Directed Study Report

Topic - Applying K-Means Clustering Analysis in Retail Industry to Explore Descriptive Price-Demand Relationship and Seasonality effect

Dr.Eugene Pinsky, Boston University

Dr.Andrew Vakhutinsky, Oracel Lab

Yaqiu Guo, Boston University

Nathan Fernandes, Sahil Khanna, Boston University

Introduction

Nowadays, one of the challenges in retail industry research is to draw meaningful inferences from multi-dimensional datasets. In this paper, we investigated and applied k-means clustering analysis method to explore descriptive price-demand relationship in retail environment, as well as seasonality effect.

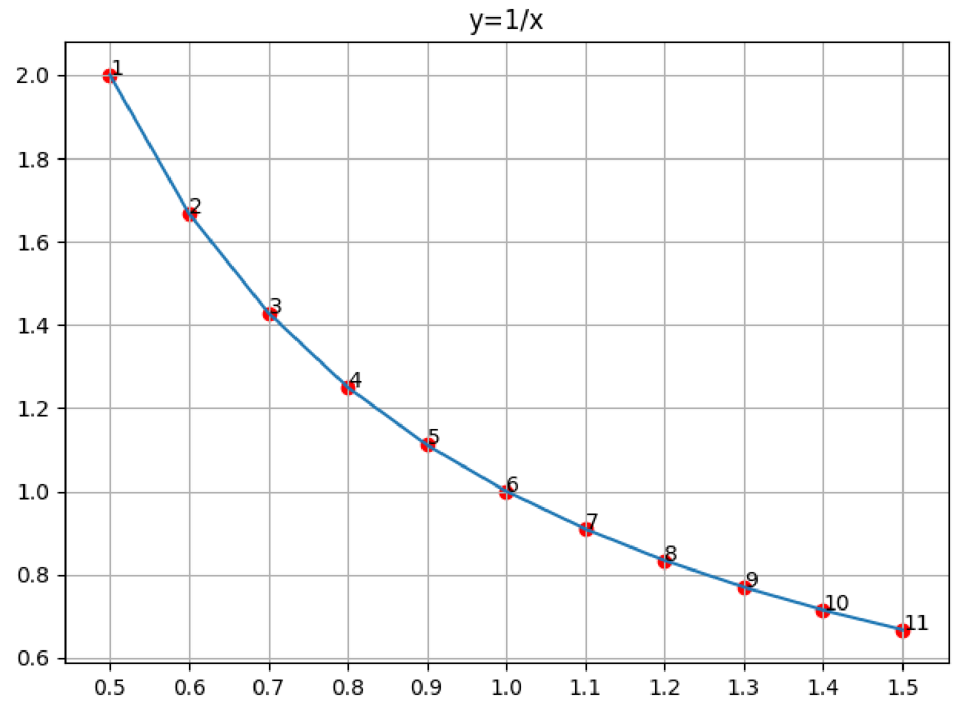
The original dataset was weekly transaction records of single product from one grocery store in Arizona state over two years period (June 2011 – May 2013).

Understanding the dataset and Demand curve

- store_id: all transaction records were from one same store in Arizona, so store_id is 1055 for all records;
- promo_week_id: records were within two years period from June 2011 to May 2013. In the dataset it used 201124 as the starting week and 201324 as the ending week, so there are 105 weeks in total;
- category_id: in retail industry, every product belongs to one category. There is a unique category_id to track different kinds of categories;
- class_id: in retail industry, every category belongs to one class. There is a unique class_id to track different kind of classes;
- upc_id: in retail industry, every product has its own unique id to differentiate from other products;
- txn_cnt: abbreviated for transaction count, total number of times that customers buy this product within this week;
- units: total number of units of this product in total customer buy within this week;

- meas_qty: measurement quantity
- mkdn_amt: total markdown amount(revenue) of this product within this week;
- net_amt: total net amount(revenue) of this product within this week;

Assigning each product's weekly (net_price_ratio, unit_ratio) with a series of predetermined label on y=1/x curve and run k-means clustering on weekly labels



On the y=1/x curve, we manually picked 11 points with x coordinate equals to [0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5] and calculated y coordinate correspondingly. Then we labeled the 11 points from 1 to 11 (from left top to right bottom). Point 6 represents (1.0, 1.0), point 1 - 5 are more elasticity to price change and point 7-11 are less elasticity to price change.

