# CLEAN SWEEP

## A study on the price elasticity of selected household cleaning brands


Group 10

Thahir Hussain MH | Uma Perumal | Prabakaran V | Bodhisattwa Sengupta | Dan Anamitra | Ullas DP


In collaboration with:

JIGSAW ACADEMY

THE ONLINE SCHOOL OF ANALYTICS

# Problem Statement

To understand the pricing dynamics for a leading household cleaning supplies brand in USA

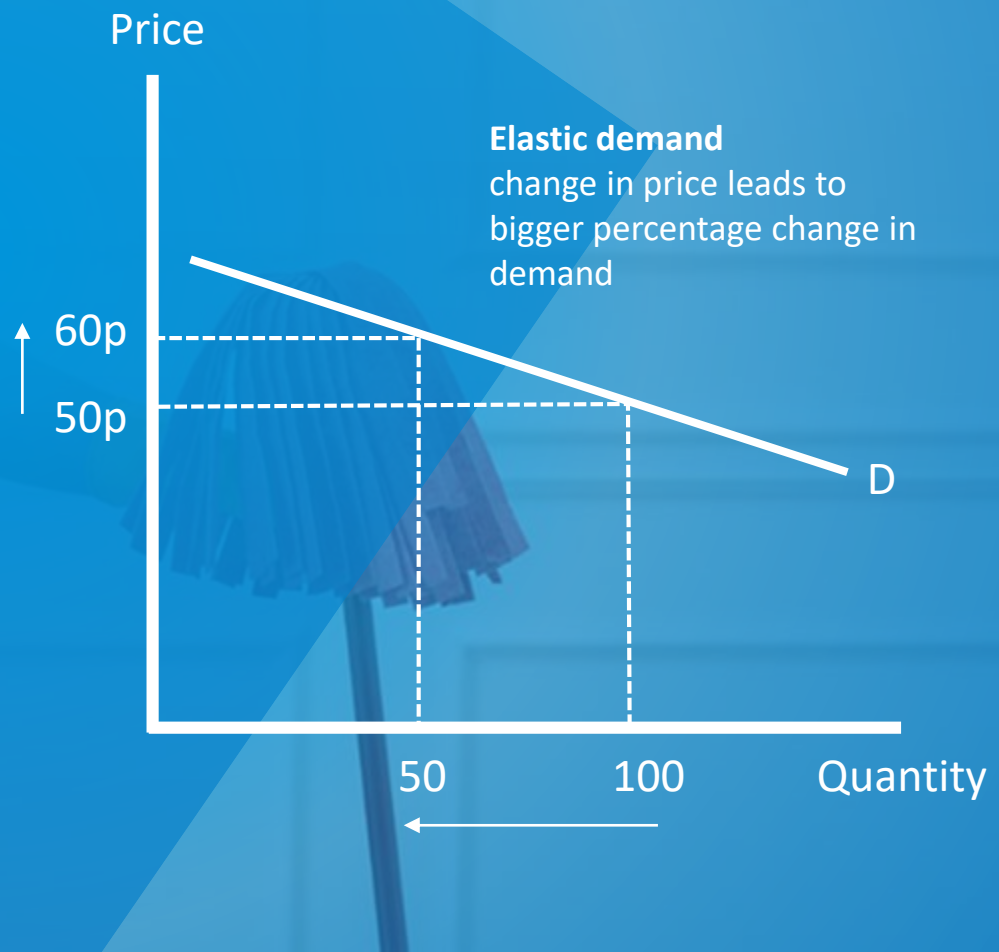
- Given 3 different customers(retailers), determine the price elasticity of the target product based on customer segments
- Identify the impact of key competitor’s pricing on the target product
- Create price elasticity curve based on time period and evaluate if the elasticity increased or decreased over time
- Estimate the elasticity of the target product based on channels/group of stores

# About Price Elasticity

Price elasticity of demand is an economic measure of the change in the quantity demanded or purchased of a product in relation to its price change.

$$\textit{Price Elasticity of Demand (E)} = \frac{\textit{\% Change in Quantity Demanded}}{\textit{\% Change in Price}}$$

| | |
|---|---|
| $E = \alpha$ | Perfectly elastic |
| $E > 1$ | Elastic |
| $E = 1$ | Unit Elastic |
| $E < 1$ | Inelastic |
| $E = 0$ | Perfectly inelastic |




**Elastic demand**
change in price leads to bigger percentage change in demand

# Statistical Methodology for Elasticity

## Linear Regression

A linear approach to model the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). In our case,

$$Q = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e_i$$

*Q – Tot. Quantity sold of the target product*
*X1 – Price of the target product*
*X2 – Price of the related product*
*X3 – Price of the competitor product*
*X4 – Time period*

**Price elasticity of demand:** $\Delta Q / \Delta Px_1$

**Cross Price elasticity of demand:** $\Delta Q / \Delta Px_2$

# Our Understanding of the data set

| Category | Customer 1 | Customer 2 | Customer 3 |
|---|---|---|---|
| **Total Products** | 7 | 7 | 10 |
| **Stores** | 3987 | Absence of Stores & States data | 1981 |
| **States** | 42 | Can be an Online/ Wholesale retailer | 55 |
| **Target Products** | WD009, WD012, WD017, WD029 | WD008, WD012, WD017 | WD008, WD009, WD012, WD017 |
| **Competitor's Products** | SP001, SPC11, SPC13 | SP001, SPC10, SPC11, SPC65 | SP001, SP018, SPC010, SPC11, SPC15 |
| **Missing Values** | 0.03% | Nil | 1.91% |
| **Transaction Details** | 61 weeks (24th April 2015 to 17th July 2016) | 66 weeks (6th March 2015 to 3rd June 2016 ) | 70 weeks (1st March 2015 to 18th June 2016) |

# Process Flow

Slice and dice data by products via Manipulation & Visualization

Data aggregation based on week/products/stores

Missing/Negative Value Treatment

Outlier detection in the aggregated data

Feature Engineering variables like Unit Price, Week, log of variables

Model building (Linear Regression)

