

The background is a dark gray textured surface. It features several large, semi-transparent spheres in red and green. In the top left, there are two thin white lines forming a stylized 'Z' or 'N' shape. In the bottom right, there is a thin white 'X' shape.

# Wholesale Data

## **Unsupervised Learning Machine**

By Hajar Kaddouri



# Structure

1. **Introduction**
2. **EDA**
3. **Preprocessing**
4. **hierarchical clustering,**
5. **Conclusion**



# Introductionnn

- **The aim of the study : is to understand the patterns of a wholesale store by implementing the clustering methods and PCA.**

N







# **Part I - EDA**

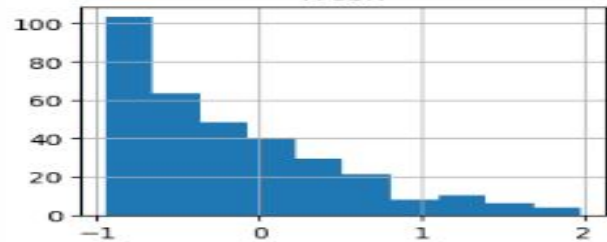


# EDA & Data Preprocessing:

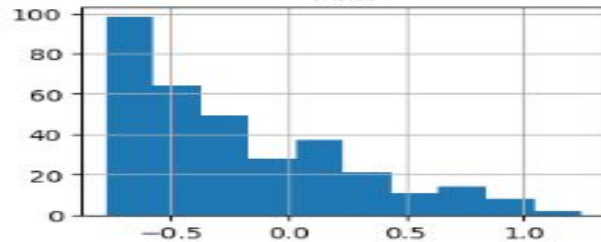
Data Understanding and Preprocessing: The dataset, comprising 'Fresh', 'Milk', 'Grocery', 'Frozen', 'Detergents\_Paper', and 'Delicatessen', underwent thorough data cleaning, missing value handling, outlier treatment, and normalization using standard scaler.



