



Unsupervised Learning Project

By

Zhiyun(Jenny) Liang



Project Objective

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



Project Flow Structure

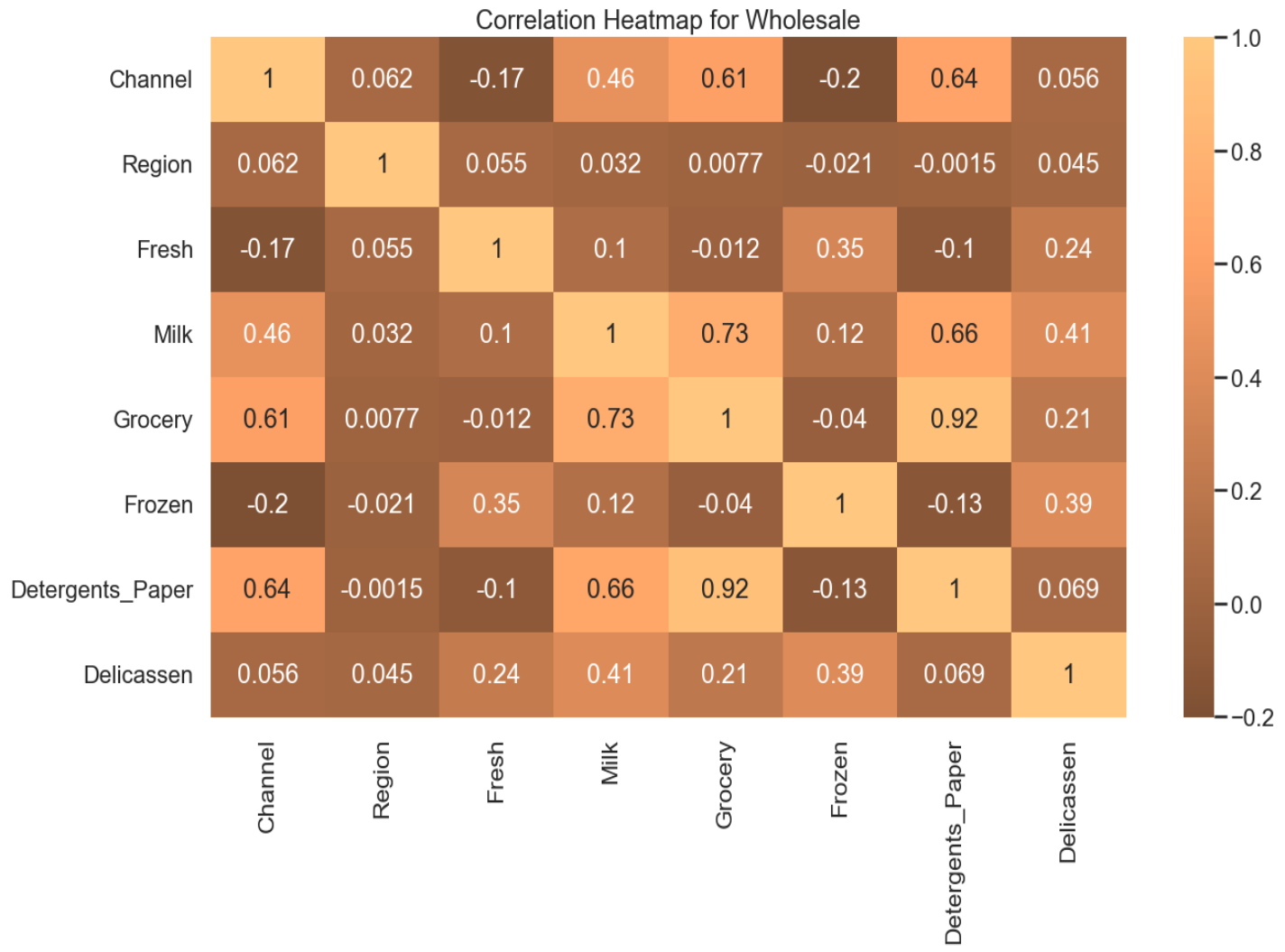
- Exploratory data analysis
- Preprocessing
- Kmeans clustering, hierarchical clustering
- PCA
- Results and Discussion