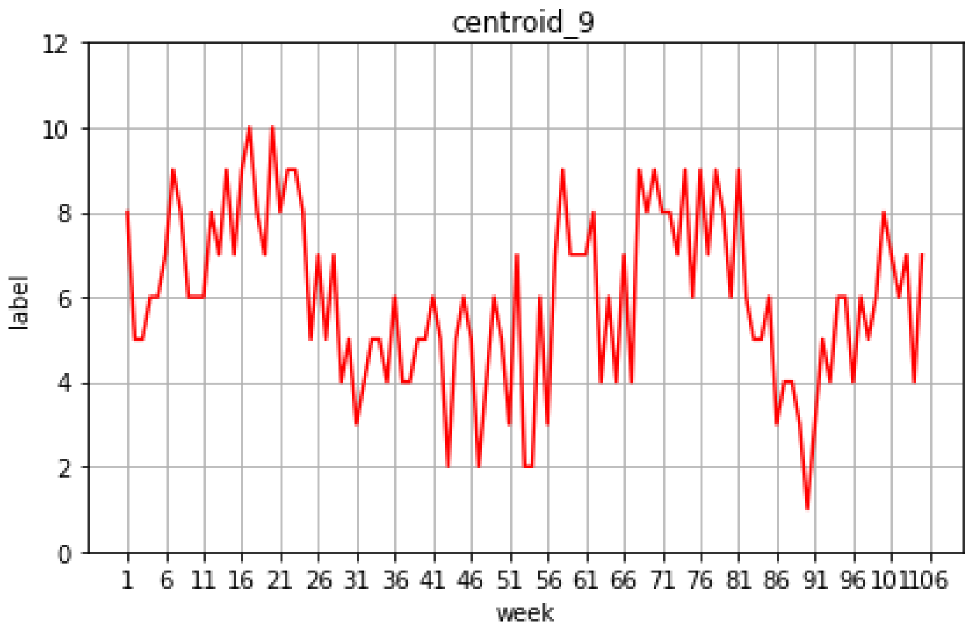
For each product, we calculated the distance between each week's (net_price_ratio, unit_ratio) and the 11 points we selected from y=1/x curve. After finding out the shortest distance, we assigned the label of that point to the original week. In this case, each product was represented by a list with 105 labels and the labels are some certain numbers from 1 to 11.

Running k-means clustering on certain products' weekly data and do trajectory analysis on k means centroids

We ran k-means on this matrix and set the number of clusters equal to 10, which gave us 10 centroids. After doing this, we calculated the distance between each centroid's weekly (net_price_ratio, unit_ratio) and the 11 points we selected from y=1/x curve. By finding out the shortest distance, we assigned the label of that point to the original week of the centroid. In this case, each centroid was represented by a list with 105 labels and the labels are some certain number from 1 to 11. Then we plotted each centroid's labels by weekly and got a trajectory path.