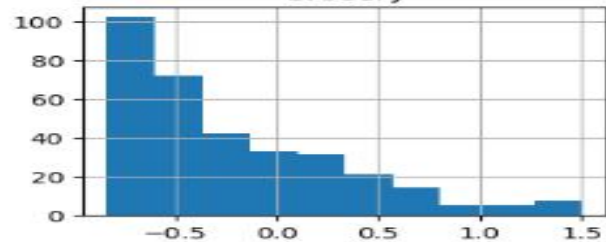
Fresh



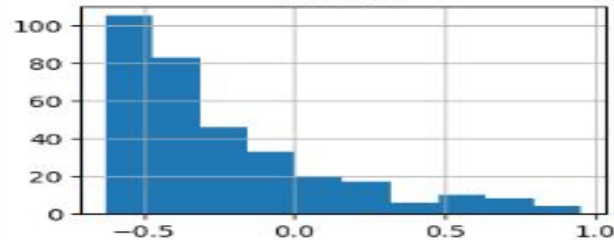
Milk



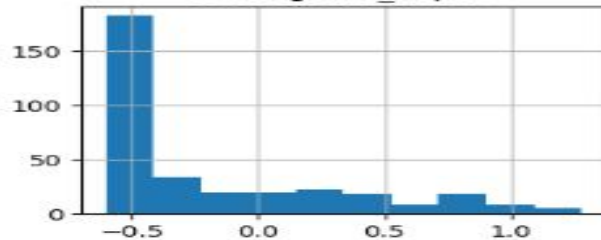
Grocery



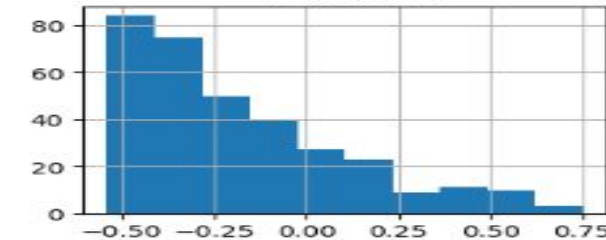
