**Technical Requirements Document (TRD):**

* **Data Sources:** Mall Customers dataset

**Dataset Description:** The dataset used in this project is the "Mall Customers" dataset, which provides information about customers from a mall. The dataset contains demographic and behavioral attributes of the customers, which can be used to perform segmentation. The dataset includes the following columns:

1. **CustomerID:** Unique identifier for each customer.
2. **Gender:** Gender of the customer (Male/Female).
3. **Age:** Age of the customer.
4. **Annual Income (k$):** Annual income of the customer in thousands of dollars.
5. **Spending Score (1-100):** Spending score assigned by the mall based on customer behavior and spending nature (1 being lowest and 100 being highest).

* **Technologies:** Python, Jupyter Notebook, Matplotlib, Seaborn, Scikit-learn, Power BI

**Python**

* **Purpose:** Primary programming language for data manipulation, analysis, and machine learning.
* **Key Features:** Extensive libraries for data handling (Pandas), numerical computations (NumPy), and machine learning (Scikit-learn).

**Jupyter Notebook**

* **Purpose:** Interactive environment for data exploration, analysis, and documentation.
* **Key Features:** Supports code execution, visualization, and narrative text in a single document format.

**Matplotlib**

* **Purpose:** Python plotting library for creating static, animated, and interactive visualizations.
* **Key Features:** Generates a wide variety of charts, graphs, and plots to visualize data distributions and trends.

**Seaborn**

* **Purpose:** Statistical data visualization library based on Matplotlib, offering enhanced aesthetic appeal and ease of use.
* **Key Features:** Simplifies complex visualizations such as heatmaps, violin plots, and categorical plots.

**Scikit-learn**

* **Purpose:** Machine learning library in Python for data mining and data analysis tasks.
* **Key Features:** Provides tools for clustering (K-means, DBSCAN), classification, regression, and model evaluation.

**Power BI**

* **Purpose:** Business intelligence tool for creating interactive reports and dashboards.
* **Key Features:** Integrates data sources, performs data modeling, and generates visualizations for business insights.
* **Architecture:**
* Data preprocessing, EDA, clustering, and visualization

**Data Preprocessing**

* **Purpose:** Prepare raw data for analysis by cleaning, transforming, and integrating datasets.
* **Key Steps:** Handle missing values, remove duplicates, standardize formats, and encode categorical variables.

**Exploratory Data Analysis (EDA)**

* **Purpose:** Understand the structure and patterns in the data before applying clustering algorithms.
* **Key Techniques:** Summary statistics, visualizations (like histograms and scatter plots), and correlation analysis.

**Clustering**

* **Purpose:** Group similar customers together based on their attributes or behaviors.
* **Key Algorithms:** K-means, hierarchical clustering, DBSCAN, or other methods depending on data characteristics.

**Visualization**

* **Purpose:** Present insights and segmented groups visually for easy interpretation and decision-making.
* **Key Tools:** Matplotlib, Seaborn, and possibly tools like Power BI for creating interactive dashboards.
* **Data Flow:** Import data → Clean data → Analyze data → Segment customers → Visualize results

This structured data flow ensures that raw data is processed, analyzed, segmented into meaningful customer groups, and visualized effectively to derive actionable insights for business strategies in customer segmentation projects.