Executive Summary

Title: K-Means Clustering of Wholesale Customers Dataset

Objective: The primary purpose of the project was to segment the dataset using K-Means clustering algorithm in python on PyCharm. The dataset is a wholesale customer dataset which was segmented based on their spending habits on various products. The segmentation provided insights on the customers’ purchasing power which will help the business roll out specific marketing strategies.

Dataset Description: The wholesale customer dataset was sourced from UCI machine learning directory. The dataset contained features such as Fresh, Milk, Grocery, Frozen, Detergents\_Paper, and Delicassen. Additional categorical features (Channel and Region) were excluded to focus on purely numerical clustering.

Methodology:

* The data was pre-processed by removing non-numerical features, removing outliers and scaling the variables (bringing all the numerical values to a common scale) using Standard Scaler.
* To use K-means to segment into data into clusters, an optimal number of clusters must be determined, and this is done using the **Elbow method.**
* K-Means clustering was then performed with 3 clusters (which was determined using the elbow method), and the cluster labels were added to the dataset for analysis.
* The dimension of the dataset was reduced to 2-D using PCA (principal component analysis) for easier visualization of the clusters.

Key Findings/Insight: Visualization revealed distinct clusters validating the effectiveness of the clustering approach. The clusters were segmented into:

* **Cluster 0**: High-Spending Generalists (e.g., Large Supermarkets, Large Retailers): High spending on Grocery, Milk, and Detergents\_Paper
* **Cluster 1**: Fresh-Focused Customers (e.g., Restaurants or Fresh Markets): High spending on Fresh
* **Cluster 2**: Grocery and Detergents Specialists (e.g., Convenience Stores): Low spending overall, slightly higher on Delicassen

Conclusion:

The K-Means clustering technique effectively categorized wholesale customers into distinct, insightful segments. These customer groupings provide valuable opportunities for businesses to implement focused marketing strategies, personalized loyalty initiatives, and tailored pricing structures. This project highlights how unsupervised machine learning can reveal underlying trends in commercial datasets and enhance data-informed strategic planning.