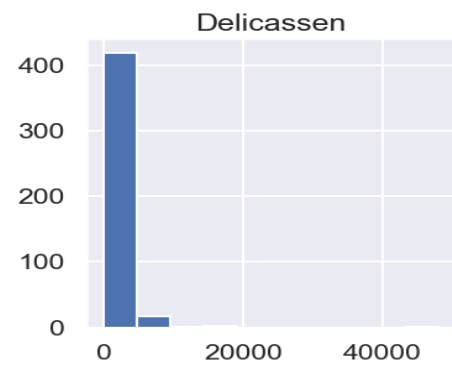
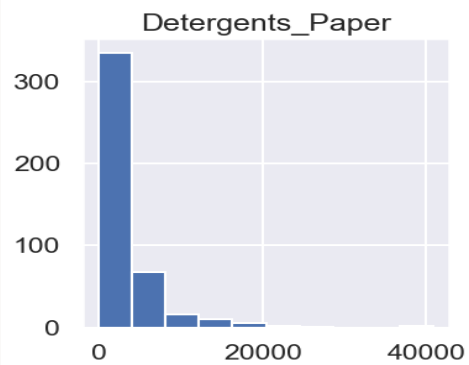
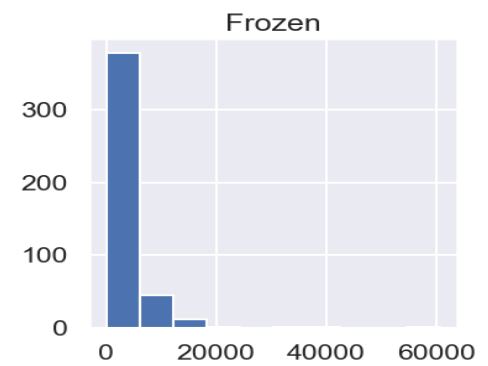
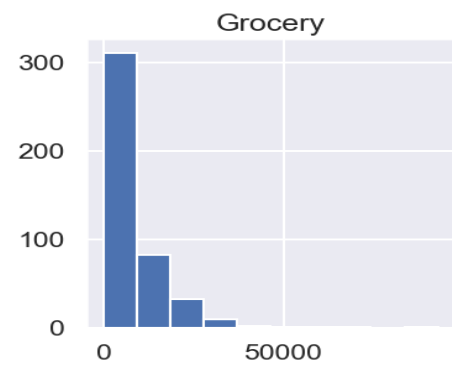
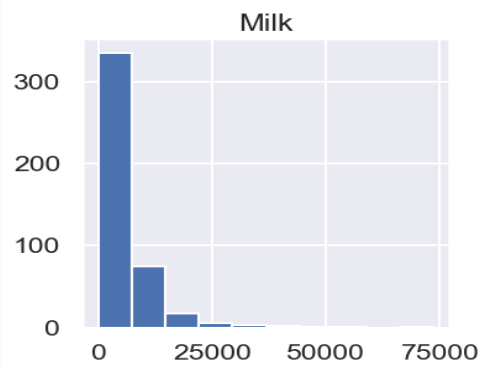
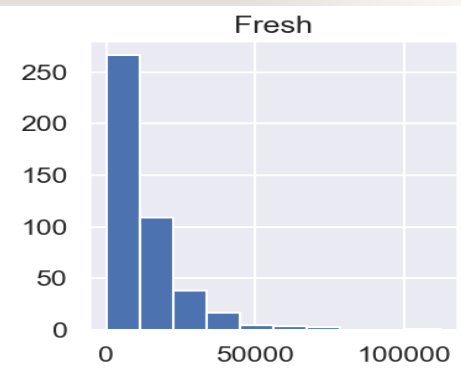
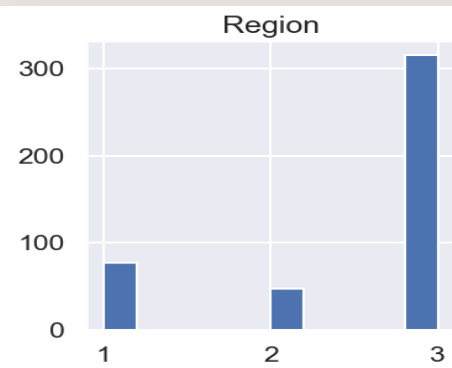
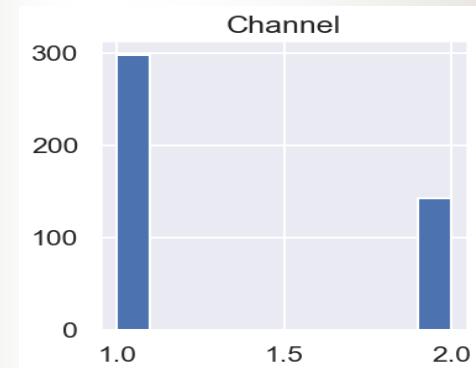


