# UNSUPERVISED LEARNING

George Edgar

PROJECT GOAL

Apply unsupervised learning techniques to a real-world data set and use data visualization tools to communicate the insights gained from the analysis.

The data set for this project is the "Wholesale Data" dataset containing information about various products sold by a grocery store. The project will involve the following tasks:

•Exploratory data analysis and pre-processing: We will import and clean the data sets, analyze and visualize the relationships between the different variables, handle missing values and outliers, and perform feature engineering as needed.

•Unsupervised learning: We will use the Wholesale Data dataset to perform k-means clustering, hierarchical clustering, and principal component analysis (PCA) to identify patterns and group similar data points together. We will determine the optimal number of clusters and communicate the insights gained through data visualization.

The ultimate goal of the project is to gain insights from the data sets and communicate these insights to stakeholders using appropriate visualizations and metrics to make informed decisions based on the business questions asked.

HIERARCHICAL CLUSTERING
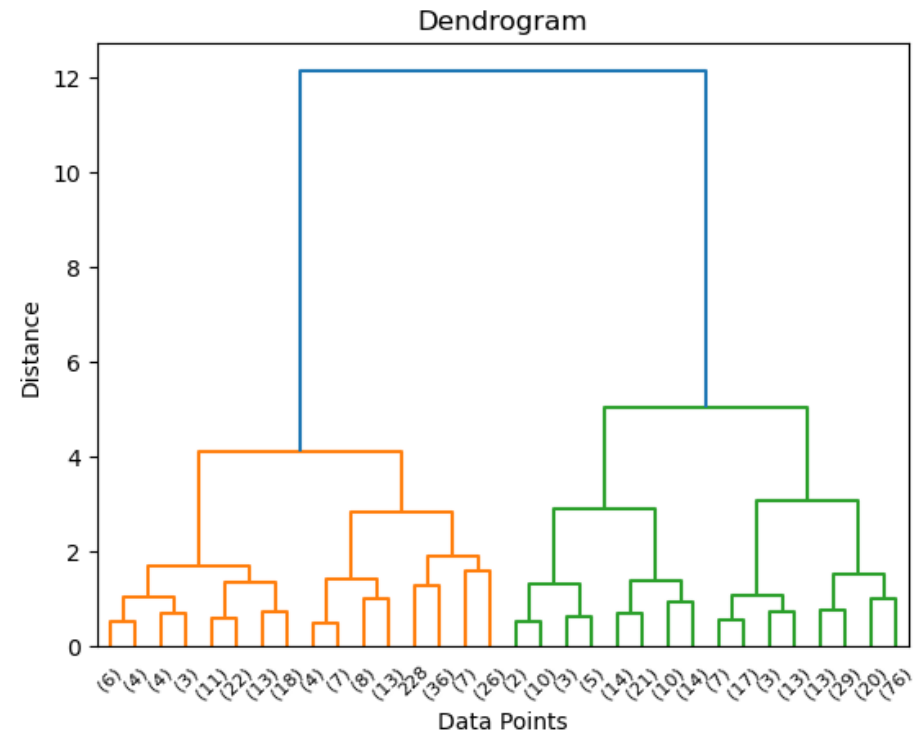
PCA

- Applied PCA to the data to find combinations of features that best describe customers. Three principal components were selected based on the explained variance ratio.

- Identified the feature weights of the first three principal components, which helped in understanding the spending habits of customers across different product categories.

CONCLUSION

Customers with higher values in the first principal component (PC1) show a strong preference for Detergents Paper, Grocery, and Milk, suggesting a focus on household and daily essential products.

The second principal component (PC2) reveals that customers with higher values tend to spend more on Delicatessen and Frozen products while having relatively lower spending on Fresh items, indicating a preference for convenience or ready-to-eat options.

The third principal component (PC3) highlights a preference for Delicatessen products and a lower spending on Fresh and Frozen items among certain customer segments, suggesting a focus on specialty products.