# Customer Segmentation using K-Means Clustering

Created by: J. TAKESHWAR

## Introduction

Customer segmentation is a technique used by businesses to categorize their customers into distinct groups based on common characteristics. In this document, we explain a Python-based approach to segment customers using K-Means clustering.

## Code Explanation

### Step 1: Import Required Libraries

The following Python libraries are used:  
- pandas: For handling structured data.  
- numpy: For numerical computations.  
- matplotlib.pyplot: For visualizing data.  
- sklearn.cluster.KMeans: For applying K-Means clustering.  
- sklearn.preprocessing.StandardScaler: For normalizing data before clustering.

### Step 2: Generate Synthetic Customer Data

We generate a dataset of 5000 customers with the following attributes:  
- customer\_id: Unique customer identifier.  
- products\_purchased: Random number of products purchased (1 to 15).  
- complains: Whether a customer has made complaints (2% chance of complaint).  
- money\_spent: Total amount spent (based on a random multiplier between 10 and 300).

### Step 3: Data Preprocessing

To ensure effective clustering, we scale the numerical features using StandardScaler.

### Step 4: Apply K-Means Clustering

We use the K-Means algorithm to segment customers into 4 clusters.

### Step 5: Save Clustered Data

We save the clustered dataset as a CSV file for further analysis.

### Step 6: Visualization

We create a scatter plot of money\_spent vs. products\_purchased, with colors representing clusters.

### Step 7: Display Cluster Distribution & Summary Statistics

The cluster distribution shows how customers are segmented. The summary statistics provide insights into customer behavior in each cluster.

## Conclusion

This project demonstrates how to segment customers based on purchasing behavior and complaints using K-Means clustering. The results help businesses understand their customer base and tailor marketing strategies accordingly.