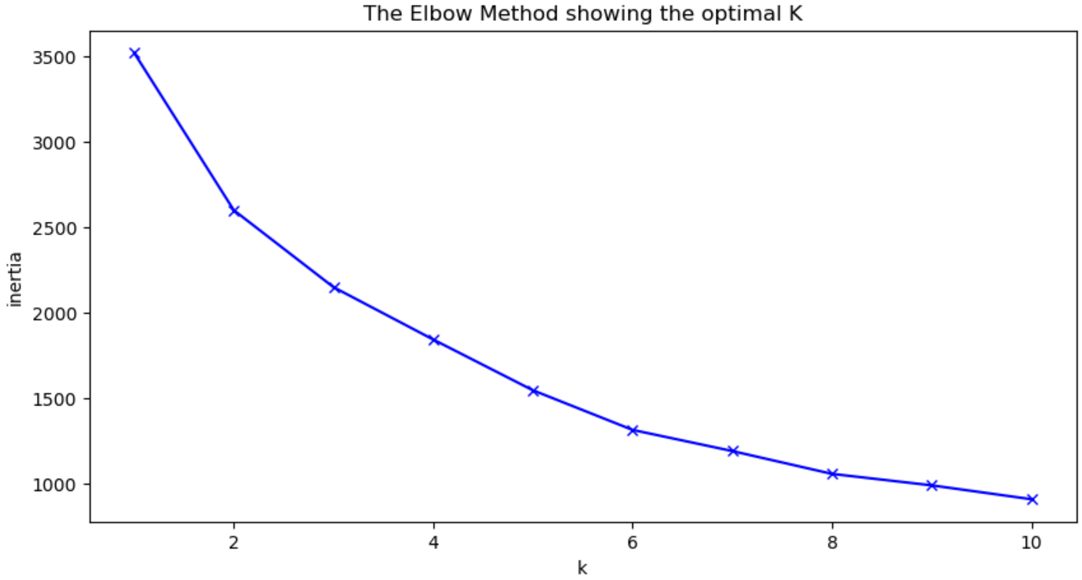
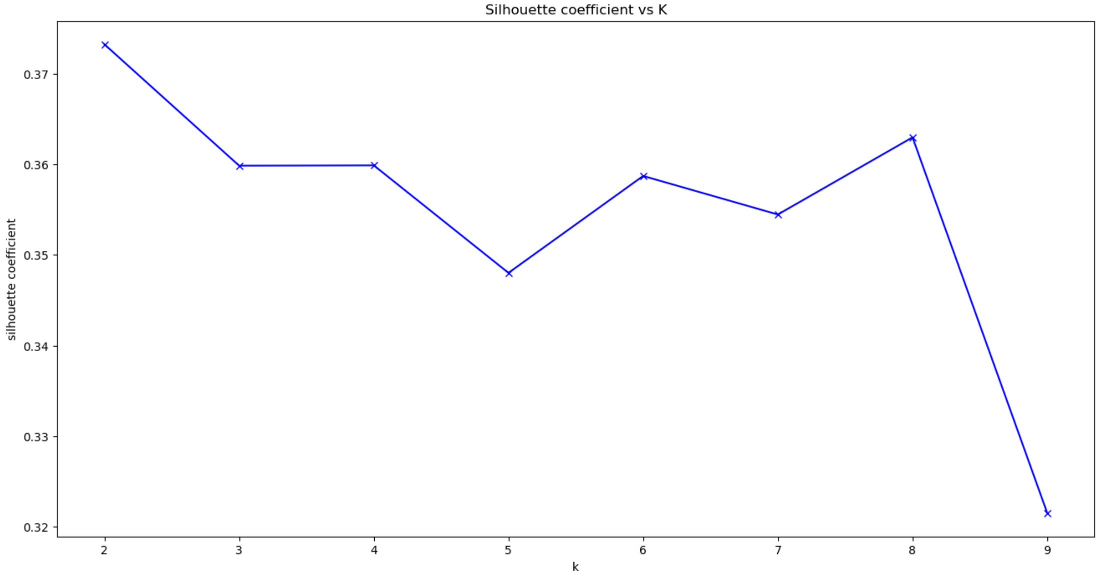
When asked to group similar data points together in unsupervised learning, the very first thing that pops in my head is K_means clustering.

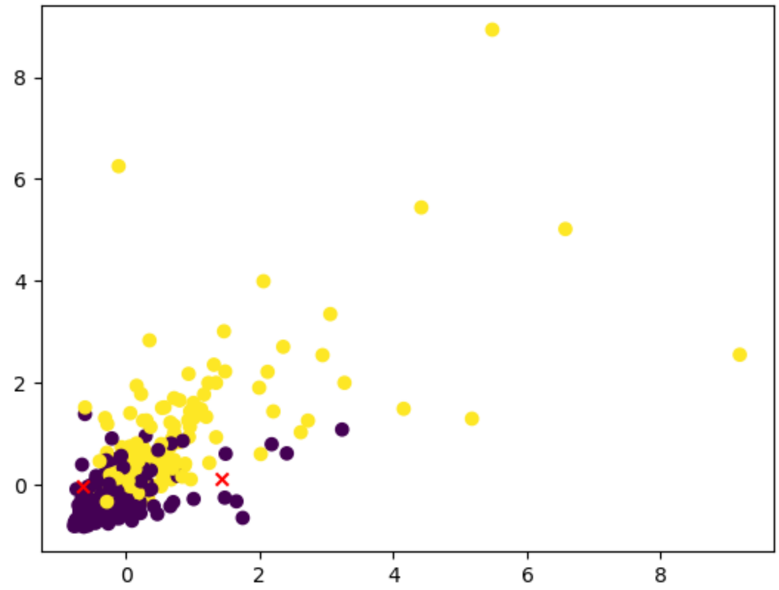
Before performing K_means_clustering we have to identity the optimal number of clusters. One way can identify this is by identifying the elbow in the inertia by # of clusters. The point where it performs an elbows is the optimal number of cluster.


