**CUSTOMER SEGMENTATION USING DATA SCIENCE**

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

**Problem Definition:**

The key aim of this project is to configure marketing strategies and improve customer satisfaction via data science methodologies, resulting in an array of benefits such as enhanced customer engagement, higher conversion rates, customer loyalty and contentment, increased revenue, and so on. This segmentation empowers organizations to manipulate their marketing tactics and services to distinct consumer segments, thus enhancing patron satisfaction and driving business prosperity.

**Design Thinking:**

**1.Data Collection:** We negotiate data from databases such as Kaggle, where the data is categorized by CustomerID, Genre, Age, Annual Income(k$), and Spending score.

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

**2.Data Preprocessing:** Eliminating errors, inconsistencies, and inaccuracies is crucial for preparing raw data for analysis, modeling, and delivering enhanced outcomes. To execute data purification and exploration techniques, we employ Python and modules such as pandas, Scikit-Learn, Seaborn, matplotlib, and Numpy.

**3.Feature Engineering:** we analyse clients behaviour and preferences, such as total spending, frequency of purchase, and the monetary unit through RFM analysis.

* **Recency**: duration of time since their last purchase
* **Frequency**: metric that reflects how adequate a customer purchases or visits.
* **Monetary**: the total value of their purchases

In marketing and consumer segmentation, RFM (Recency, Frequency, Monetary) scores are used to more effectively recognise and categorize clients based on their commercial behaviour.

**4.Clustering Algorithm:** Here we’re productively engaged with unsupervised machine learning algorithm i.e., the K-means clustering algorithm. It employs multiple iterations to segment the unlabeled data points into k distinct groups.

**5.Visualization:** We actively leverage Python-based data visualization tools because of their broad range of alternatives for building visualizations lined up to marketing data and targets. Plotly, Matplotlib, Seaborn, Bokeh, and Pandas are some of the popular Python tools that we will use to visualize data. Data profiling advances such as OpenRefine and Trifacta can assist with data quality control.

**6.Interpretation:** We use machine learning frameworks like scikit-learn (Python) to develop predictive models by identifying key elements influencing outcomes for data interpretation. IBM SPSS, SAS, and Alteryx sophisticated Analytics Platforms provide sophisticated analytics and data mining capabilities for in-depth data interpretation. Jupyter Notebooks are interactive workspaces for Python data research.