Unsupervised Learning

Wholesale market data

By: Jacob Cyr

Goals and intentions

- -The goal is to use unsupervised clustering techniques to do classification on sales data.
- -The methods used are Exploratory Data Analysis and Object Oriented Pipelines.
- -Use of classification algorithms such as K –means and hierarchical Clustering

K means clustering

- -K means clustering algorithms found that the optimal number of clusters was 5.
- -The method chosen was the elbow method

Hierarchical Clustering

- -Although I was able to get the heirachial algorithms to work, the concept was black box for me.
- -I understand I was able to find dendrograms which use an agglomerate up function to specify clusters starting with individual points.

PCA

- -Principle component analysis was applied and the default value of 2 components was chosen for initial analysis.
- -this pipeline calculated the explained variance ratio.
- -The PCA helps visualize patterns in a reduced dimensional feature space.

Conclusion.

- Through the use of object-oriented pipelines, clustering techniques, and Principal Component Analysis (PCA), we gained valuable insights into the wholesale customer data.
- K-means clustering determined that 5 clusters were optimal for grouping similar customers based on their purchasing behavior.
- Hierarchical clustering helped identify patterns and relationships among the data points through dendrograms.
- PCA revealed the most important components of the data matrix, aiding in understanding the underlying structure.
- This analysis provides actionable information for targeted marketing, inventory management, and personalized customer experiences.