CUSTOMER SEGEMENTATION

# INTRODUCTION

The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes. The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction. This project involves data collection, data preprocessing, feature engineering, clustering algorithms, visualization, and interpretation of results.

In this phase the building and loading of data flow of customer segmentation is going

to be done.

## PREREQUISITES FOR BUILDING A CUSTOMER SEGMENTATION MODEL

* The data is obtained from [https://www.Kaggle.com/data](https://www.kaggle.com/data)
* Have the following libraries installed — Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Kneed, and Scipy.
* Columns Required from dataset

1. CustomerID
2. Gender
3. Age
4. Annual Income
5. Spending Score

## UNDERSTAND THE SEGMENTATION DATA

Before starting any data science project, it is vital to explore the dataset and understand each variable.

* Libraries Imported :

1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn

* Loading the Data

**df=pd.read\_csv(‘/kaggle/input/mall-customers/Mall\_Customers.csv’)**

* let’s look at the head of the dataframe:

**df.head()**

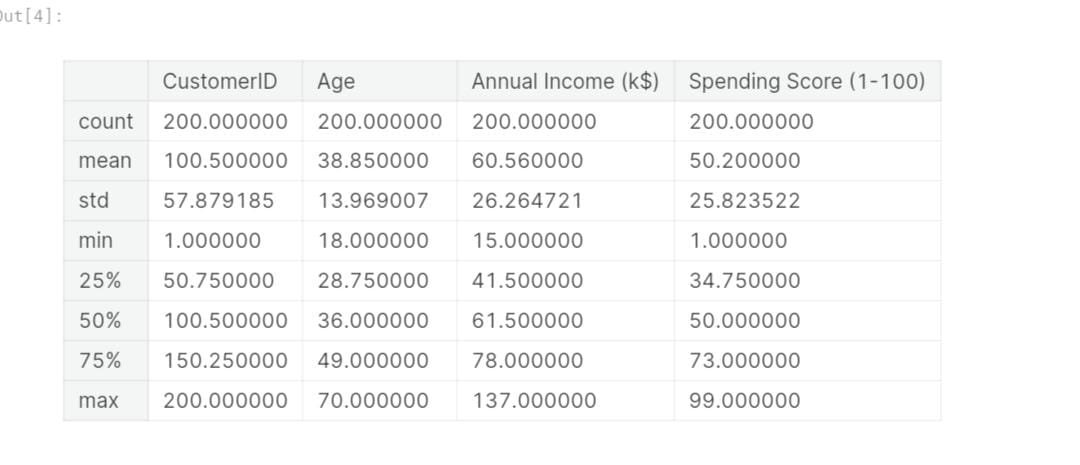


## PREPROCESSING DATA FOR SEGMENTATION

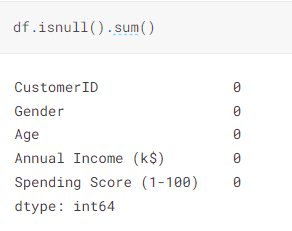
The raw data we downloaded is complex and in a format that cannot be easily ingested by customer segmentation models. We need to do some preliminary data preparation to make this data interpretable.