Frozen



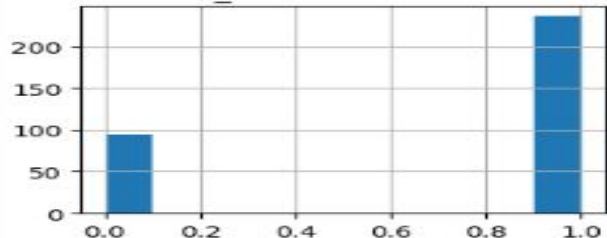
Detergents\_Paper



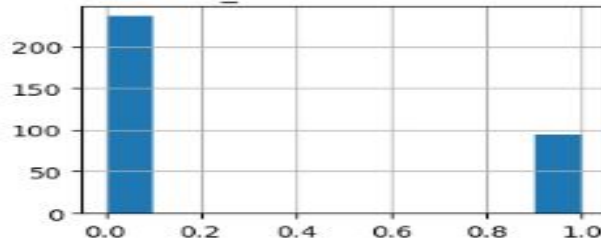
Delicassen



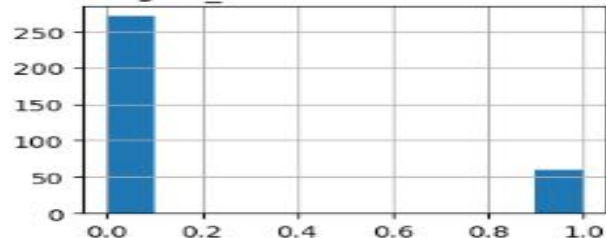