Considering seasonal products such as

Beer

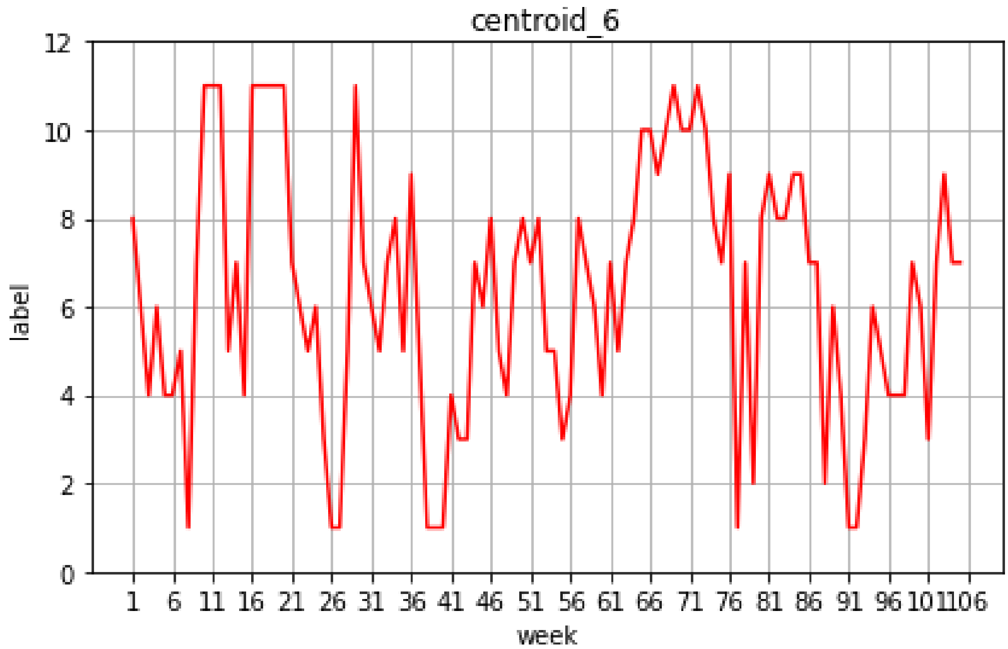


The trajectories showed us that most weeks range from between 4 – 10 normally and sometimes in the year are very elastic that vary from 1 – 3. As we can see, this particular product is very elastic towards price change towards the last 15 weeks of the year.

Increase in Profit per week due to discounts = $3.988517

Items sold per week = 5.1628

Candy



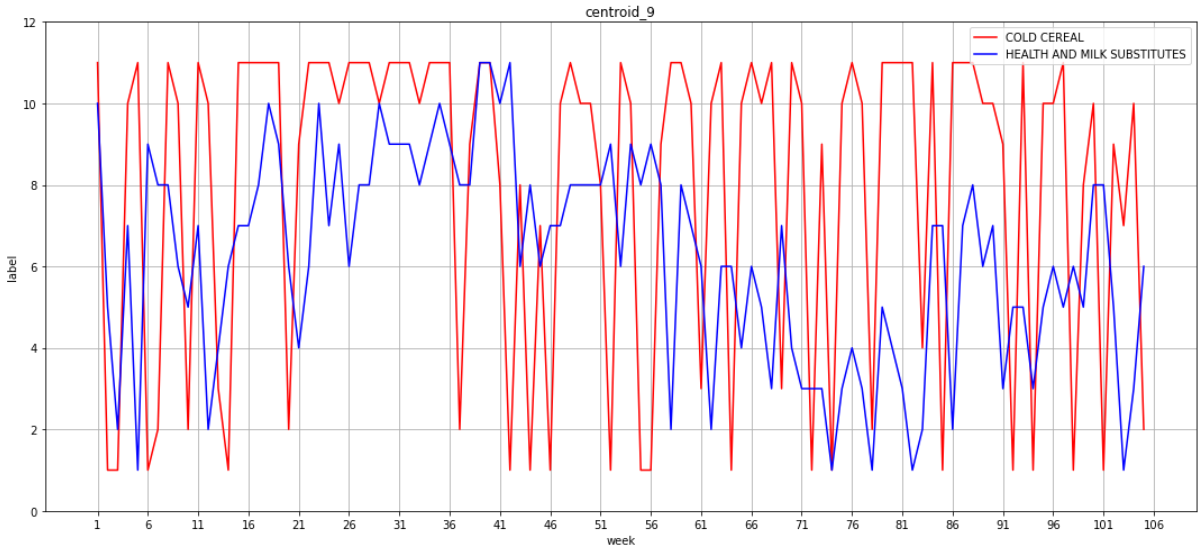
The trajectories of this cluster shows us that there are multiple discounts being offered during the year for this product. We can also see that the discounts being offered are around the same time in both years, meaning that there could be some seasonality in effect here. There are discounts around the 46th to 50th week, that is usually around Halloween and Christmas.

Increase in Profit per week due to discounts = $7.130494

Items sold per week = 4.1163.

