Technical Requirements Document

Data Sources : Mall Customers dataset

Technologies:

- Programming Language :
  Python

- Development Environment :
  Jupyter notebook

- Libraries and Tools :

Tableau : for dashboard

Pandas : for data manipulation

Numpy : for numerical operations

Matplotlib : for visualization

Seaborn : for visualization

Scikit-learn : For building the model using machine learning algorithm, training and Evaluation.

Architecture :

- Data Collection : Import the Mall Customers dataset

- Data Preprocessing : Handle missing values, normalize data formats, and remove any outliers to ensure data quality.

- Feature Engineering : Modify and select features to improve effectiveness of clustering

- Exploratory Data Analysis :

Visualize the statistical data to understand data distribution

Identify patterns and gain insights from data

- Clustering :

Apply K-Means clustering algorithm to segment customers

Train the model on cleaned data

Evaluate clusters using inertia and silhouette scores to improve their quality

- Visualization :

Create plot to identify different customer segments

Create Dashboard to summarize insights

Data Flow :

Import Mall Customers data -> Clean Data -> Perform EDA for analysis -> Segment Customers using K-Means Clustering -> Visualize Results

Performance considerations :

Optimize data processing and clustering to handle large datasets efficiently.

Intuitive visualizations to easily interpret the results

Security and Compliance :

Ensure that the data is handles in accordance with privacy regulations.

Maintain accuracy and consistency throughout the process.