* Description

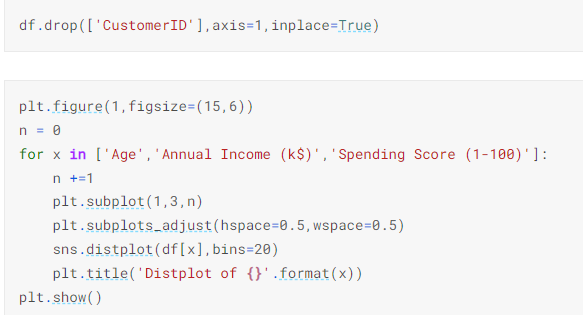
**df.describe()**

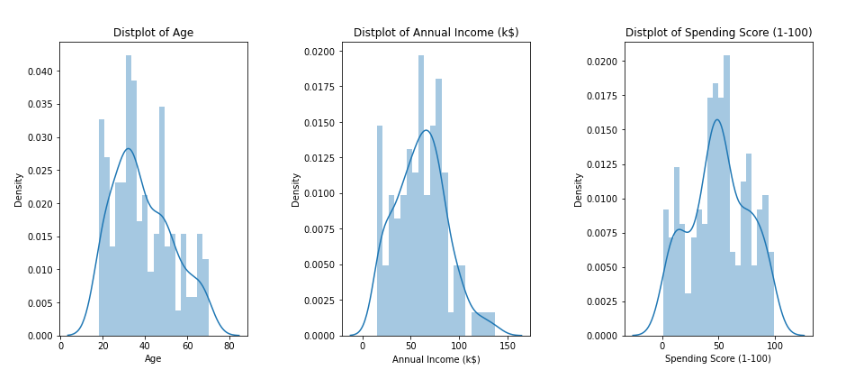


* Null Values

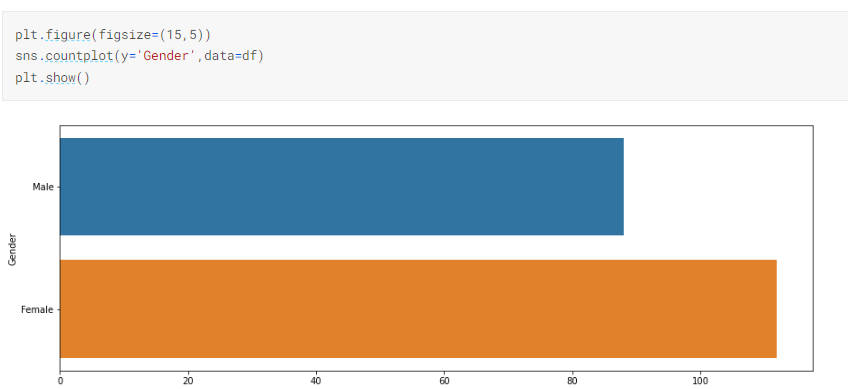


* Dropping





* CounterPlot

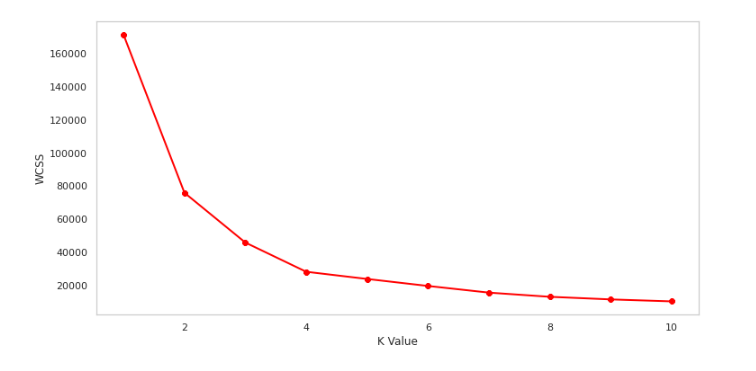


## BUILDING THE CUSTOMER SEGMENTATION MODEL

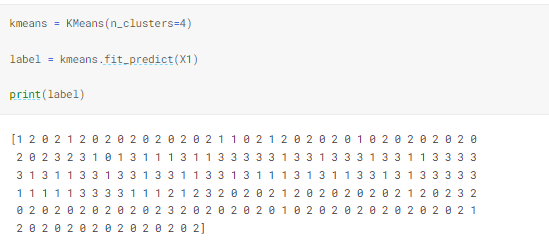
  We are going to create a K-Means clustering algorithm to perform customer segmentation.The goal of a K-Means clustering model is to segment all the data available into non-overlapping sub-groups that are distinct from each other.

* K-Means

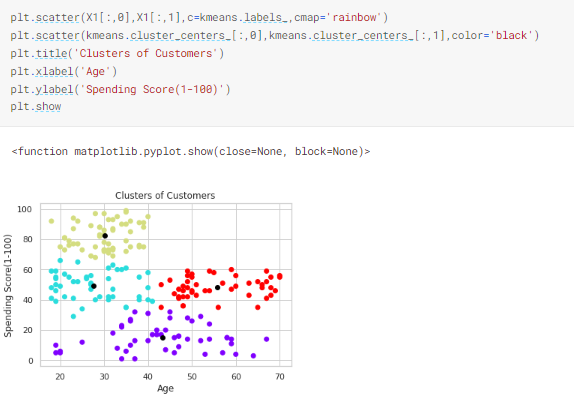




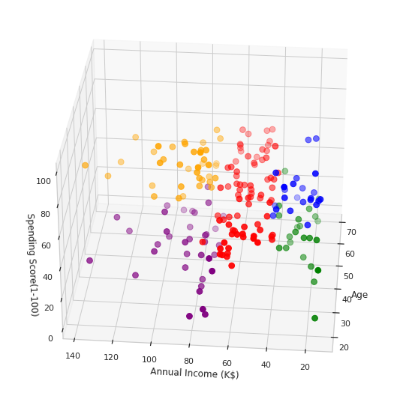
* K-Means Cluster



* Clusters of Customers

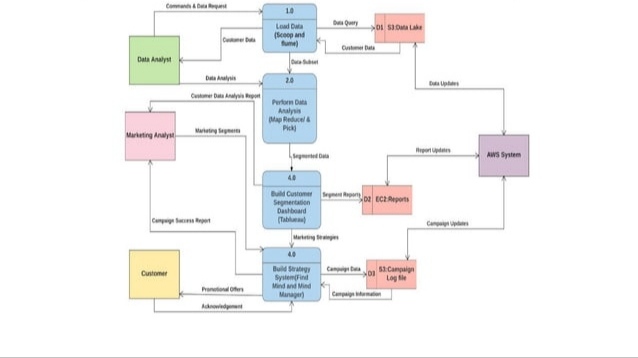


* 3D Model



DATA FLOW OF CUSTOMER MODEL

1.Physical Flow



2.Logical Flow

