Machine Learning

"Customer Segmentation Model"

Presented by Tina Pham

Table of Content

1 — 2 — 3 — 4

Project Goal Process Results Future Goal
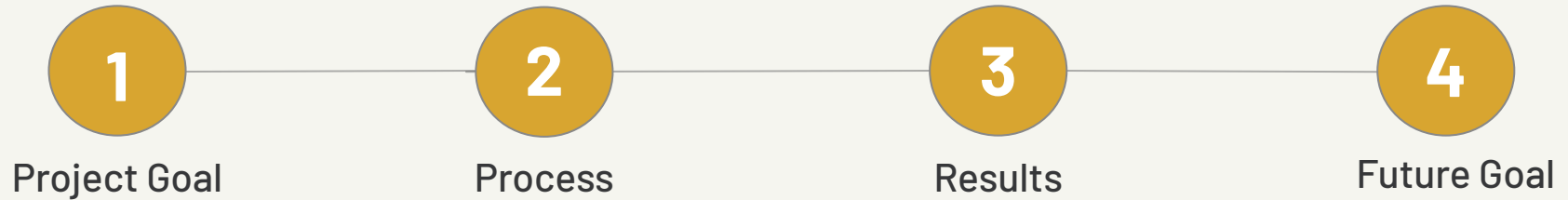
1. Project Goal

" To analyze shoppers' behavior by creating clusters of similar products base on their attributes"
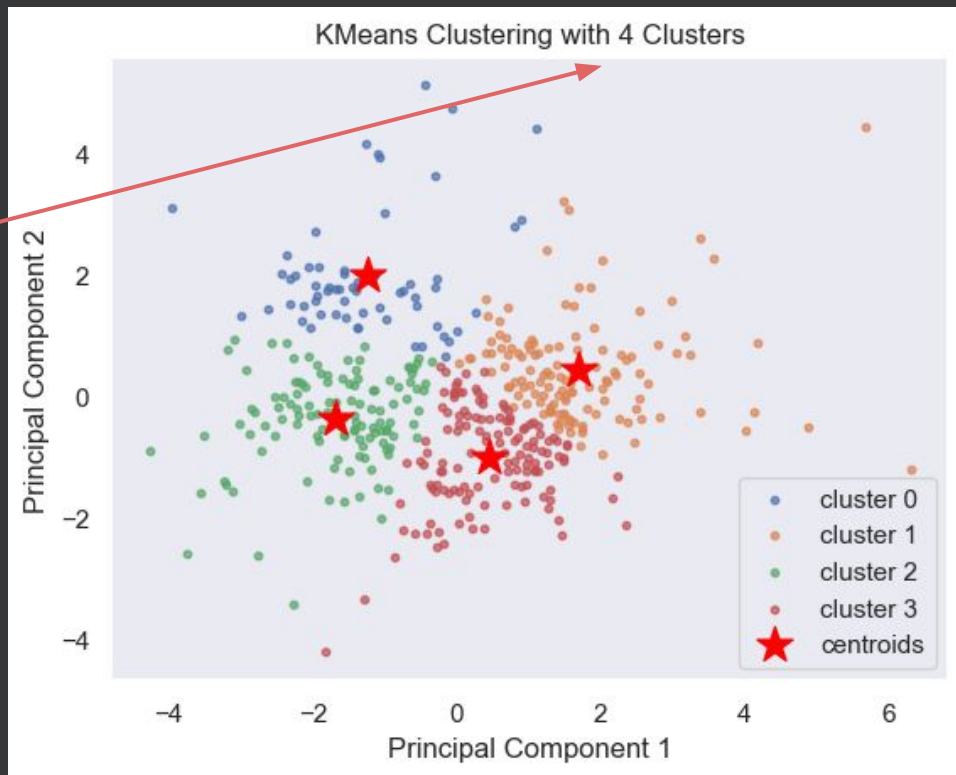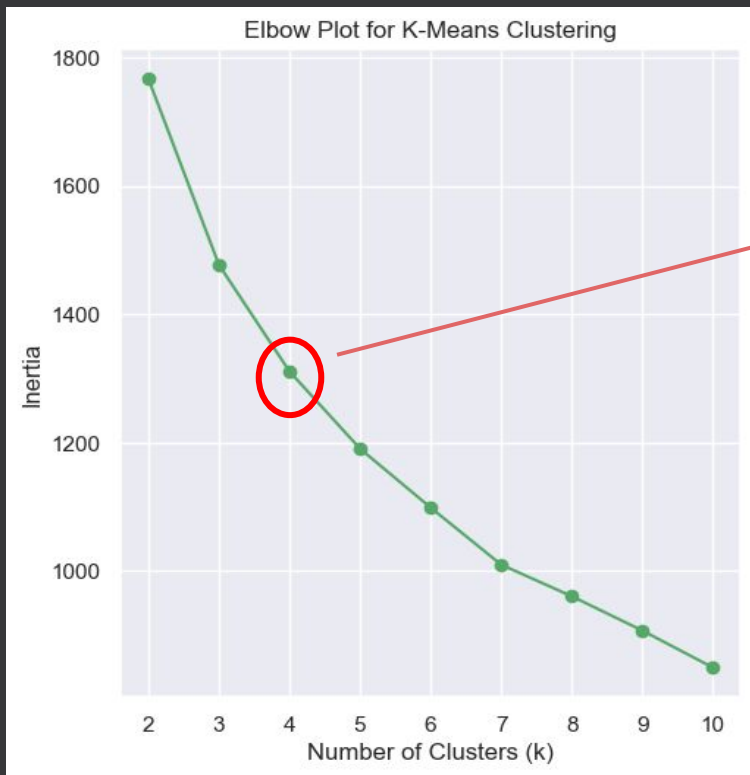
2. Process

