



**ANALYTIX**LABS

## **SAS Master Case Study-2: Sample Solution- Methodology**

Disclaimer: This material is protected under copyright act AnalytixLabs ©, 2011-2016. Unauthorized use and/ or duplication of this material or any part of this material including data, in any form without explicit and written permission from AnalytixLabs is strictly prohibited. Any violation of this copyright will attract legal actions

# Outline of the Business Problem

- **The Business Objective:**

- Determine product strategy and pricing policies that will maximise company's projected revenues in 2009.

- **Expectations from the Trainees:**

- Uncover any information in the available data that may be useful in meeting the business objective, and make specific recommendations to management that follow from this.
- Also assess the relevance of the data provided, and suggest how the company can make better use of data in 2010 to shape this aspect of their business strategy and operations.
- Clearly define approach and provide the insights and approach in presentation and analysis in Excel with proper formatting. Also provide detailed codes with sufficient comments

# Data

An analysis was performed to uncover key drivers behind company annual revenue and volume.

Data was collected along the following dimensions which were made available

TRANSACTIONAL	PRODUCT	STORES	POSTCODES
<ul style="list-style-type: none"><li>• 297,572 transactions across various stores</li></ul>	<ul style="list-style-type: none"><li>• Product categories and dimensions</li></ul>	<ul style="list-style-type: none"><li>• Store locations defined by postcodes</li></ul>	<ul style="list-style-type: none"><li>• Lat/Long information relating to postcodes</li></ul>

Transactional information had 14 rows with missing information on retail price which were removed.

# Steps for data preparation

1. Create POS file by appending “POS\_Q1”, “POS\_Q2”, “POS\_Q3” and “POS\_Q4” files
2. Create Final file by merging POS, Laptops (Models & configuration information), store post codes (rename os\_x as store\_x and os\_y as store\_y) and London postal codes (rename os\_x as customer\_x and os\_y as customer\_y).

[Note: 1. You can merge all other files(laptops, store post codes, London post codes) with POS using left join by taking POS as left table 2. os\_x, os\_y are the x and y co-ordinates of the locations 3. London postal codes file can be used as customer post codes]

3. Create New variable called “distance” by calculating Euclidian distance between store postal codes (store\_x and store\_y) and customer postal codes(customer\_x and customer\_y)

# Sample Insights from the analysis

[Note: The sample outputs/insights are provide direction to solve the questions but these are not the final solutions. Please note that, you may get different outputs/insights. As you know that, every problem can be solved in many ways and try with all possible ways]

# Are the prices changing?

- Does the Laptop price change with time?

- ✓ There is price drop of ~\$5 every month across all models. This results average price change is ~35% (Between Jan and Dec) which is very high.

[Note: It may require to analyze the average price by configuration and month for all stores or individual store level. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]

- Are prices over retail outlets consistent?

- ✓ Prices over retail outlets are consistent except CR78IE, SW1P 3AU and W43PH. These stores are offering 30% discounts across all models at end of each quarter (i.e March, June, September & December).

[Note: It may require to analyze the average price by configuration and month for each individual store level. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create reports. And observe the pattern of price drop and possible discounts at month level and quarter level etc.]

- How does price change with configuration?

- ✓ The price variation between different configurations ~70%. This variation is increasing over the period of time (Jan – 66% and Dec- 81%) and is very high in end of each quarter.

[Note: It may require to analyze the variation/standard deviation, range of price by configuration and month for all stores or individual store level. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create reports. ]

# How location is influencing sales?

- Where are the stores and customers located?

- ✓ The minimum distance between two stores is ~1.6KM. Customers visiting stores different places (up to 20 KM)

[Note: You can use distance variable to find average distance and the percentage of customers visiting store by each kilometer. You calculate total sales, proportion of sales, Total volume and proportion of volume by each kilometer distance. You can also find maximum, minimum, and average distance of customers and stores. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]

- Which stores are selling most?

- ✓ The maximum sales(~71%) contributing below 5 stores

- SW1P 3AU – 19%
- SE1 2BN – 16%
- SW1V 4QQ – 15%
- E2 0RY – 11%
- NW5 2QH – 10%

[Note: You can calculate total sales, proportion of sales and cumulative proportion of sales by each store. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]

- How far would customers travel to buy a laptop?

- ✓ Customers visiting stores from long distances up to ~20KM. But maximum sales(~94%) is coming from the customers who are around 1 to 7 KM radius.

[Note: You can use distance variable to find average distance and the percentage of customers visiting store by each kilometer. You calculate total sales, proportion of sales, Total volume and proportion of volume by each kilometer distance. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]

# How configuration is influence prices of laptops?

- What are the details of each configuration, and how does this relate to price?
  - ✓ All features including Screen Size, Battery Life, RAM, Processor Speeds, Wireless, HD Size and Applications Bundling are influencing prices.
  - ✓ Prices are correlated positively with configuration.

[Note: You may choose to create buckets for the different configuration parameters(e.g: RAM can be divided into 3-4 buckets like High/Medium/Low RAM using percentiles/quartiles as cut-offs) and you can calculate total sales , Total volume, proportion of sales and proportion of volume by each configuration group or combination of the groups. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create reports.]