Channel\_-0.6902970861260346



Channel\_1.4486516314475937



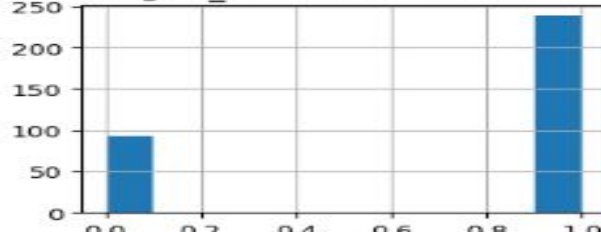
Region\_-1.9953421178949025

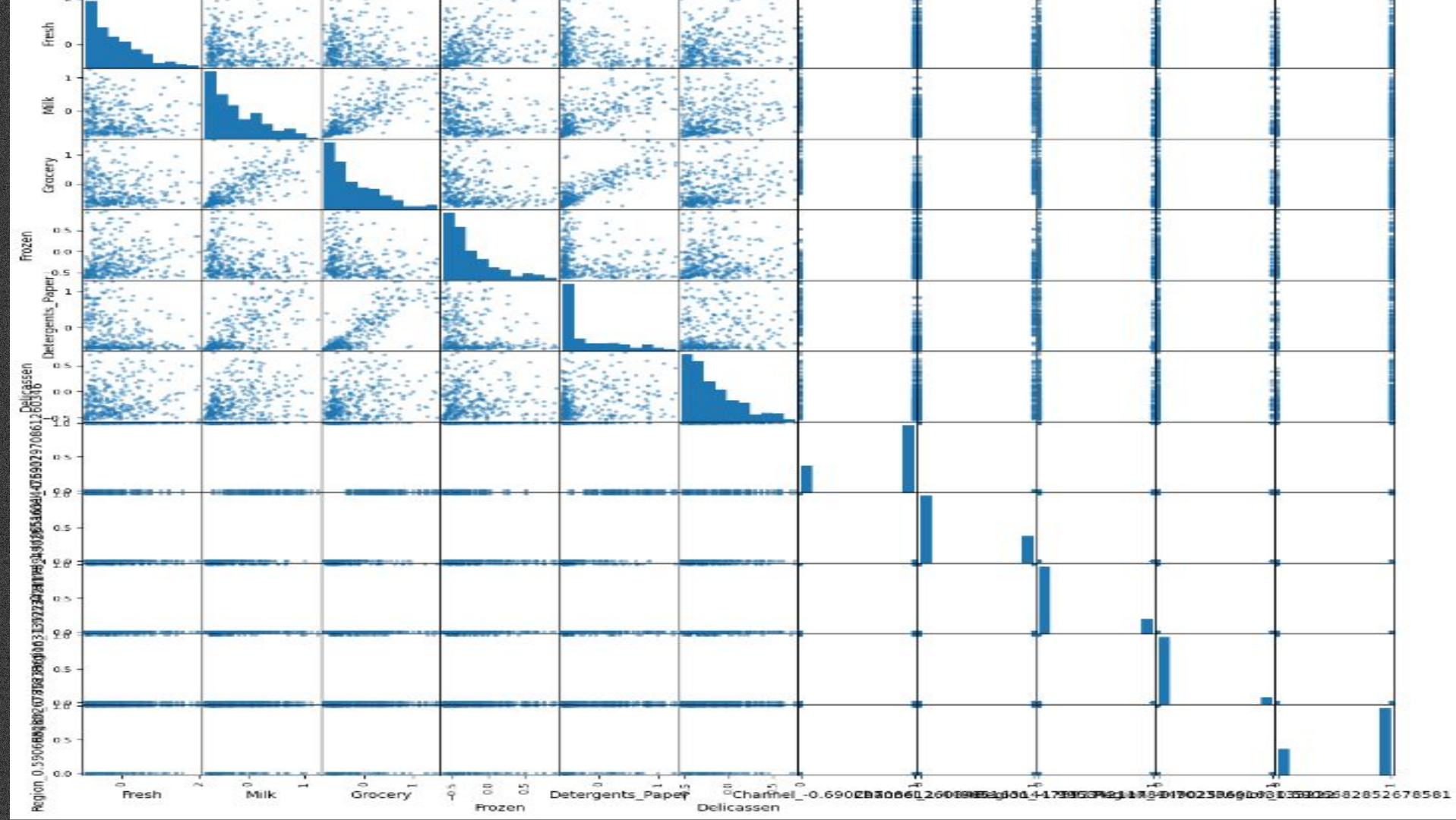


