1. **INTRODUCTION**

**1.1 Problem Statement** -

Customer Segmentation is a popular application of unsupervised learning. Using clustering, identify segments of customers to target the potential user base. They divide customers into groups according to common characteristics like gender, age, interests, 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 incomes and spending scores.

Language: Python

Dataset/Package: https://drive.google.com/file/d/19BOhwz52NUY3dg8XErVYglctpr5sjTy4/view

**1.2 Motivation -**

A good application of clustering algorithms can be its application In the field of marketing, where applying clustering can or may be used to identify various customer groups with existing customer data. Based on that, customers can be provided with discounts, offers, promo codes etc. Let's say one owns a supermarket or a shopping center like a mall and via membership cards or other possible means you have some basic data about your customers such as customer number, age, gender, annual income and spend score, which is something you assign to a customer based on parameters you define such as customer behavior and purchase data.

An important objective we can look to achieve is to learn the purpose of customer segmentation concepts, also known as market basket analysis, to try to understand customers and segment them into different groups according to their preferences, and once the division is done, this information can be utilized by to the marketing team to plan various selling strategies.

Mall Customer data is an interesting dataset that has tabulated customer data. It puts you in the shoes of the owner of a supermarket. You have customer data, and on this basis of the data, you have to divide the customers into various groups.

**2. IMPLEMENTATION**

### **2.1 Environment and Tools**

Coding language - Python

Libraries Used -

* Pandas
* Numpy
* Matplotlib
* Seaborn
* Sklearn (for k-means)

Algorithm - K-means for Clustering

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 applications 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.

**2.2 Results and Code**

Google colab link -

<https://colab.research.google.com/drive/1K0fjQfPvaOHyr2qdkNi83GmoOaFIGP3O?usp=sharing>

**3. CONCLUSION**

KMeans Clustering is a powerful algorithm that helped us to analyze customer segmentation for mall customers data. It helped us to understand the behavior of different customers and thus we can plan a good marketing strategy accordingly. There isn't much difference between the spending score of women and men, which leads us to think that our behavior when it comes to shopping is pretty similar. We also observed that the ones who spend more money in malls are young people. That is to say they are the main target when it comes to marketing, so doing deeper studies about what they are interested in may lead to higher profits. Although younglings seem to be the ones spending the most, we can't forget there are more people we have to consider, like people who belong to the pink cluster, they are what we would commonly name after "middle class" and it seems to be the biggest cluster. Promoting discounts on some shops can be something of interest to those who don't actually spend a lot and they may end up spending more.

