PHASE 1 : Problem Definition and Design Thinking

Project Title: Customer Segmentation with Data Science

Problem Statement: Implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes, enabling businesses to personalize marketing strategies and enhance customer satisfaction.

Dataset Link: https://www.kaggle.com/datasets/akram24/mall-customers

Problem Definition: The problem at hand is to leverage data science techniques to effectively segment customers based on various aspects such as their behavior, preferences, and demographic attributes. The ultimate goal of this project is to empower businesses with the ability to tailor their marketing strategies in a personalized manner, ultimately leading to improved customer satisfaction and potentially increased sales.

The process involved in achieving this objective can be broken down into several key steps :

➔ Data Collection
➔ Data Preprocessing
➔ Feature Engineering
➔ Clustering Algorithms
➔ Visualization
➔ Interpretation

Data Collection: The first step is to gather customer data from various sources. This data encompasses a wide range of attributes, including but not limited to purchase history, demographic information (age, gender, location), and customer interaction behavior (e.g., website visits, social media

engagement, email response rates). Collecting diverse data points is crucial as it provides a holistic view of each customer.

Data Preprocessing: Once the data is collected, it needs to be cleaned and preprocessed. This involves tasks such as handling missing values, dealing with outliers, and ensuring data consistency. Additionally, categorical features like gender or product categories may need to be converted into numerical representations (one-hot encoding, label encoding) that machine learning algorithms can work with.

Feature Engineering: Creating new features is a vital step in understanding customer behavior. Features such as total spending, frequency of purchases, and average order value can be engineered from the raw data. These features help in capturing meaningful patterns and characteristics of each customer.

Clustering Algorithms: The heart of the project lies in applying clustering algorithms to segment the customer base. Common algorithms like K-Means, DBSCAN, or hierarchical clustering can be employed for this purpose. These algorithms group customers with similar attributes and behavior into distinct clusters or segments.

Visualization: Visualization plays a crucial role in presenting the results in an understandable and actionable manner. Techniques like scatter plots, bar charts, and heatmaps can be used to visualize the identified customer segments. This step helps stakeholders grasp the differences and similarities between segments intuitively.

Interpretation: Finally, the project requires in-depth analysis and interpretation of the identified customer segments. This involves understanding the characteristics, needs, and preferences of each segment. By doing so, businesses can derive actionable insights that can inform marketing strategies, product offerings, and customer engagement tactics.