- "Wholesale Data" dataset
- The project involves four main parts:
  1. Exploratory Data Analysis and and Preprocessing
  2. KMeans clustering
  3. Hierarchical clustering
  4. Principal components Analysis

3. Result

KMeans Cluster
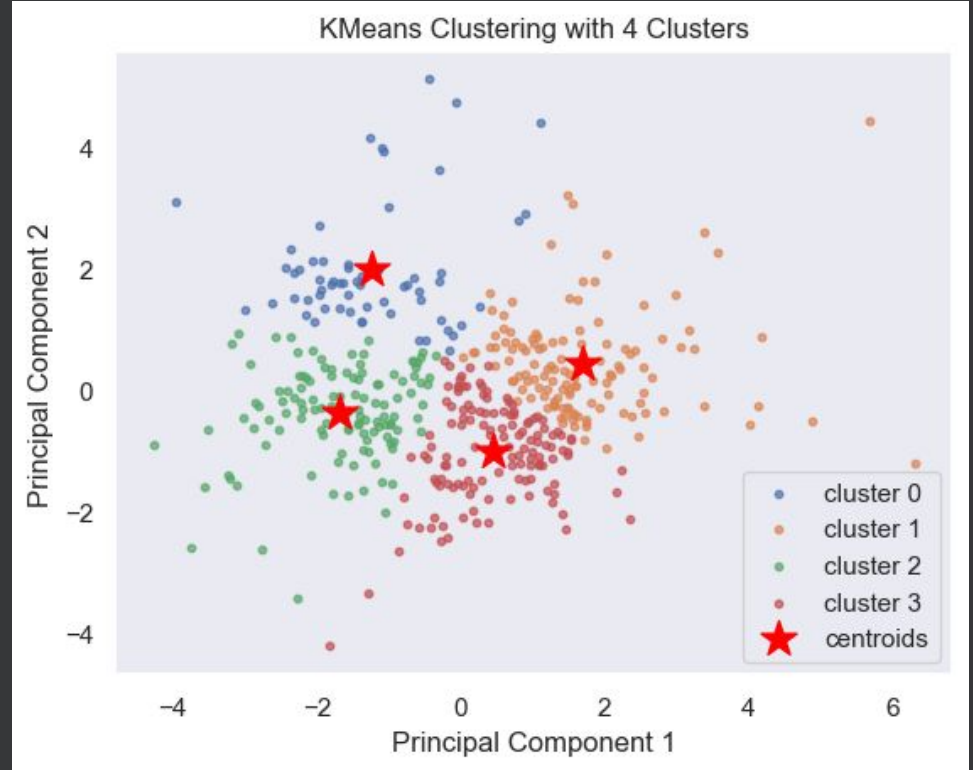

Elbow Plot for K-Means Clustering

KMeans Clustering with 4 Clusters

KMeans Cluster

- KMeans produce more distinct clusters.
- The optimal number of clusters obtained from elbow rule is 4.

→ Four distinct group of customer base on their shopping pattern.

KMeans Clustering with 4 Clusters

Hierarchical Cluster

- The optimal number of clusters obtained from elbow rule is 3.

 Not a distinct grouping of customer base on their shopping pattern.

