**Phase 1: Problem Definition and Design Thinking**

In this part you will need to understand the problem statement and create a document on what have you understood and how will you proceed ahead with solving the problem. Please think on a design and present in form of a document.

**Problem Definition:**

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.

**Dataset Link:**[**https://www.kaggle.com/datasets/akram24/mall-customers**](https://www.kaggle.com/datasets/akram24/mall-customers)



**Design Thinking:**

Data collection:

Data collection is the process of collecting and analyzing information on relevant variables in a predetermined, methodical way so that one can respond to specific research questions, test hypotheses, and assess results.

Data preprocessing:  
 Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task.

Feature Engineering:

Feature Engineering is the process of creating new features or transforming existing features to improve the performance of a machine-learning model. It involves selecting relevant information from raw data and transforming it into a format that can be easily understood by a model.

Clustering Algorithms:

Clustering algorithms are procedures for partitioning data into groups or clusters such that the clusters are distinct, and members of each cluster belong together.

Visualization:

Visualization or visualisation may refer to: Visualization (graphics), the physical or imagining creation of images, diagrams, or animations to communicate a message. Data and information visualization, the practice of creating visual representations of complex data and information.

Interpretation:

Analyze and interpret the characteristics of each customer segment to derive actionable insights for marketing strategies.

Program:

import numpy as np linear algebra

import pandas as pd

import os

for dirname,. filenames in os.walk(' C:\Users\Students\Documents\New folder\Mall Customers '):

for

filename in filenames:

print(os.path.join(dirname, filename))

import numpy as np

import pandas as pd import matplotlib.pyplot as plt

import seaborn as sns

df pd.read\_csv('/kaggle/input/mall-customers/Mall Customers.csv')

df.rename(columns=('Genre: 'Gender' }, inplace=True)

df.head()

**output:**