Now let's compare and see pairs of related products

Milk and Cereal



For cold cereal, we can see no clear pattern from the analysis from centroid 9's trajectory. It varies from 1-11 without any clear pattern.

Increase in Profit per week due to discounts = $-7.063825

Items sold per week = 1.2143

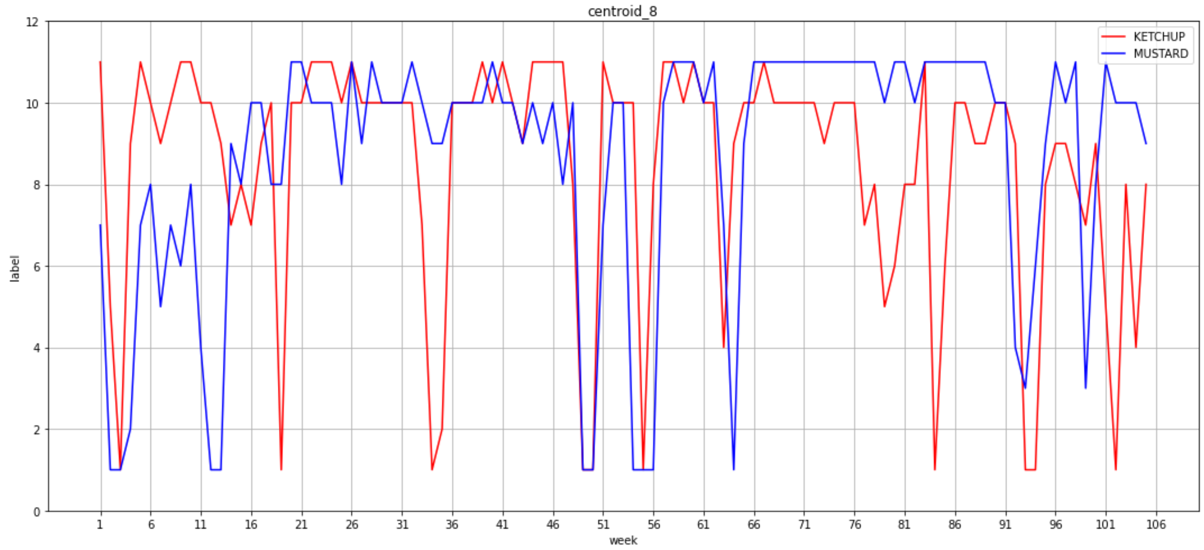
For health and milk substitutes, we can see a sharp increase in elasticity at the beginning of both the years.

Increase in Profit per week due to discounts = $-7.080034

Items sold per week = 1.2308

There is no clear visible display of correlation between the trajectories of cold cereal and health and milk substitutes. We can attribute that due to the cloudy pattern displayed by the trajectories of cold cereal even though health and milk substitutes shows a sharp increase in elasticity at the beginning of the years.

Ketchup and Mustard



For Ketchup, the trajectory fluctuates from less elastic to very elastic. As we can see, there is no clear pattern within both the years.