Model Evaluation

Estimating Price Elasticity and Cross-Price Elasticity

Comparing Price Elasticity

# Key Assumptions and Imputation logic

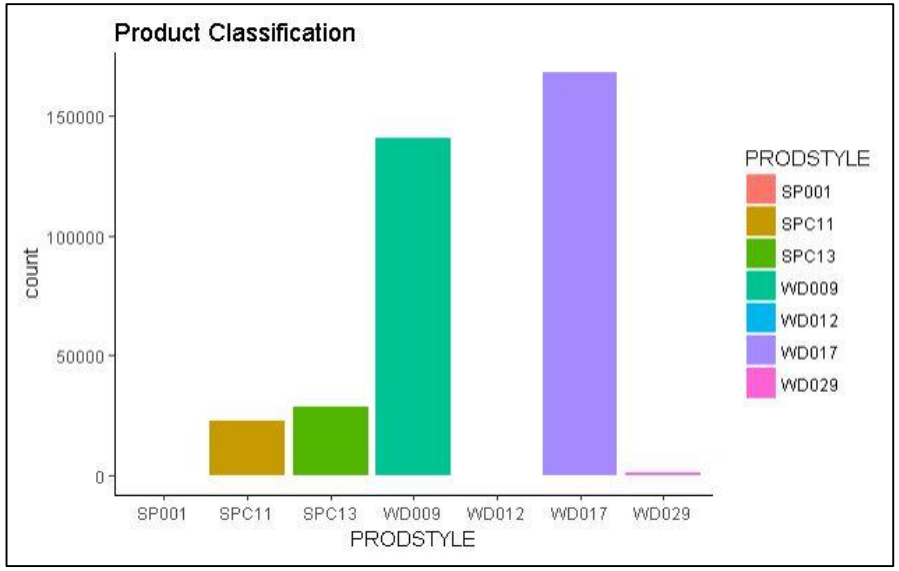
- PRODSTYLE variable was used as the key differentiating variable for product selection
- Removed observations with negative values in Price/Quantity sold. They comprise <1% of the entire dataset. Such negative values might represent product returns/cashbacks/discounts.
- Being a weekly sales data, outliers detected were not removed as they might represent the actual sales during a particular week.
- In Customer 3, the missing sales data of WD017 was imputed using the data from WD009 that was stopped selling in mid-way. WD017 might be a replacement for WD009
- Store clustering was done based on WD008 product only, as it was the most elastic.
- The missing data of certain products in the segmented clusters were imputed based on the average of previous 4 weeks

# Evaluation metrics used

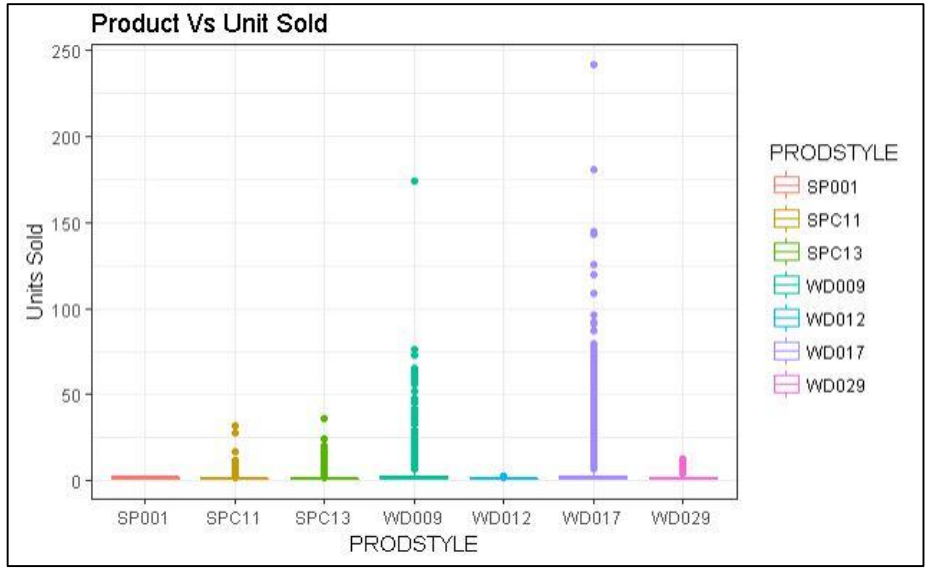
- Linear Regression for elasticity – RMSE, MAE, R2, VIF
- K-Means Clustering for store segmentation – Elbow method, Within Sum of Squares