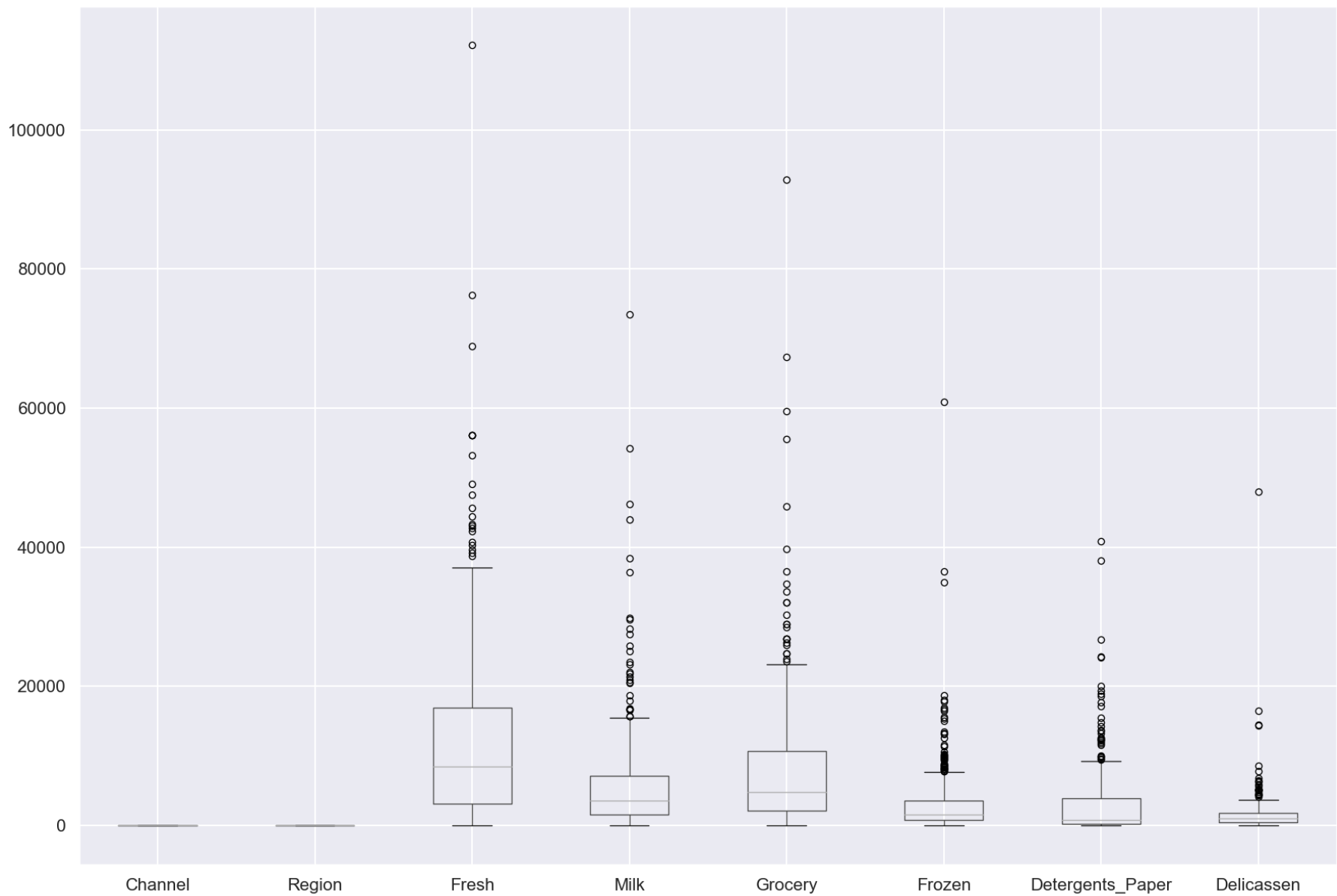
Exploratory Data Analysis

- Exploratory data (null value, outlier, etc.)
- Distribution of each variable
