Industry Standard Documentation

Project Charter:

- **Project Title**: Mall Customer Segmentation
- **Project Manager:** Sanjeev Thakur
- **Start Date:** 13-07-2024
- **End Date:** 17-07-2024
- **Objectives:**. This project aims to perform customer segmentation on a Mall customer dataset using the K-Means clustering algorithm. The dataset contains information about customers such as their age, gender, annual income, and spending score. The goal of this project is to cluster the customers based on their purchasing behavior and demographic characteristics. Also aims to to segment customers into distinct groups based on their purchasing behavior

- **Scope:**

  Data cleaning: Collection and preprocessing of the data which includes handling missing values, ensuring correct data type, remove duplicates, handling outliers, and normalization

  Customer segmentation using K-Means :

  This Clustering Analysis gives us a very clear insight about the different segments of the customers in the Mall. There are clearly 5 segments of Customers based on their Annual Income and Spending Score which are reportedly the best factors/attributes to determine the segments of a customer in a Mall.

  Visualization using Matplotlib and Power BI.

  Matplotlib is a Python library for creating static, animated, and interactive visualizations. It's widely used for plotting graphs, charts, histograms, and more in Python scripts or notebooks.Power BI, on the other hand, is a

business analytics tool by Microsoft that lets you visualize and share insights from data. It connects to various data sources, allows for interactive dashboard creation, and supports powerful data visualizations.Combining Matplotlib with Power BI means leveraging Python's robust plotting capabilities within Power BI's data analysis and visualization environment.

Exploratory Data Analysis and Descriptive Statistics :

Performing EDA to understand relationship between features, analyze trends in data

Data Visualization:

Use matplotlib and seaborn to present insights and clusters

Unsupervised Machine Learning (K-Means Clustering):

- **Deliverables:**

Useful insights from EDA ,

Customers group report consisting of detailed interpretation , characteristics ,conclusions, and recommendations.