Unsupervised Learning Project:

Wholesale Data Analysis

Rodrigo Rezende

Goals:

- Perform exploratory data analysis on the "Wholesale Data" dataset to understand the structure of the data.
- Apply KMeans Clustering to group similar products into clusters based on their attributes.
- Apply Hierarchical Clustering to identify patterns and group similar data points together in a hierarchy.
- Use Principal Component Analysis to identify which combinations of features best describe customers.

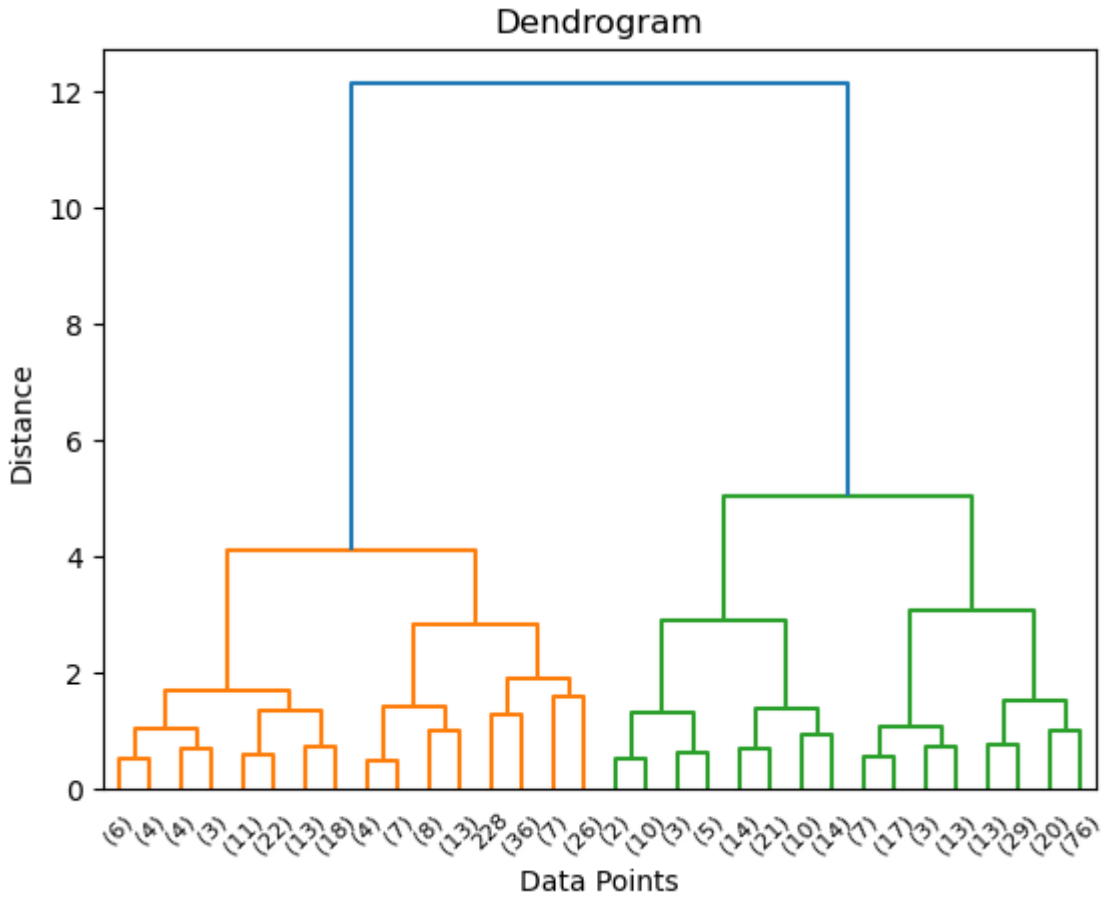
KMeans Clustering:

- 4 clusters were chosen based on the silhouette score and cohesion.

- Evaluated the clusters using the silhouette score and cohesion, which provided an effective measure of the quality of the clustering.

Hierarchical Clustering :

- Based on the dendrogram, 4 clusters were chosen because it provided a good balance between the number of clusters and the within-cluster sum of squares.



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 PC1 values prefer household and daily essential products like Detergents Paper, Grocery, and Milk.

- Customers with higher PC2 values prioritize Delicatessen and Frozen products, showing a preference for convenience or ready-to-eat options over Fresh items.

- Certain customer segments, characterized by higher PC3 values, exhibit a preference for Delicatessen products and relatively lower spending on Fresh and Frozen items, indicating a focus on specialty products.