Customer 1

# Customer 1 – Product Classification

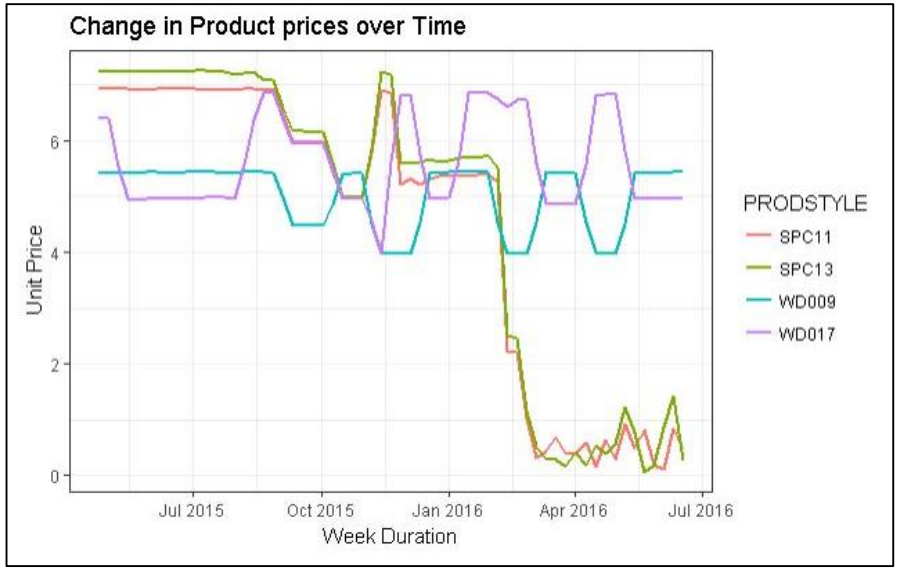


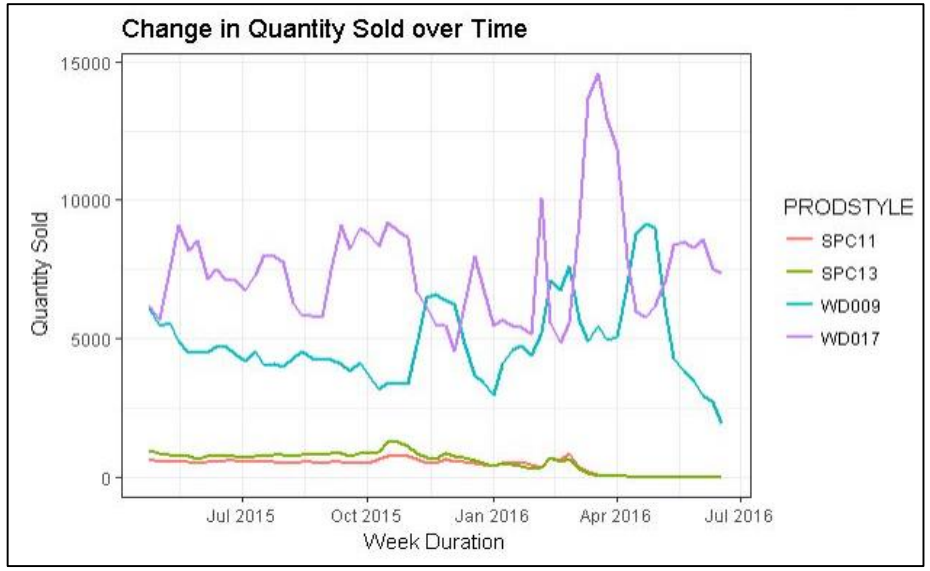

| SP001 | SPC11 | SPC13 | WD009 | WD012 | WD017 | WD029 |
|---|---|---|---|---|---|---|
| 39 | 22820 | 28325 | 140900 | 30 | 168199 | 1074 |




- Products - SP001,WD012, WD029 do not have enough history of data and needs to be removed.
- WD017 is the most preferred product among all.

# Customer 1 – Price Vs Demand

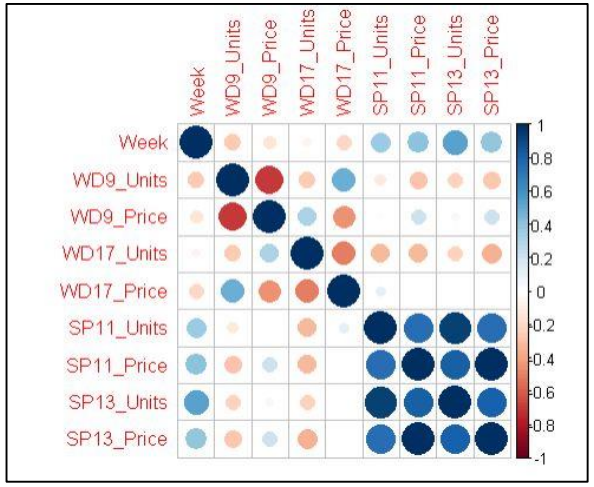



- In the initial weeks, the price of all the products were competitively priced around $6. But the price of SPC11 and SPC13 sharply declined to $1 after Mar 2016
- The demand for WD009 and WD017 have responded positively to their decrease in price. The demand of competitive products remain stagnant over the entire time duration

# Inference from outputs of Customer 1

## Correlation Matrix

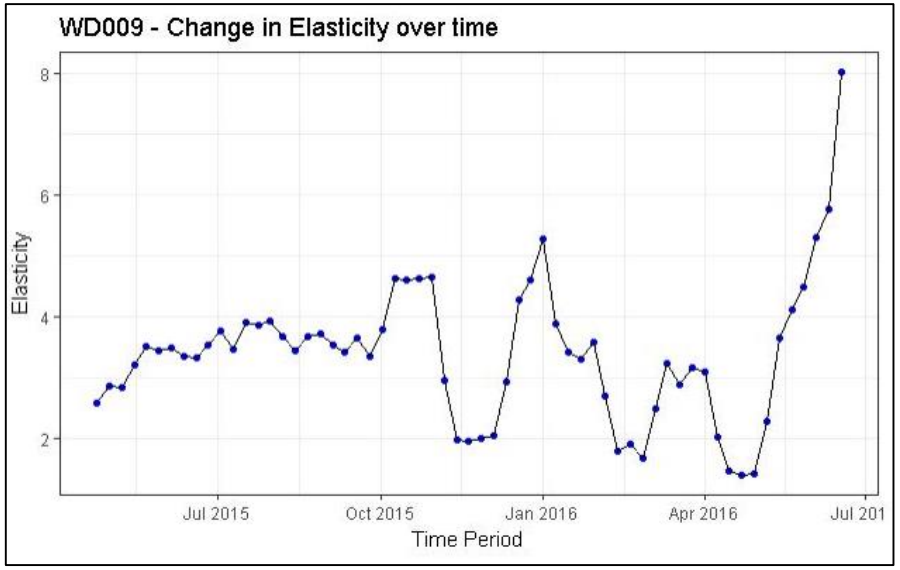


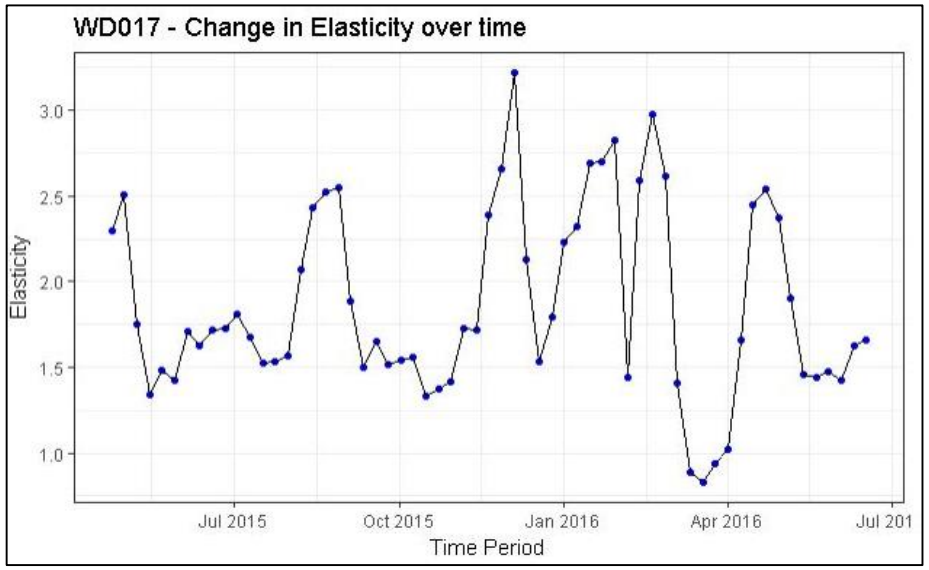

## Elasticity Matrix

| | WD009 | WD017 | SPC11 |
|---|---|---|---|
| WD009 | **-3.08** ⬇ | NA | NA |
| WD017 | NA | **-1.73** ⬇ | NA |
| SPC11 | NA | -0.16 | NA |