- Correlation between the variables

Heat Map







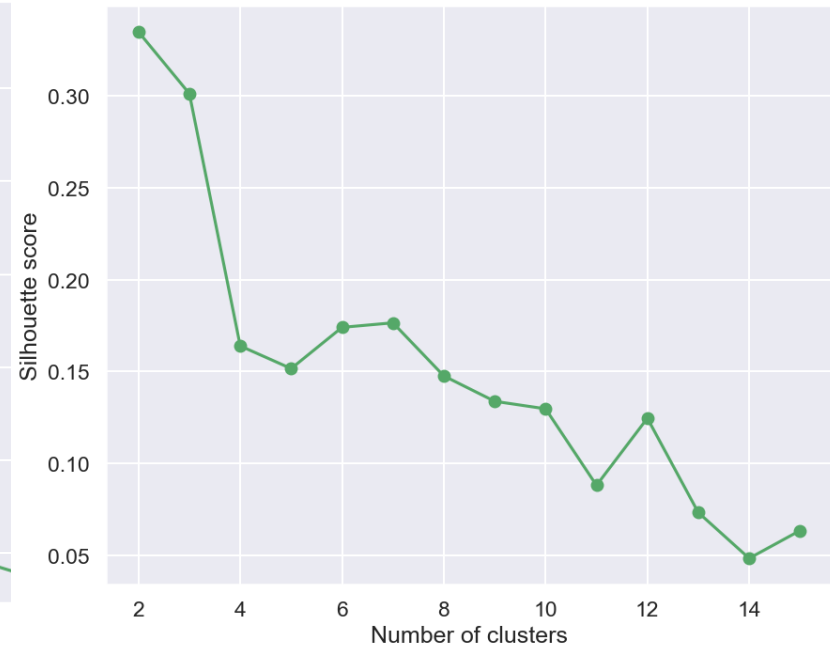
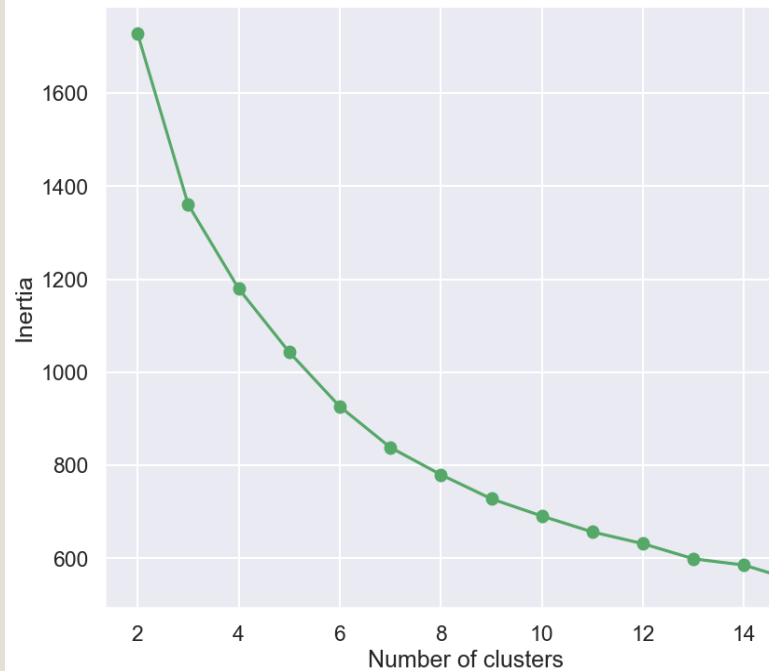