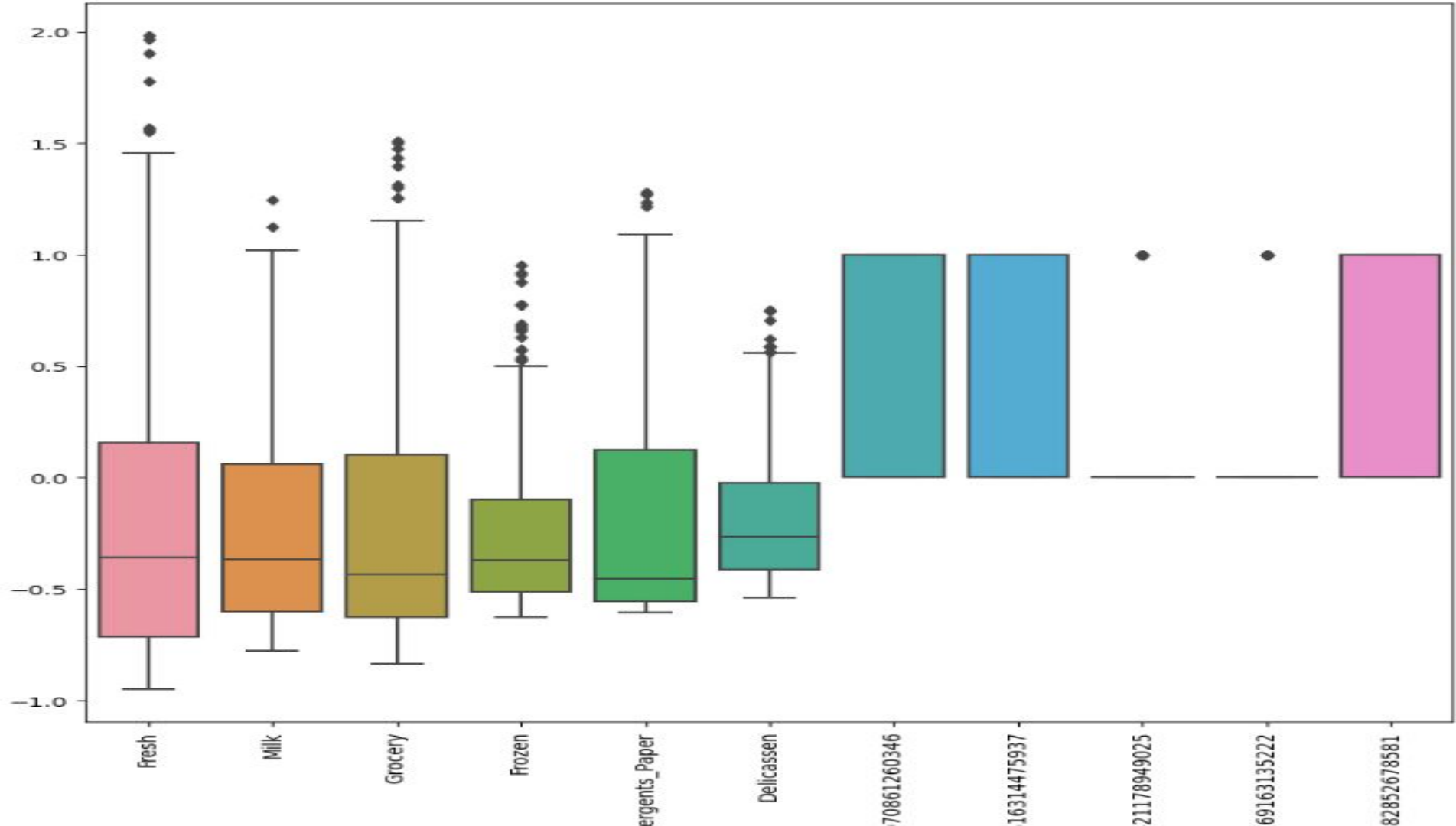
Region\_-0.7023369163135222



Region\_0.5906682852678581







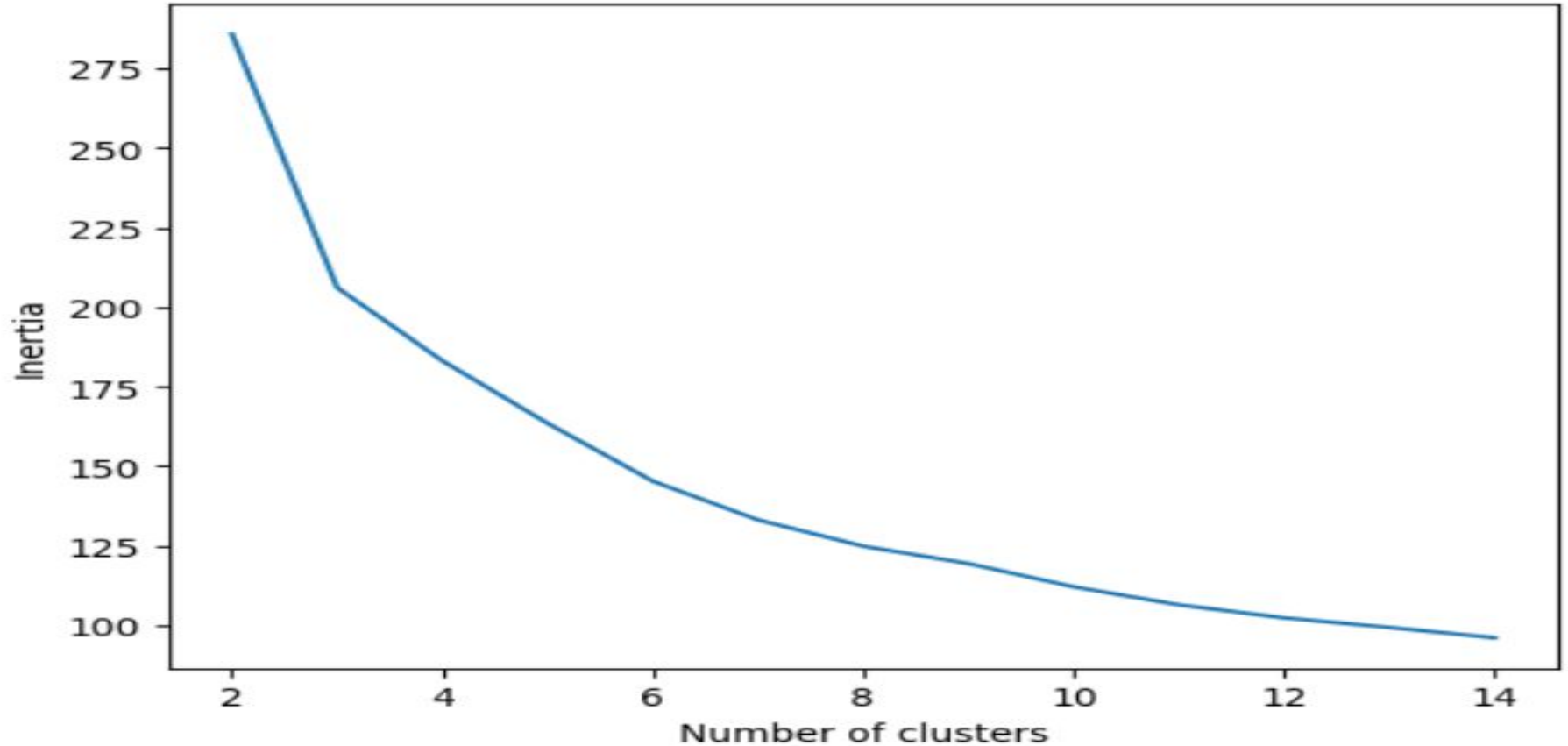




