

PHASE-3

3.1 Dataset: <https://www.kaggle.com/datasets/akram24/mall-customers>

This dataset is used in retail and marketing analytics to understand customer behavior and preferences. It includes the following types of information:

- Customer ID
- Gender
- Age
- Annual Income
- Spending Score

A "customer ID" (Customer Identification) is a unique identifier assigned to each customer in a database or system. It is used to distinguish one customer from another and track their activities, purchases, interactions, and other relevant information.

Gender is one of the key factors in segmenting customers into distinct groups. For example, stores may tailor their product offerings and marketing strategies differently for male and female customers.

Age is a fundamental factor for segmenting customers into groups. Different age groups may have distinct preferences, shopping behaviors, and income levels. For example, retailers often distinguish between teenagers, young adults, middle-aged individuals, and seniors.

The annual income of mall customers is a crucial demographic variable that helps businesses and mall operators understand the spending capacity and shopping preferences of their customer base.

Spending score is a metric used to assess and quantify a customer's purchasing behavior within a mall.

3.2 Loading the dataset

3.3 Preprocessing Dataset

3.4 Performing different analysis

Executed in Python Notebook (.ipynb) file in git repository.

Git repository link:

<https://github.com/adrieljoshua/CustomerSegmentation-AppliedDataScience.git>