- WD009 is the most elastic product when compared to WD017 and SPC11.
- A 10% decrease in price of WD009 will lead to 30.8% increase in their demand. Assume if 1000 products are sold each at $5 in a week, reducing the price by 10% i.e. keeping it at $4.5, the demand of WD009 will increase by 1308. Subsequently the revenue in a week will increase from $5000 to $5886

# Elasticity based on Time Period

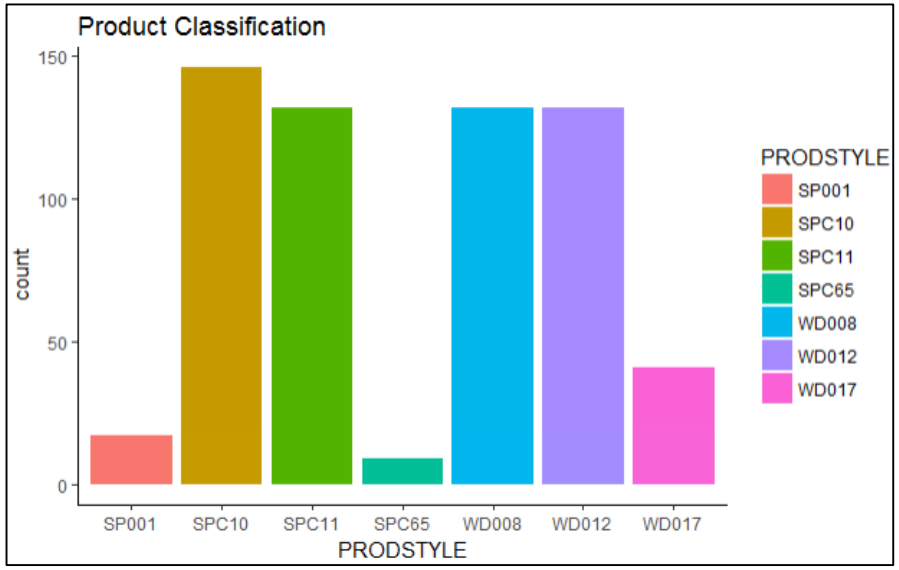


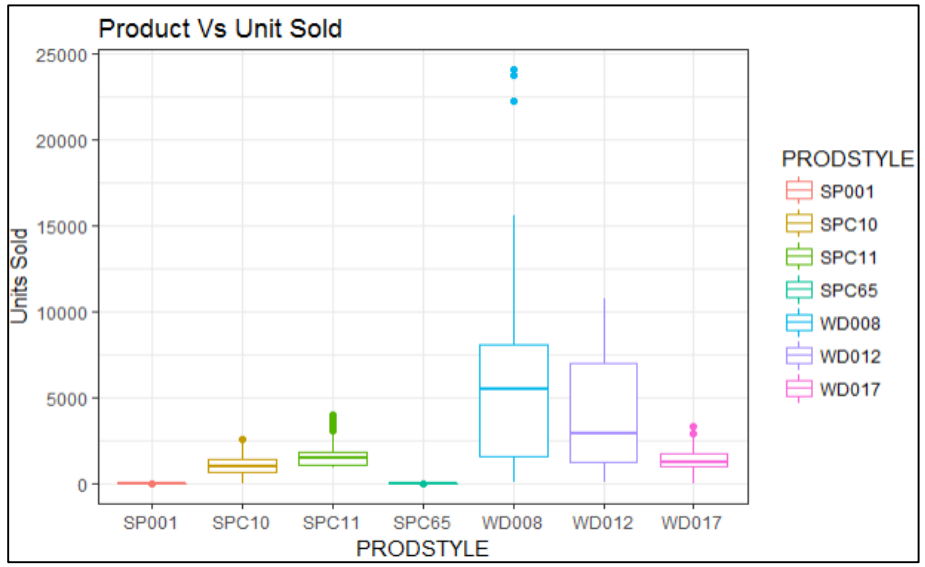

- Elasticity of WD009 has increased over the time and almost reached a peak elasticity of 8 in Jun 2016.
- For WD017, the elasticity has been in the range of 1 and 3, but has not increased over the period of time.