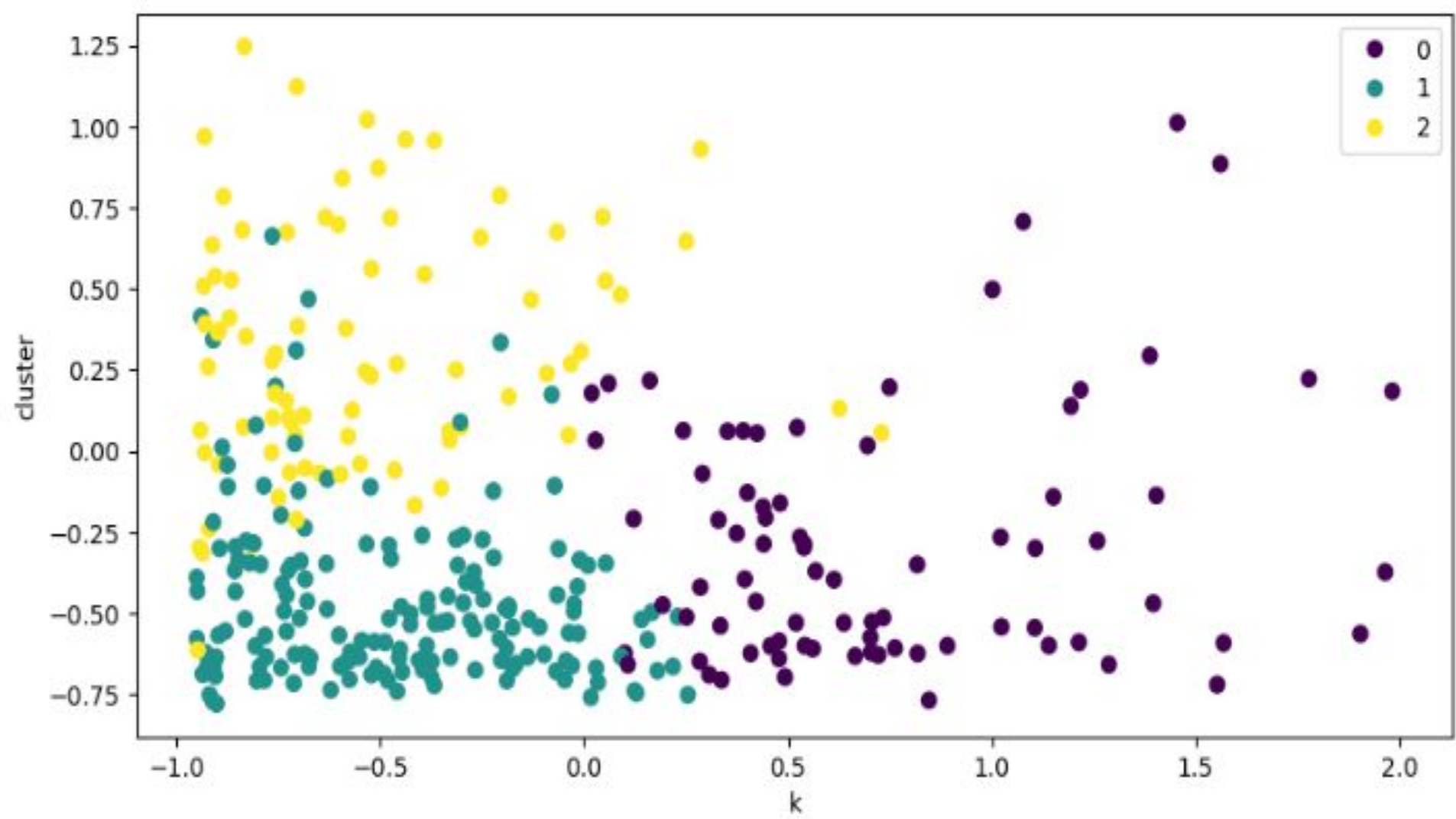
# **Part II - KMeans Clustering**



In the k-means clustering process, the optimal number of clusters was determined using the elbow method,