The Elbow Method showing the optimal K

But the “elbow” isn’t easy to identify in the scenario, In my personal opinion an elbow forms at k equals 2,3 and 6. Another method to identifying the optimal number of clusters is figuring out the K that has the highest Silhouette Coefficient.


Silhouette coefficient vs K

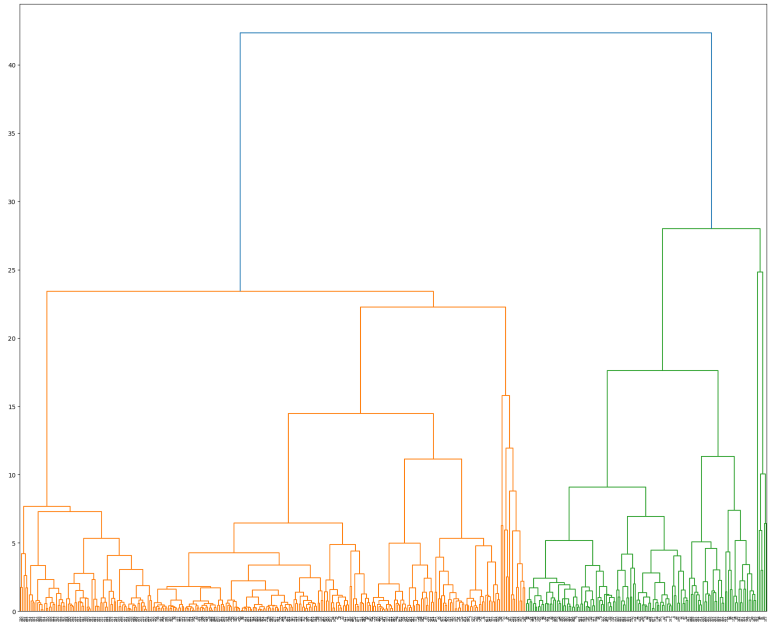
From the Silhouette Coefficient, we can clearly identify that there should be two clusters. Performing k means clustering on 2 clusters, we get this plot.



The centroids in the plot is the result of initializing the centroids, assigning data points to clusters, updating the centroids, and repeating until convergence. The Silhouette score for the k means clustering is also noted below.

Silhouette score: 0.37323636511581165

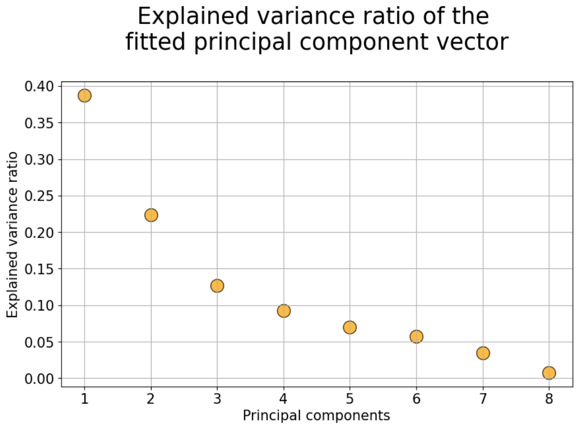
Using a dendrogram to find the optimal number of cluster for Hierarchical Clustering, we can clearing identify that there are two.

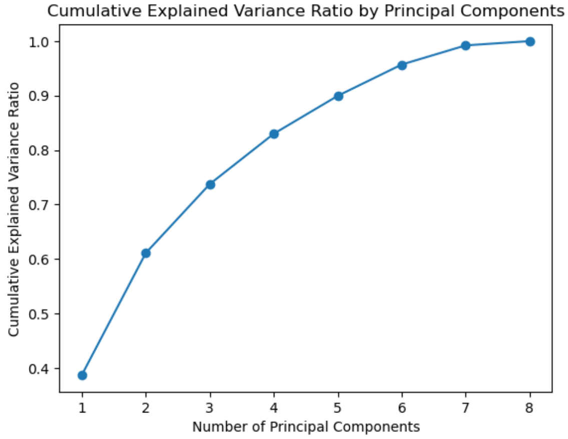


Using 2 clusters for Agglomerative Clustering we get Silhouette scores of:

Silhouette Scores:
Single: 0.8263511877226627
Complete: 0.8263511877226627
Average: 0.8263511877226627
Ward: 0.8263511877226627

Using PCA (principal component analysis) we get only 2 PCA that can explain more than 15% on the variance.





To summarize what we have learned
Agglomerative Clustering performs better than K means clustering
There are 2 principal components the can explain the variance in the data more than 15%
The Silhouette scores for the 4 different types of linkages are equivalent
PCA #1 is heavily influenced by Grocery, Detergents Paper, Milk, and Channel, while
PCA#2 heavily influenced by Frozen, Delicassen, Fresh