Preprocessing

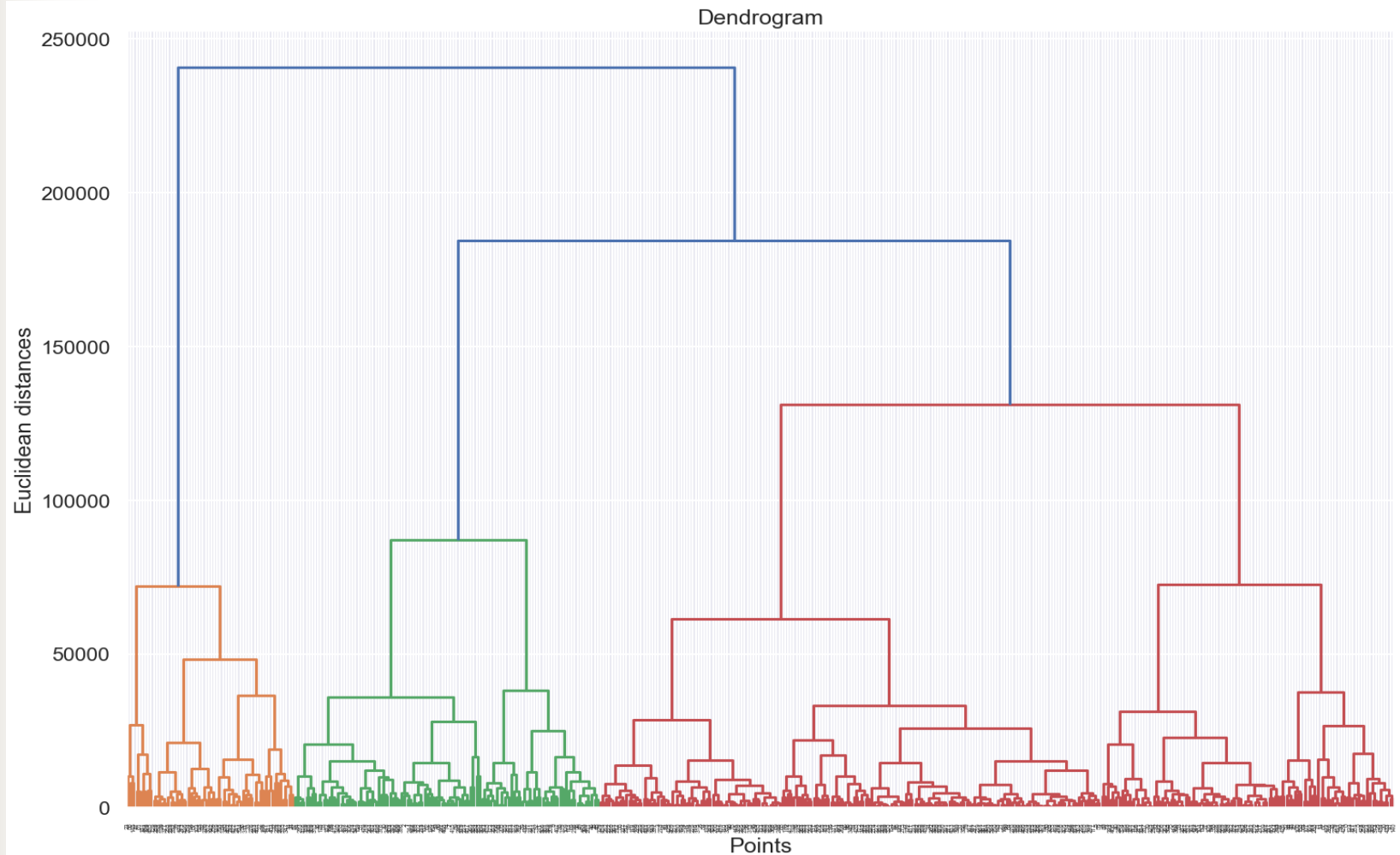
- Handling outliers
- Scaling and normalization variable

