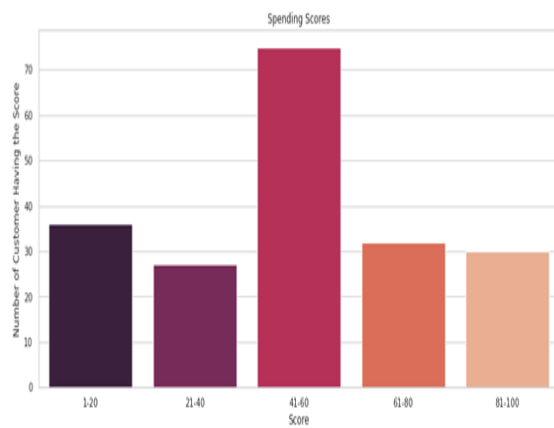
# 8
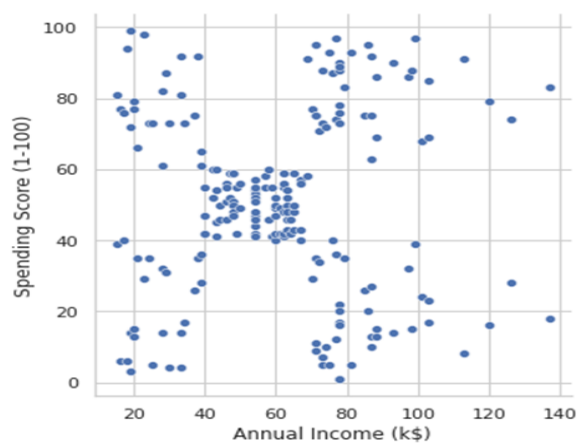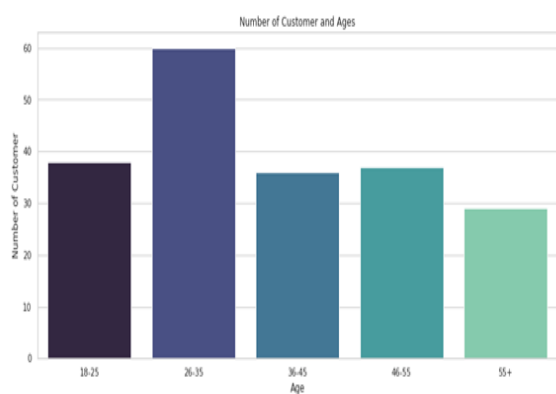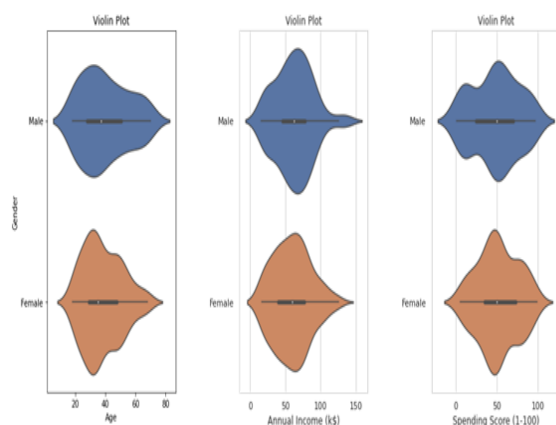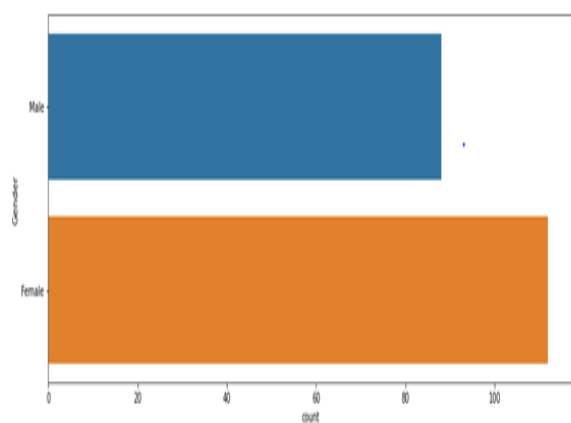
# Conclusion

So we concluded that the , The Highest income , high spending can be target these type of customers as they earn more money and spend as much as they want. Highest income, low spending can be target these type of customers by asking feedback and advertising the product in a better way. Average income, Average spending may or may not be beneficial to the mall owners of this type of customers. Low income, High spending can be target these type of customers by providing them with low-cost EMI's etc. Low income, Low spending don't target these type of customers because they earn a bit and spend some amount of money. So high income, high spending are the most beneficial ones to the mall owners which increases the owner's business.

# 9

# References

[1] Cooil, B., Aksoy, L. Keiningham, T. L. (2008), 'Approaches to customer segmentation', Journal of Relationship Marketing 6(3-4), 9–39.

[2] D. Aloise, A. Deshpande, P. Hansen, and P. Popat,"The Basis Of Market Segmentation" Euclidean sum-of-squares clustering," Machine Learning, vol. 75, pp. 245-249, 2009.

[3] I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.