# Customer 2

# Customer 2 – Product Classification

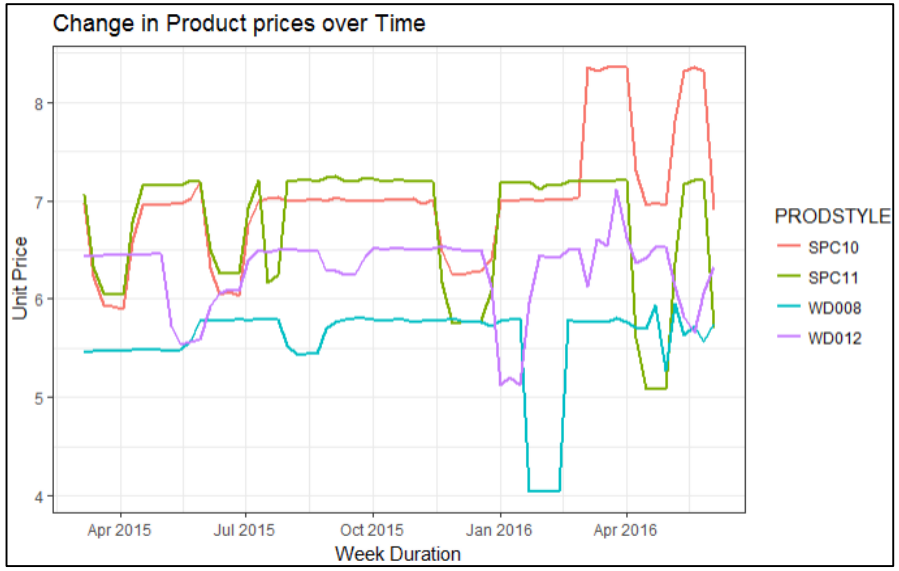


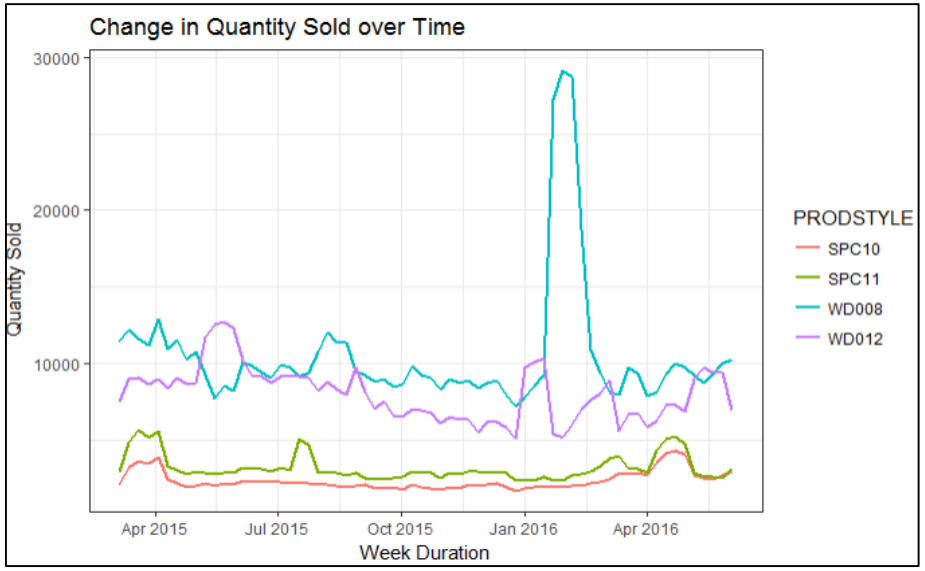

| SP001 | SPC10 | SPC11 | SPC65 | WD008 | WD012 | WD017 |
|---|---|---|---|---|---|---|
| 17 | 146 | 132 | 9 | 132 | 132 | 41 |




- Products - SP001, SPC65, WD017 do not have enough history of data and needs to be removed.
- WD008 is the most preferred product among all.
- WD008 has a higher average units sold per week – in excess of 5000 units

# Customer 2 – Price Vs Demand

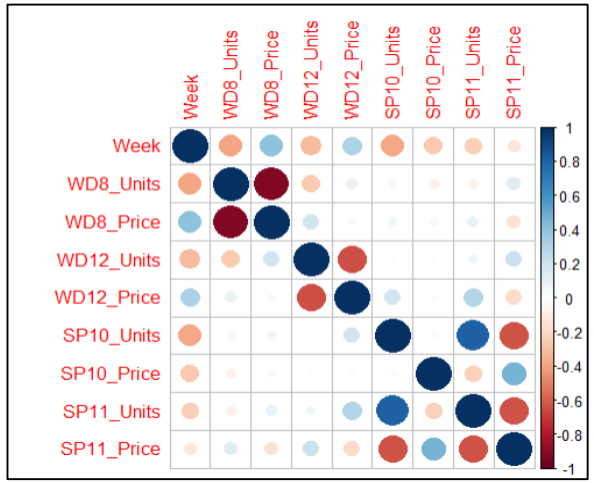



- The price of all products are competitively priced over the entire time duration.
- Despite same price, the demand for SP products are relatively low compared to WD products.
- The peak demand in WD008 in Feb 2016 coincides with the price reduction. Heavy discounts and promotion strategies were probably employed during this time.

# Inference from outputs of Customer 2

## Correlation Matrix



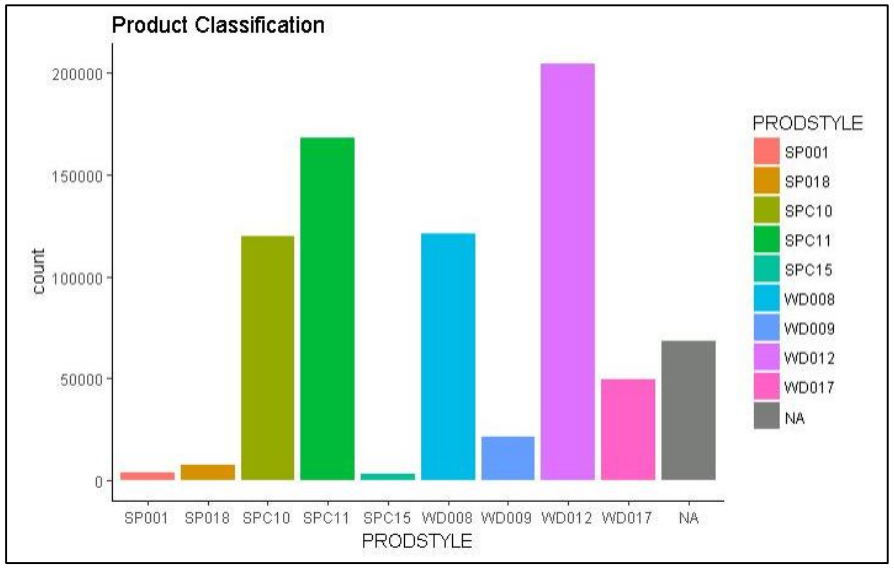

## Elasticity Matrix

| | WD008 | WD012 | SPC10 | SPC11 |
|---|---|---|---|---|
| WD008 | **-0.47** | 0.14 | NA | NA |
| WD012 | 0.16 | **-0.40** | NA | 0.23 |
| SPC10 | -0.08 | -0.20 | **0.20** | -0.20 |
| SPC11 | NA | 0.35 | -0.24 | NA |

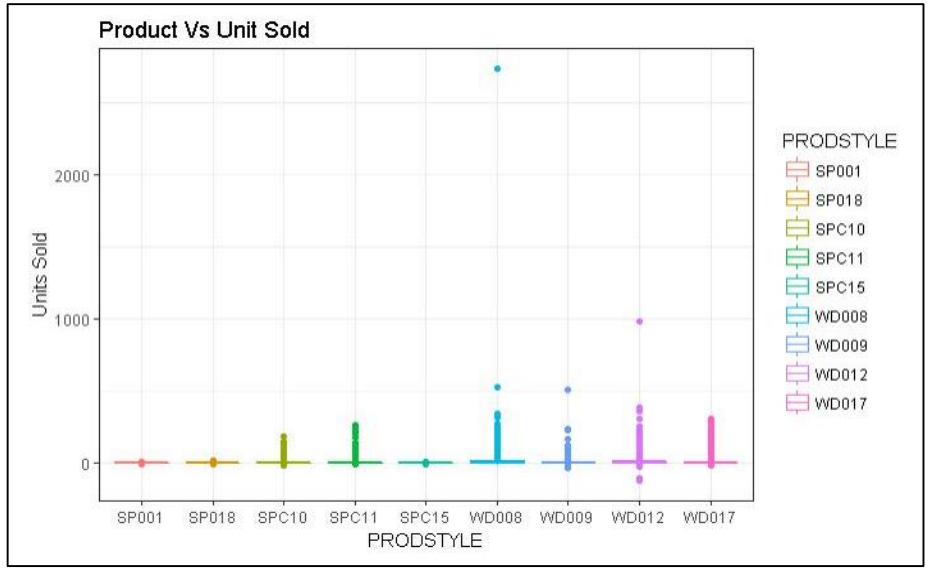
- Price elasticity of all products are between 0 and 1. So all the products are inelastic in nature.
- There is only a slight change in quantity sold with change in their price .