KMeans clustering

The Elbow Method

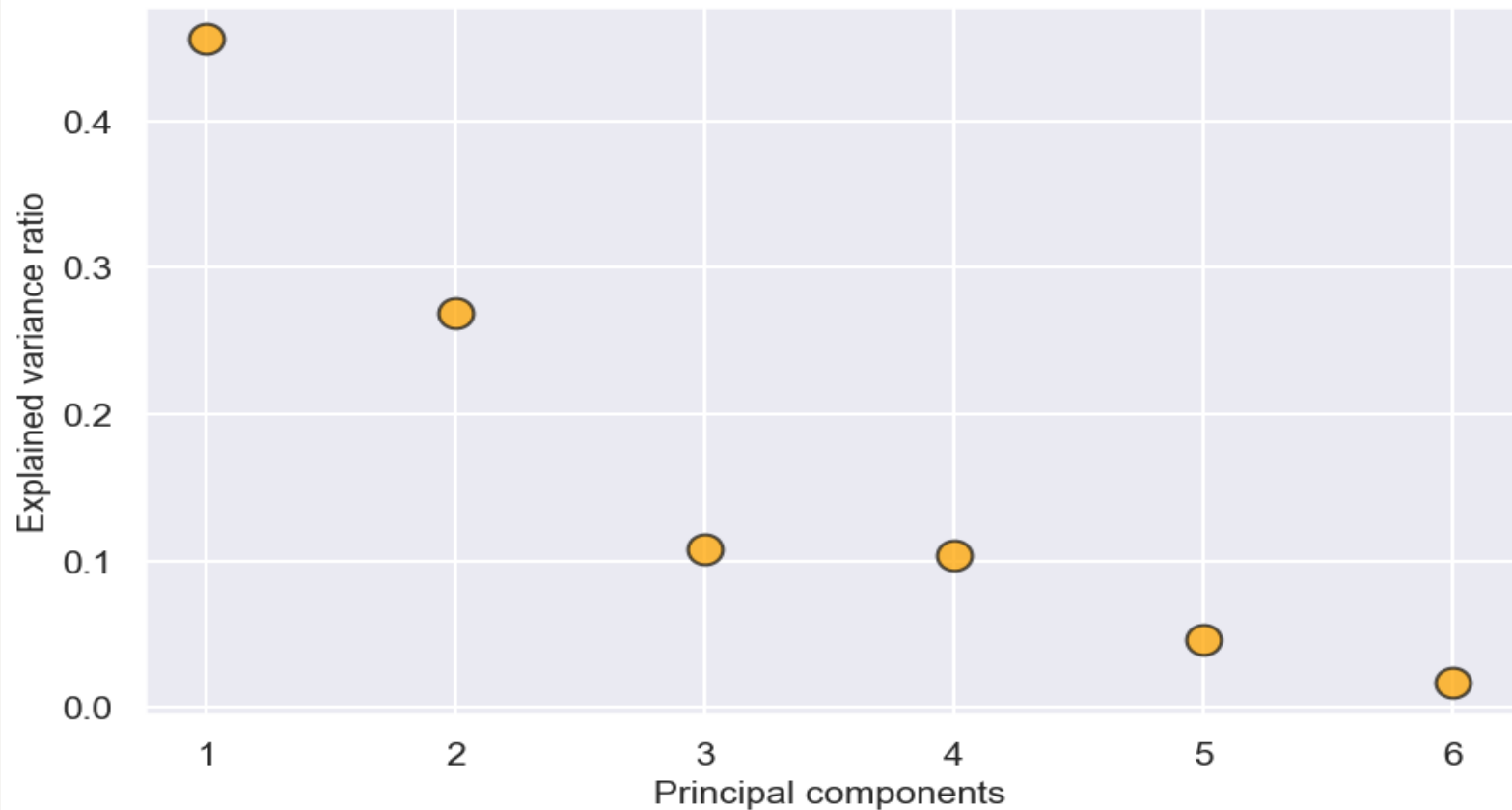


Hierarchical clustering



PCA

Explained variance ratio of the fitted principal component vector





Results and Discussion

- Dataset has 20-40 numbers of outlier for each column
- K-Mean clustering presents 3 or 7 clustering solution
- Hierarchical clustering presents 3 clustering solution
- PCA performs six variables reduction to three variables