### Insights from K-means:

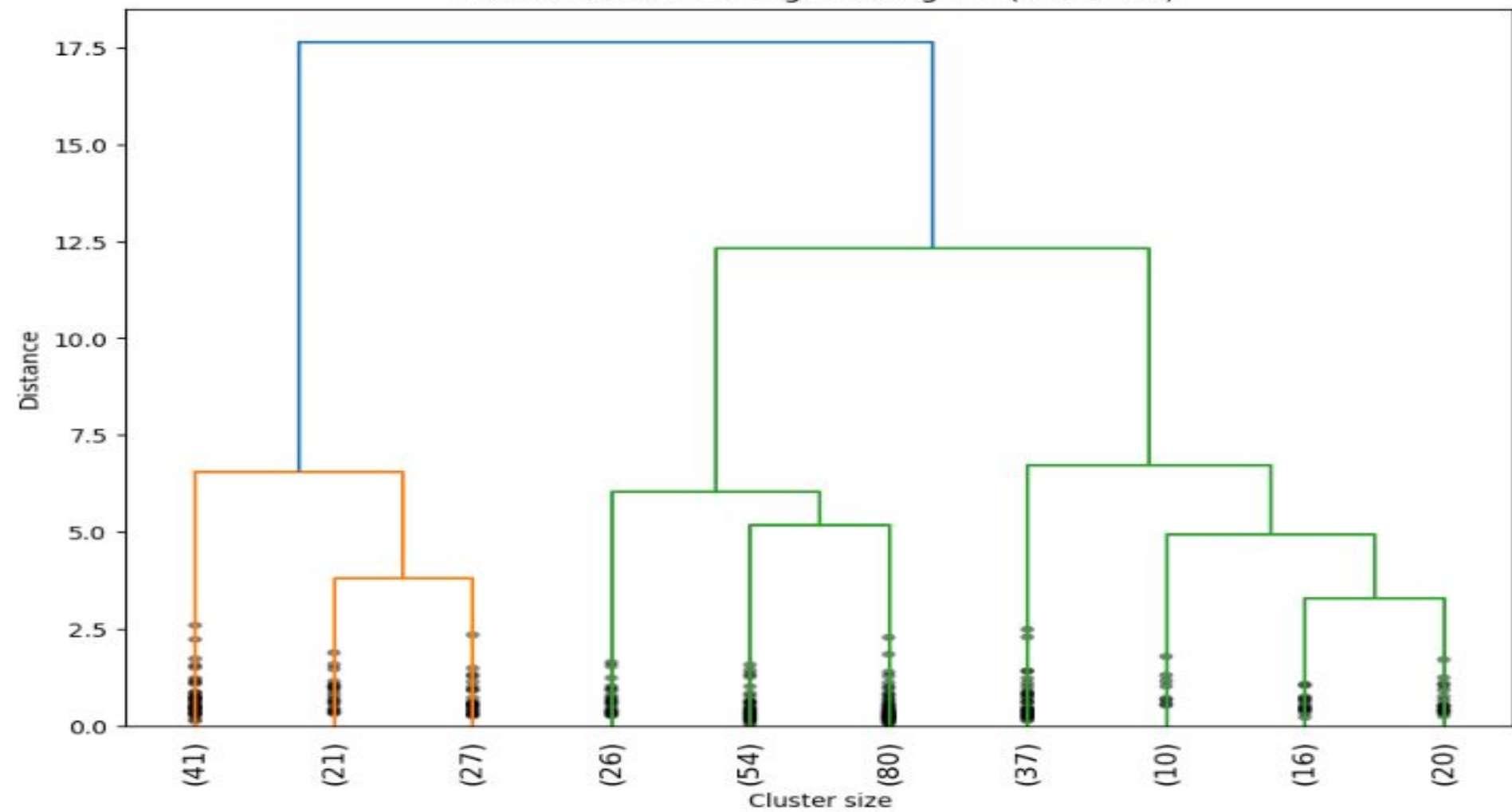
- Customers buying more "Fresh" and "Frozen" goods are grouped together indicating they might represent businesses like restaurants or cafes which require fresh food to serve customers.
- Customers who purchase more "Grocery", "Milk", and "Detergents\_Paper" products are in another cluster, possibly indicating grocery stores or supermarkets.
- The analysis also suggested that "Fresh" and "Frozen" products are often purchased together, and likewise for "Grocery", "Milk", and "Detergents\_Paper", it can be for individual usage





# **Part III - Hierarchical Clustering**

Hierarchical Clustering Dendrogram (truncated)







