**Ex No: 12 Implementation of clustering techniques –Hierarchical**

**Problem Scenario**

A retail store wants to understand its customers better in order to design targeted marketing strategies. The management has collected data on **Annual Income** and **Spending Score** of customers at the mall. However, the customer base is diverse, and there is no prior information on natural groupings within the data.

Link: [Dataset](https://gist.githubusercontent.com/pravalliyaram/5c05f43d2351249927b8a3f3cc3e5ecf/raw/8bd6144a87988213693754baaa13fb204933282d/Mall_Customers.csv)

To address this, we aim to perform **Hierarchical Clustering**, an unsupervised machine learning technique that groups data points into clusters based on similarity. Customers with similar income levels and spending habits will be grouped together, allowing the store to identify segments such as **high-income, high-spending customers** or **low-income, low-spending customers**.

1. **Import Required Libraries**  
   Load required Python libraries like numpy, pandas, matplotlib, scipy.cluster.hierarchy, and sklearn.cluster.
2. **Load Dataset**
   * Read the CSV file into a Pandas DataFrame.
   * Display the first few rows to confirm successful loading.
3. **Feature Selection**
   * Select **Annual Income** and **Spending Score** columns for clustering.
   * Store them in a variable X.
4. **Construct Dendrogram**
   * Use scipy.cluster.hierarchy.linkage with method = ward.
   * Plot dendrogram to find the optimal number of clusters.
5. **Apply Hierarchical Clustering**
   * Use AgglomerativeClustering from scikit-learn.
   * Set n\_clusters based on dendrogram result (e.g., 5).
   * Fit model on dataset and get cluster labels.
6. **Visualize Clusters**
   * Plot the clusters using scatter plot with different colors.
   * Add labels, legends, and axis names for clear interpretation.