# Customer 3

# Customer 3 – Product Classification

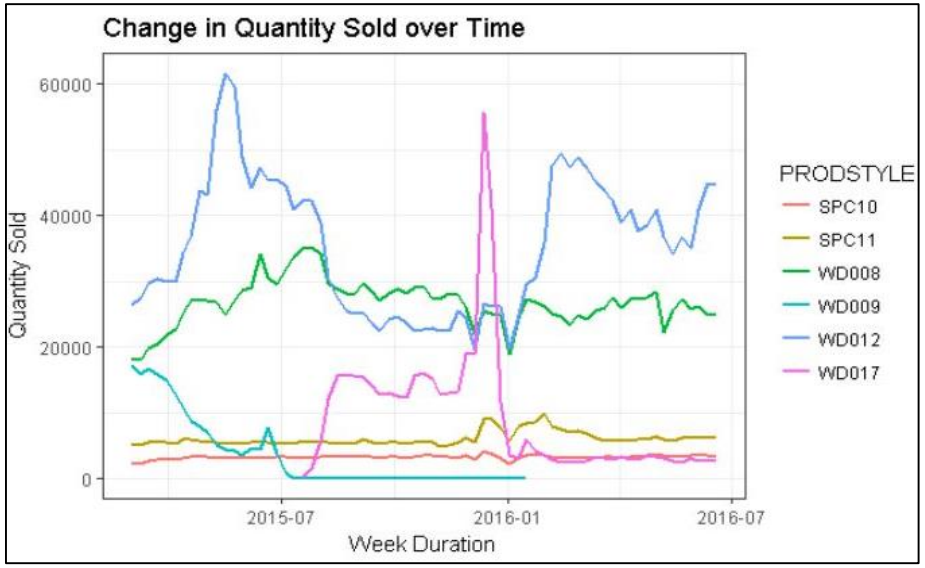



| SP001 | SP018 | SPC10 | SPC11 | SPC15 | WD008 | WD009 | WD012 | WD017 |
|---|---|---|---|---|---|---|---|---|
| 3866 | 7830 | 119495 | 168171 | 3273 | 121276 | 21599 | 204543 | 49815 |




- Products - SP001,SP018, SPC15 do not have enough history of data and needs to be removed.
- WD012 is the most preferred product among all. The next most preferred product is WD008.
- Among the competitors, SPC11 is the most preferred.

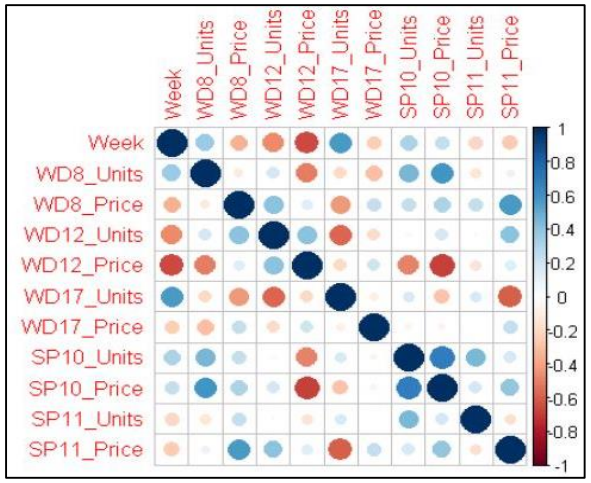
# Customer 3 – Price Vs Demand


- For Customer 3 also, the demand for SP products are stagnant despite having a competitive price
- Unlike in Customer 1&2 , the distribution for WD009 was discontinued after Sept 2015. WD017 was introduced around the same time which might be a replacement for WD009

