**Industrial Standard Documentation**

**Project Charter:**

* **Project Title:** Mall-Customer Segmentation for a Retail Store
* **Project Manager:** [Akancha Kumari Sah]
* **Start Date:** [13-07-2024]
* **End Date:** [17-07-2024]

**Introduction**

**Objective and Use Case:**

**Objective:** The primary objective of this project is to analyze customer data from a retail store and segment customers into distinct groups based on their purchasing behavior. By identifying these segments, the retail store can tailor its marketing strategies to better meet the needs of each customer group, ultimately enhancing customer satisfaction and boosting sales.

**Use Case:** Customer segmentation is a crucial aspect of customer relationship management (CRM) and marketing strategies. In the context of a retail store, understanding different customer segments allows the store to:

* **Develop Targeted Marketing Campaigns:** Tailor promotions and advertisements to specific customer groups based on their purchasing habits and preferences.
* **Personalize Customer Experiences:** Offer personalized recommendations and services to improve customer satisfaction and loyalty.
* **Optimize Product Offerings:** Adjust inventory and product offerings to align with the preferences of different customer segments.
* **Increase Customer Retention:** Implement strategies to retain high-value customers and reduce churn rates.
* **Enhance Sales and Revenue:** Identify opportunities for cross-selling and up-selling to maximize sales and revenue.

By leveraging customer segmentation, the retail store can implement more effective marketing strategies, improve operational efficiency, and ultimately achieve a competitive advantage in the market.

**Overview of the Datasets:**

**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).

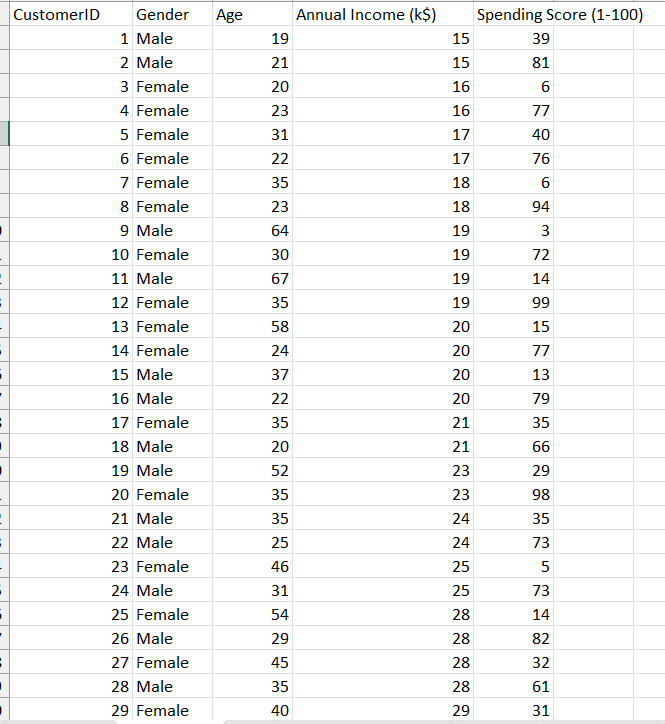
**Importing the Dataset:**

* The dataset is imported into the analysis environment (e.g., Jupyter Notebook) using Python libraries like Pandas.

**Brief Overview of the Dataset:**

* An initial examination of the dataset is performed to understand the structure and contents, including the number of records, columns, and data types.

**Datasets Snapshot**



**3. Data Cleaning**

**Handling Missing Values:**

* Check for any missing values in the dataset and handle them appropriately, either by imputing them or removing the affected records.

**Data Transformation:**

* Transform the data if necessary, such as converting categorical data (e.g., Gender) into numerical format using encoding techniques.

**Handling Outliers:**

* Identify and handle outliers that may affect the analysis. This can be done using statistical methods or visual inspection.

**4. Exploratory Data Analysis (EDA)**

**Descriptive Statistics:**

* Calculate and display descriptive statistics such as mean, median, mode, standard deviation, and range for numerical columns.

**Visualizing Distributions and Relationships Using Matplotlib:**

* Use Matplotlib to create visualizations like histograms, box plots, scatter plots, and bar charts to understand the distribution of individual variables and relationships between them.

**Insights from Visualizations:**

* Extract insights from the visualizations, such as identifying trends, patterns, and anomalies in the data.

**5. Customer Segmentation**

**Feature Selection:**

* Select the relevant features (Age, Annual Income, Spending Score) for clustering. These features will be used to identify distinct customer segments.

**Using K-Means Clustering for Segmentation:**

* Apply K-Means clustering algorithm to segment customers into distinct groups. Determine the optimal number of clusters using methods like the Elbow Method.

**Evaluating Cluster Quality:**

* Evaluate the quality of the clusters using metrics such as silhouette score and visualize the clusters to ensure meaningful segmentation.

**6. Visualization with Matplotlib**

**Visualizing Clusters:**

* Visualize the clusters using Matplotlib to understand the distribution of customers in each cluster. Create scatter plots or other relevant plots to show the clusters.

**Detailed Analysis Using Various Plots:**

* Perform a detailed analysis of each cluster using various plots to understand the characteristics of customers in each segment.

**7. Visualization with Power BI**

**Importing Data to Power BI:**

* Import the cleaned and segmented data into Power BI for creating interactive dashboards.

**Creating Interactive Dashboards:**

* Create interactive dashboards in Power BI to visualize the customer segments, with filters and drill-down capabilities to explore the data.

**Sharing Insights:**

* Share the insights derived from the analysis and segmentation through the Power BI dashboards with stakeholders for actionable decisions.

**8. Conclusion**

**Summary of Findings:**

* Summarize the key findings from the customer segmentation analysis, highlighting the distinct customer groups identified.

**Recommendations and Next Steps:**

* Provide recommendations based on the findings, such as targeted marketing strategies for each customer segment. Outline the next steps for implementing these strategies and potential areas for further analysis.