Increase in Profit per week due to discounts = $4.879418

Items sold per week = 3.2874

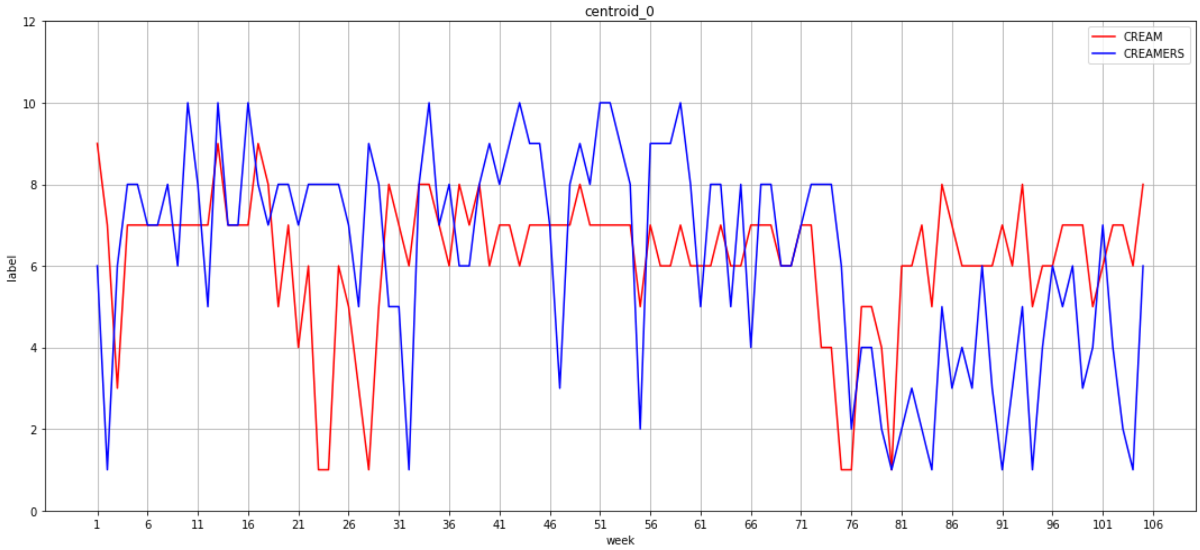
For Mustard, this particular trajectory showed us that most of the weeks were between 8 to 12. That means Mustards from this cluster had less discounts/promotions during most of the year. As we can also see, There are sometimes before the end of the year and at the start of the year that weekly label varies between 1 to 3. This means that during those weeks, the trajectories of Mustard are more elastic.

Increase in Profit per week due to discounts = $-2.686589

Items sold per week = 3.7826

We can also notice that when the trajectories of ketchup becomes very elastic, the trajectories of mustard also become very elastic. That means that there could be some seasonal effect.

Cream and Creamers



For cream, most of the trajectories are in the range of 5 to 8. That also means that cream did not get much discounts throughout the year. But during the start of the year and around the 20th week of both years, the weekly labels vary between 1 and 3 again. This means that the trajectories at those times are more elastic.

Increase in Profit per week due to discounts = $49.195079

Items sold per week = 21.4857

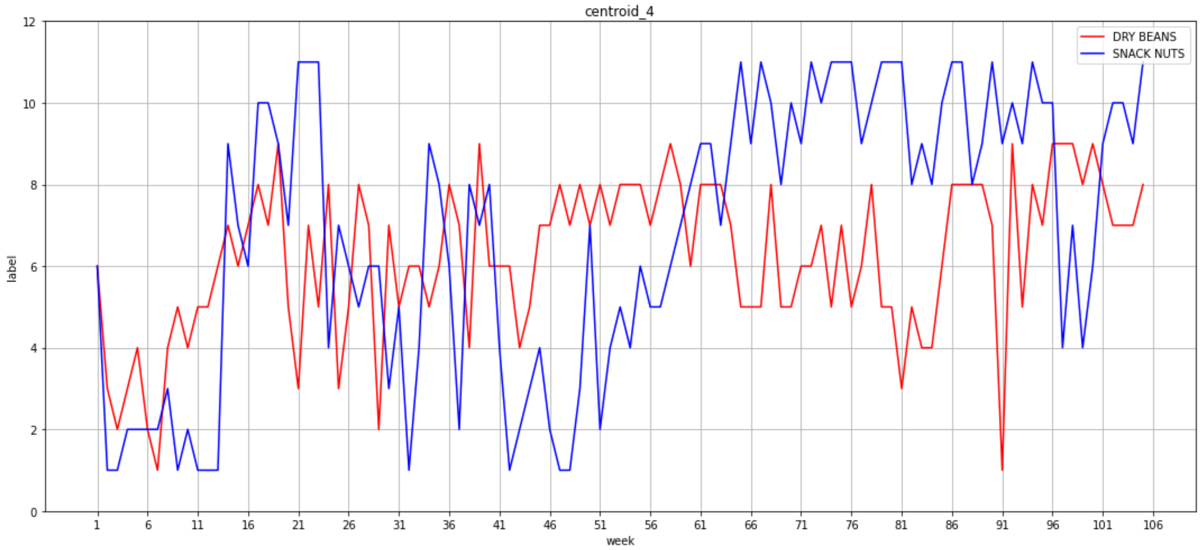
For creamers, the trajectories are not too elastic in the first year and very elastic in the second half of the second year.