# Inference from outputs of Customer 3

## Correlation Matrix

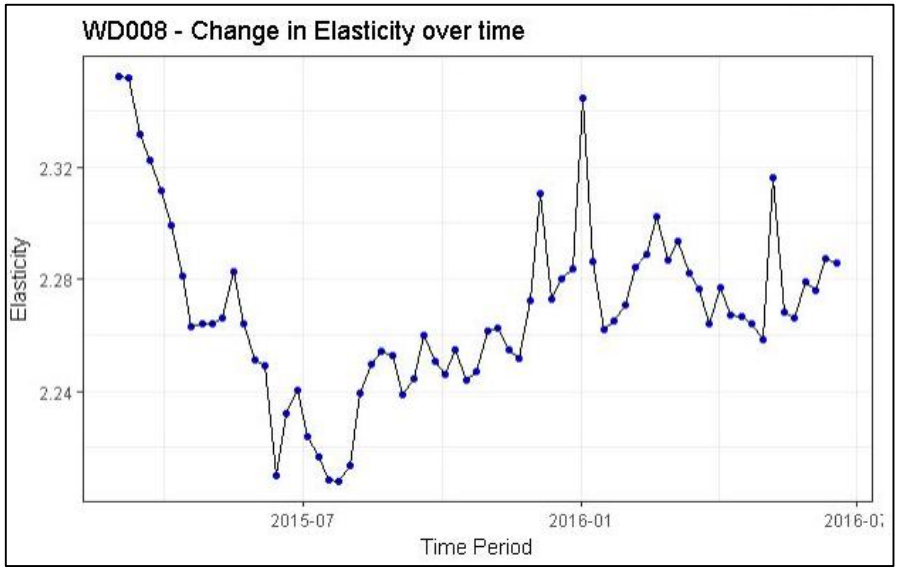


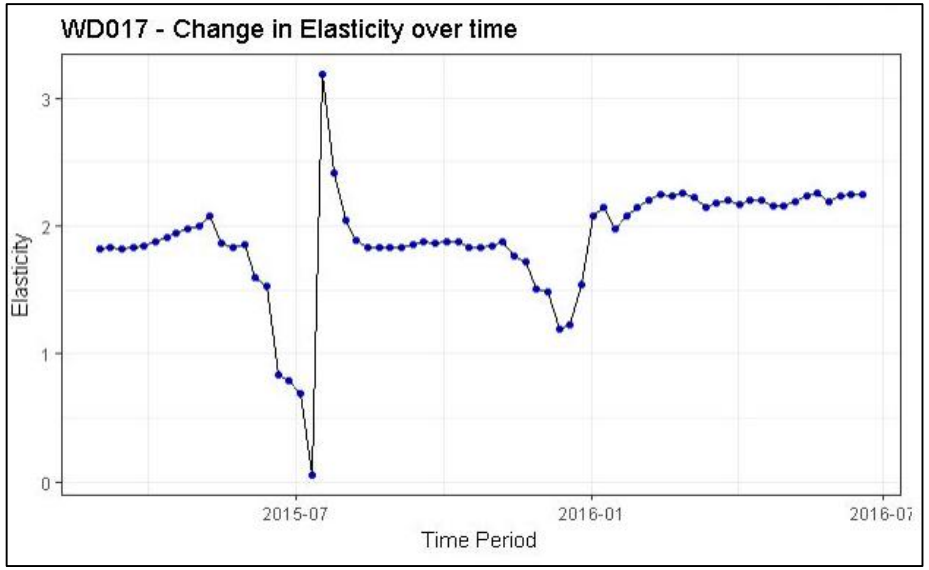

## Elasticity Matrix

| | WD008 | WD012 | WD017 | SPC10 | SPC11 |
|---|---|---|---|---|---|
| WD008 | **-2.27** ↓ | NA | -0.22 | NA | NA |
| WD012 | -0.14 | **0.61** | NA | NA | NA |
| WD017 | NA | 0.00 | **1.89** ↑ | NA | NA |
| SPC10 | NA | 0.00 | NA | **1.26** ↑ | NA |
| SPC11 | NA | NA | -1.33 | NA | **-2.93** ↓ |

- SPC11 is the most elastic product. The second most elastic is WD008
- A 10% decrease in price of WD008 will lead to 22.7% increase in their demand.
- In WD017, a 10% increase in their price will lead to 18.9% increase in demand. But, 10% increase in the SPC11 price will lead to 13.3% decrease in demand for WD017. SPC11 is a substitute for WD017.

# Elasticity based on Time Period



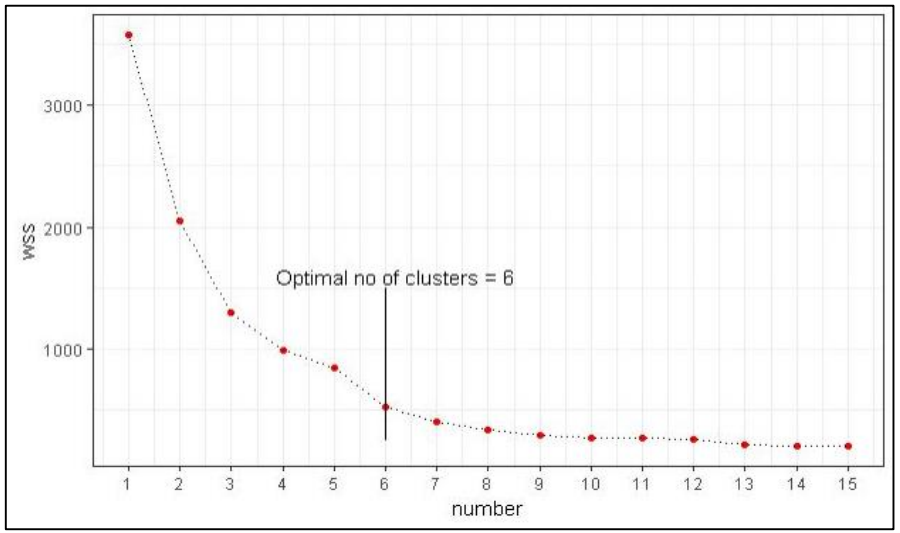

- Elasticity of WD008 has decreased over the period, but not a big shift in the elasticity.
- Elasticity for WD017 has remained constant around 2, except during the heavily promoted period in Dec 2015

# Customer 3

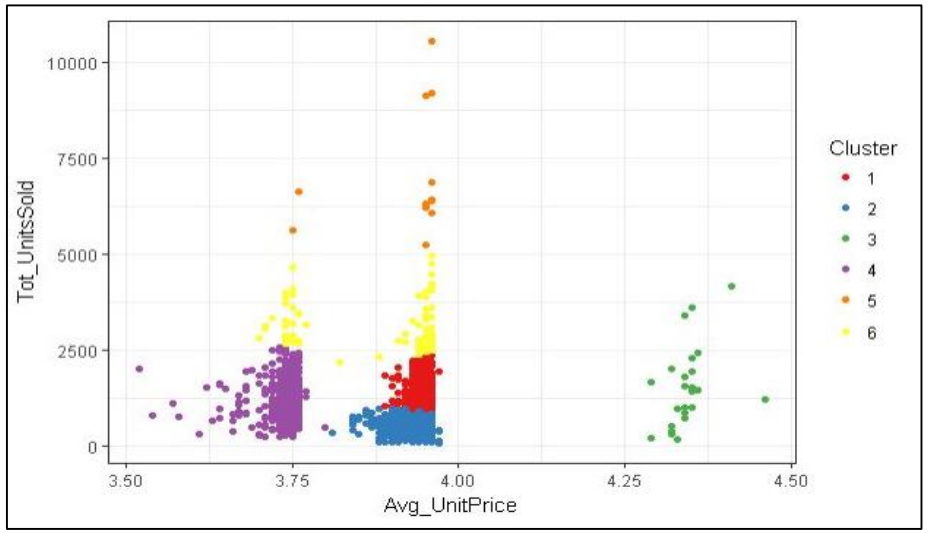
## (Store Segmentation)

# Customer 3 – Store Clustering

## Scree Plot




## Store Segmentation




- Optimal clusters of 6 were formed using K-Means clustering.

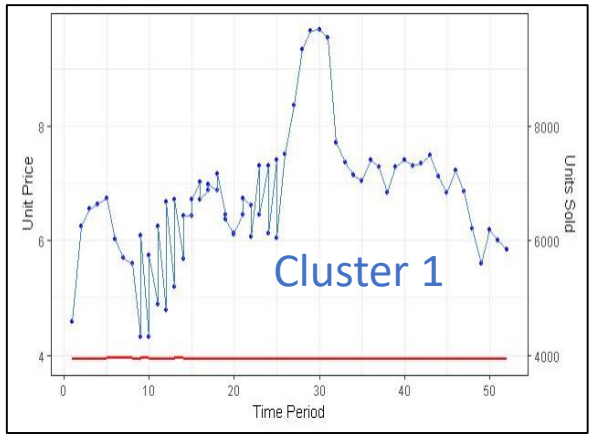
# Customer 3 – Store Elasticity matrix

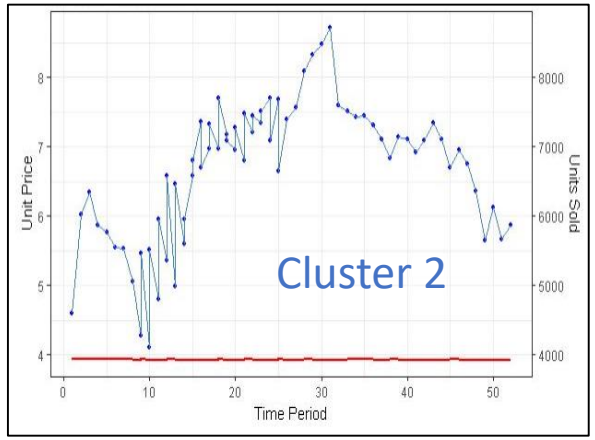
## Price elasticity for WD008 within segments

