Technical Requirements Document (TRD):

1. Data Sources:

1.1 Primary Data Source

Mall Customers Dataset: This dataset includes customer demographic information (age, gender, income) and purchasing behaviour (spending scores, purchase history).

1.2 Data Format

The dataset is typically in CSV format but can also be in other formats such as Excel, JSON, or a database table.

2. Technologies:

2.1 Programming Language

- **Python**: Used for data analysis, clustering, and visualization.

2.2 Development Environment

- **Jupyter Notebook**: For interactive data analysis and visualization.

2.3 Libraries and Frameworks

- **Matplotlib**: For creating static, animated, and interactive visualizations in Python.
- **Seaborn:** For statistical data visualization built on top of Matplotlib.
- **Scikit-learn:** For machine learning, including clustering algorithms.

2.4 Visualization Tools

- **Power BI:** For creating dashboards and detailed visual reports.

3. Architecture:

3.1 Data Preprocessing

- **Data Cleaning**: Handling missing values, removing duplicates, correcting inconsistencies, and normalizing data.
- **Feature Engineering**: Creating new features or transforming existing ones to improve the performance of clustering algorithms.

3.2 Exploratory Data Analysis (EDA)

- **Descriptive Statistics**: Summarizing the main features of the dataset.
- **Visualization**: Using plots (e.g., histograms, box plots, scatter plots) to understand data distributions and relationships.

3.3 Clustering

- **Algorithm Selection**: Choosing appropriate clustering algorithms such as K-means.
- **Model Training**: Training the clustering model on the preprocessed dataset.
- **Evaluation**: Assessing the quality and performance of the clustering results.

3.4 Visualization

- **Visual Representation**: Creating visualizations to represent customer segments, including bar charts, pie charts, and scatter plots.

4. Data Flow:

4.1 Import Data

- Load the mall customers dataset into the working environment.

4.2 Clean Data

- Handle missing values, remove duplicates, and normalize the data for consistent analysis.

4.3 Analyse Data

- Perform exploratory data analysis to understand data distributions and identify patterns.

4.4 Segment Customers

- Apply clustering algorithms to segment customers based on their demographics and purchasing behaviour.

4.5 Visualize Results

- Create visualizations to represent the customer segments and develop interactive dashboards for detailed insights.