Hierarchical Clustering with 3 Clusters
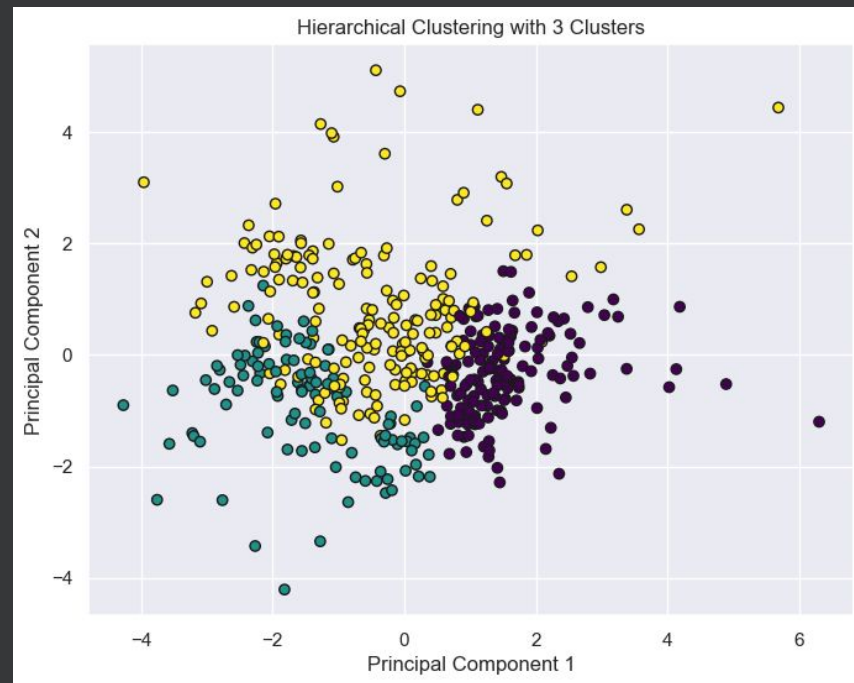
Principal Components Analysis

- With **3 principal Componen**t can preserve about **90% of the prediction** of the model
- **Fresh** has the **highest** average **sales** follow by Grocery

Principal Components:

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	0.104627	-0.542274	-0.571694	0.138351	-0.551338	-0.212235
1	-0.590474	-0.133145	0.006282	-0.589535	0.068624	-0.530389
2	0.631894	0.076076	0.133450	0.033630	0.197258	-0.732852
3	-0.488525	0.061386	0.095672	0.791609	0.077347	-0.340286
4	-0.041160	0.761634	-0.098101	-0.074144	-0.618326	-0.144124
5	-0.027447	0.313958	-0.797835	0.005889	0.513904	0.002238

Explained Variance Ratio vs. Number of Principal Components
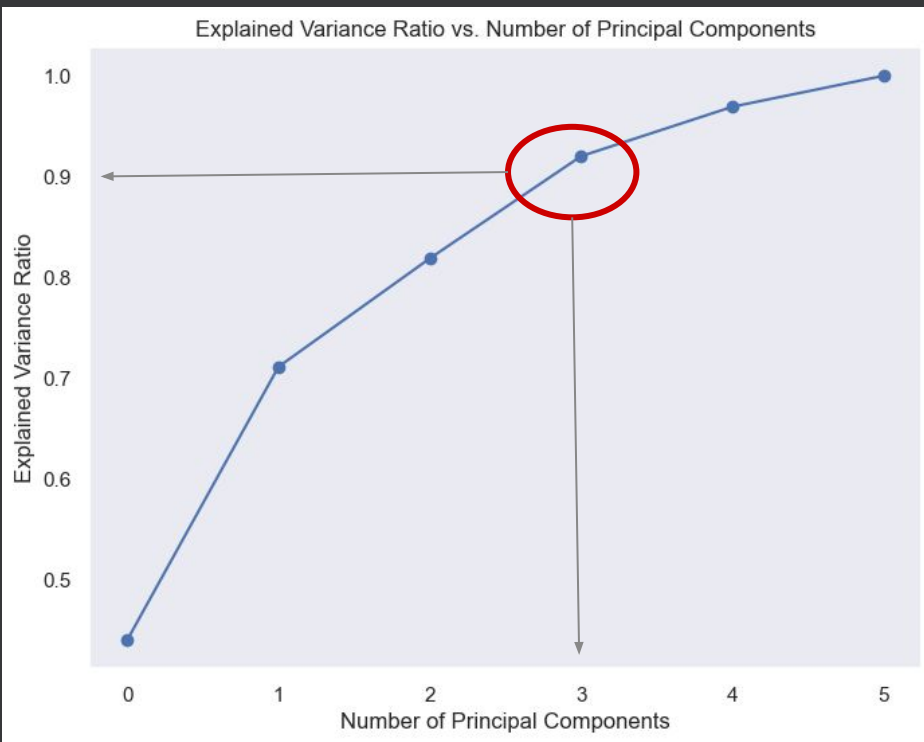
Future Goal

1 Remove Outlier with IQR Score

Enhance the robust of the cluster model by removing outlier with IQR Score

2 Evaluate the Model

Create a for loop to update centroid until convergence

3 Examine PCA

Provide business insights with the 3 products that give 90% of the prediction

Thank you!

Write a closing statement or call-to-action here.