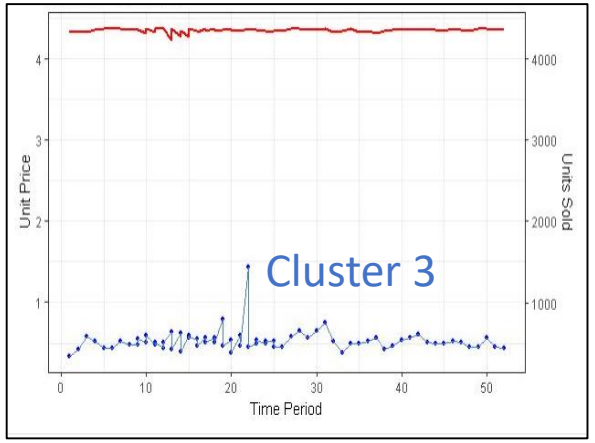
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| **No of Stores** | 335 | 890 | 24 | 445 | 12 | 82 |
| **WD008** | -3.58 ⬇ | -2.6 ⬇ | -1.29 ⬇ | NA | NA | NA |

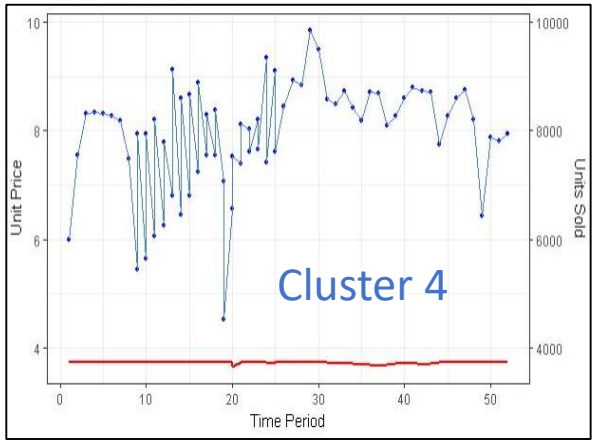
- WD008 is most elastic in Cluster 1 when compared to other Clusters with a score of 3.58
- A 10% price increase for WD008 will have a 35.8% decrease in demand in Cluster 1
- Out of 1981 stores for Customer 3, 1249 stores i.e. 63% of the stores have price elasticity for WD008(score of >1)
- For Clusters 4,5 & 6 – elasticity of WD008 cannot be determined as there were no significant relationships between price change and quantity sold.

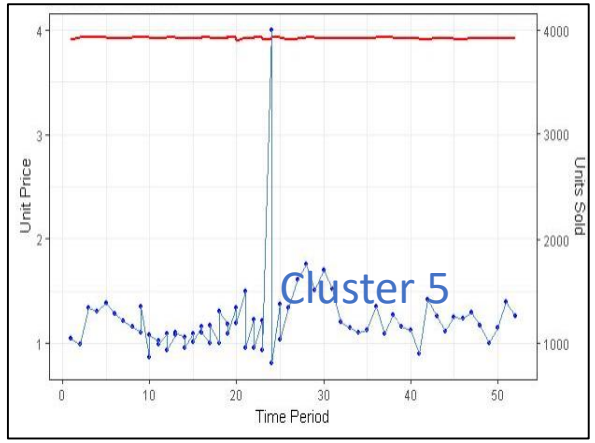
# Customer 3 Store – Price Vs Demand

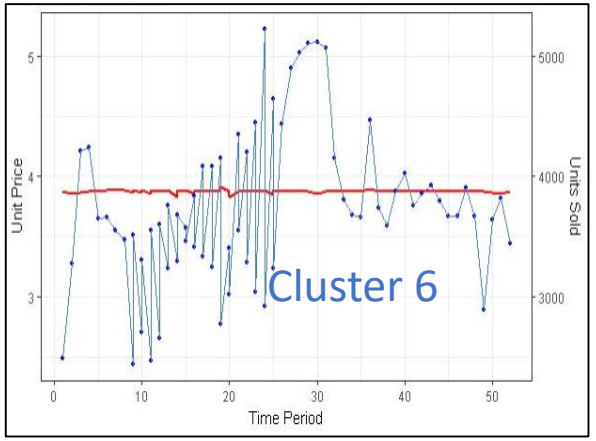



- There is only a slight shift in changes of quantity sold for Cluster 4 & 6, which might be due to the changes in the price of other products.

# Summary & Recommendations

- WD008 and WD009 are the most elastic products and are sensitive to the changes in their own price. Their pricing strategy needs to be carefully considered by factors such as consumer sentiments, time period and other external factors.
- WD012 is inelastic, so the manufacturer can increase the price marginally for incremental profits.
- WD017 is least elastic, except during the promotion periods. They may have substitutes in the market and may need to be priced competitively to gain market share.
- WD products are inelastic for customer 2. The manufacturer has a cash cow in this segment.
- Store segments having higher elasticity can be focused while creating a pricing strategy.

This study can be further extended for price optimization, demand forecasting and inventory prediction.

# Annexure for reference

Data Dictionary & Key details across customers from dataset

1. EndDate: Time period between 2015 -2016
2. CustomerNumber: 1493, 44575, 34460
3. StoreNumber: Unique id of the stores
4. StateTerritory: Unique names of States in USA
5. Postalcodes: Unique postal codes
6. PartNumber: Alphanumeric Variable - Product Identifier
7. ProdRptCat: 811,835, 850 - representing product part codes
8. PRODSTYLE: SP001, SPC10, SPC11, SPC13, SPC15, SP018, SPC65, WD008, WD009, WD012 , WD017, WD029 - representing products
9. UnitsSold: No of Units sold
10. AmtSold: Amount sold in US dollars


THANK YOU