### Insights from Hierarchical Relationships:

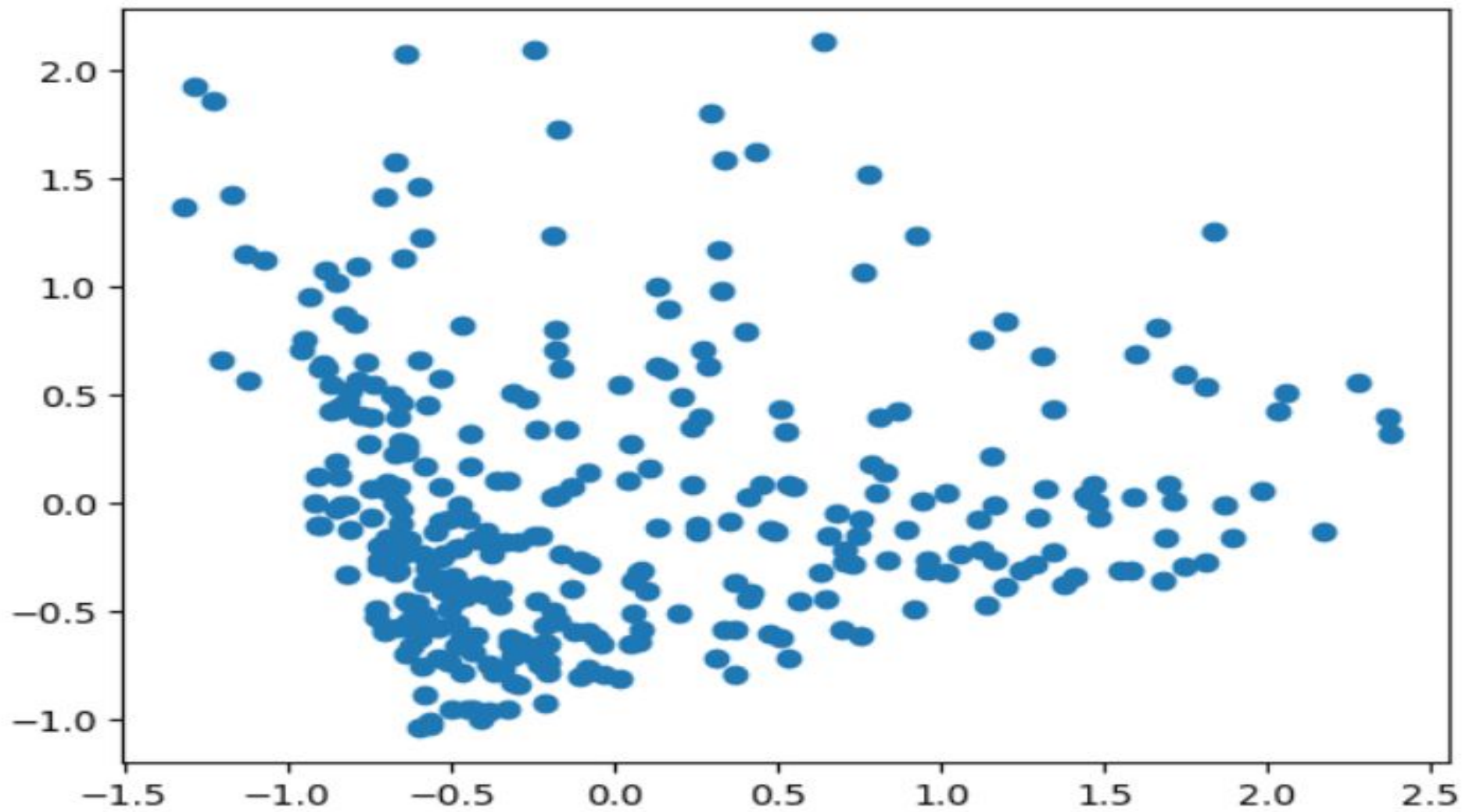
- The dendrogram resulting from hierarchical clustering can reveal relationships between clusters. For instance, two closely related clusters might represent customers who purchase similar products but in different volumes or proportions.
- silhouette score of 0.342 for 3 clusters, suggests that the wholesale customers can be grouped into three moderately distinct clusters.





# **Part III - PCA**









### Insights from PCA:

- The first principal component was heavily influenced by "Detergents\_Paper", "Grocery", and "Milk" indicating that these three categories vary together. If one increases, then the remaining ones tend to as well.
- The second principal component increased with increasing "Fresh", "Frozen", and "Delicatessen", this suggests that these three categories vary together, and they might be a measure of whether the customers are restaurants or not.



# Conclusion

Insights were drawn from both clustering methods and PCA, which can help understand customer segmentation and purchase patterns better. The relationships among different product categories and the identified customer types can serve as a foundation for devising personalized marketing strategies, targeted promotional offers, and optimizing the supply chain management.





**THANKS**



# RESOURCES

- <https://github.com/hajar-kaddouri/ml-project-unsupervised-learning>