Increase in Profit per week due to discounts = $142.412603

Items sold per week = 35.1333

We can also see that there is a relatively high correlation between the trajectories of both the products throughout the year. The trajectories of creamers seem to mimic the trajectories of cream in a few weeks.

Dry Beans and Snack Nuts



Dry Beans did have some high elastic phases throughout both the years and we can see it's lowest trajectory has reached 1.

Increase in Profit per week due to discounts = $-7.137762

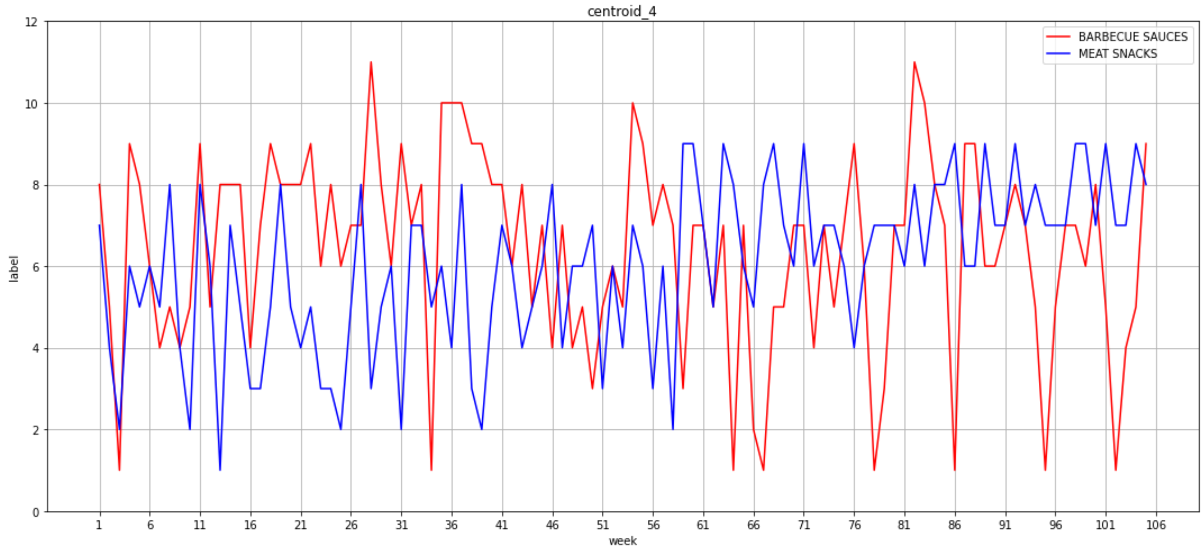
Items sold per week = 2.4302

Snack nuts on the other hand was very elastic during year one but not nearly as elastic in year 2. There is no obvious correlation between the trajectories of both the products as well.

Increase in Profit per week due to discounts = $-3.223986

Items sold per week = 2.4625

Barbecue Sauces and Meat Snacks



We can see that there is no obvious trend within the two years for both these products.

Barbecue Sauce

Increase in Profit per week due to discounts = $-3.340444

Items sold per week = 4.2

Meat Snacks

Increase in Profit per week due to discounts = $-1.389337

Items sold per week = 1.7547

[Reference]

1. SHOPPER: A PROBABILISTIC MODEL OF CONSUMER CHOICE WITH SUBSTITUTES AND COMPLEMENTS

By Francisco J. R. Ruiz, Susan Athey and David M. Blei

2. Pattern Recognition and Machine Learning

By Christopher M. Bishop

3. A Prescriptive Analytics Approach to Markdown Pricing for an E-Commerce Retailer

By Andrew Vakhutinsky, Kresimir Mihic and Su-Ming Wu

4. Estimating Price Elasticities in the Travel Industry under Revenue Management Controls

By Pelin Pekgun, Paul M. Griffin and Pinar Keskinocak

5. QUANTIFYING PROMOTION EFFECTIVENESS FOR A RETAILER

By Aleksi Pesonen