- Do all stores sell all configurations?
  - ✓ Maximum stores sold at least one configuration in the year except few stores (E2ORY, CR78LE, E78NW, KT2 5AU, N17 6QA, N3 1DH, S1P 3AU and W4 3PH). S1P3AU sold only 138 different types of configuration laptops.

[Note: It may require to calculate count of models sold for every individual store and model level. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]



# How Revenue is influenced by different factors?

- How do the sales volume in each store relate to company's revenues?

- ✓ 5 stores contributing 71% revenue and 60% volume and remaining 11 stores contributing remaining 30% revenue and 40% volume

[Note: You can calculate total sales, proportion of sales and cumulative proportion of sales by each store. The same metrics can be created for the volume. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create report.]

- How does revenue/volume depend on the configuration?

- ✓ Revenues are contributing from across all configuration however maximum coming from average configuration.

[Note: You may choose to create buckets for the different configuration parameters(e.g: RAM can be divided into 3-4 buckets like High/Medium/Low RAM using percentiles/quartiles as cut-offs) and you can calculate total sales , Total volume, proportion of sales and proportion of volume by each configuration group or combination of the groups. For this, you may require to use proc tabulate and/or proc means and/or proc report and/or proc sql procedures to create reports.]

- What statistical technique should be applied to predict the sales of the company in 2010?

- ✓ We can apply Time Series forecasting Technique. Last 7 months contributing 88% revenue.

- ✓ Since the data is only existed for 12 months, predicting 2010 sales will be challenging.

[Note: This is open question. You may not require to do any analysis however think about what possible patterns can be derived using different statistical techniques]

# Analytics Initiatives

- We can find key drivers for revenue or profit by store.
  - ✓ Y (DV)= Revenue or Profit or Growth Rate
  - ✓ X's (IV's) = Laptop Features, customer radius, discounts, Distance, nearest store distance etc
  - ✓ Technique = Linear Regression

[Note: You may think different business problems and different techniques to solve with available data.]

# Recommendations

- ✓ **Since Maximum revenue is coming from customers who are around radius**
- ✓ **Discounts not influenced in sales of these stores**
- ✓ **Product price reduction of older configurations at stores generating most sales.**
- ✓ **Concentrate product inventory of newer configurations initially at stores generating most sales to maximize appeal to customer location preferences\***
- ✓ **Consider opening additional stores at mid point locations where customers travel a great distance. This however will be based on evaluating the cost/benefit associated with it.**

**[Note: These are some of the sample recommendations to guide you. However you can think and provide Some more recommendations/inferences]**

\*This is a hypothesis based on initial data pointing to it. It needs to be validated with further validation by comparing distance travelled by consumers over stores.

# Further analytics and relevant data capturing to shape business strategy....

## **DATA QUALITY:**

Further transactional information, store information and cost information will increase the value of data to achieve the business objectives such as–

- discounts
- whether product on sale
- whether salesman assisted in sale
- No of salesmen in store
- Cost information

## **FURTHER ANALYTICS:**

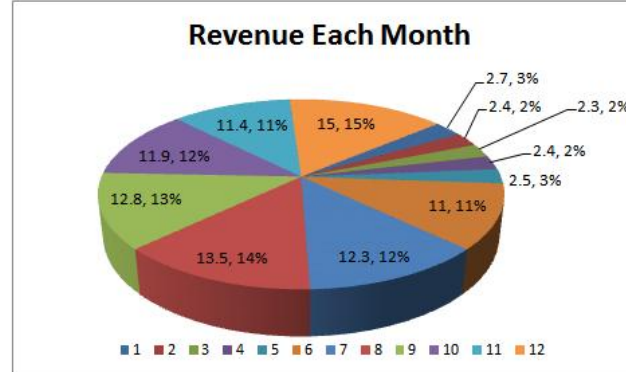
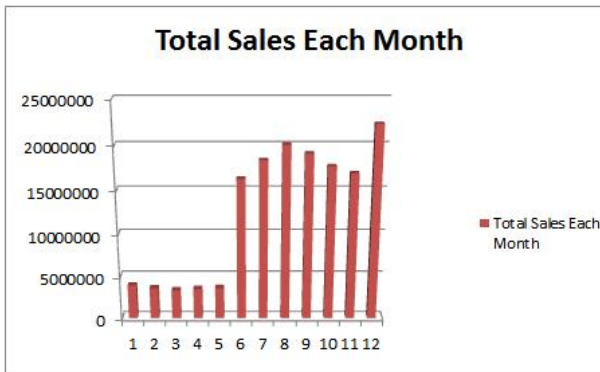
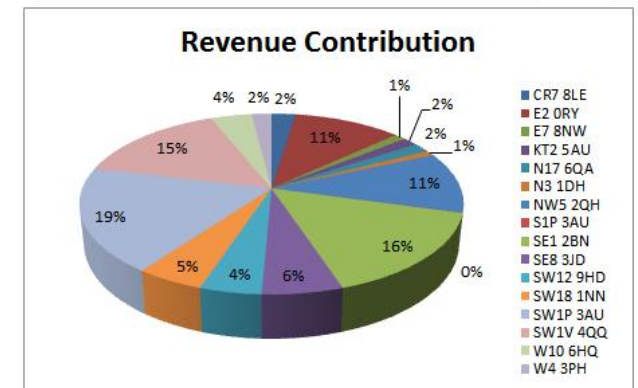
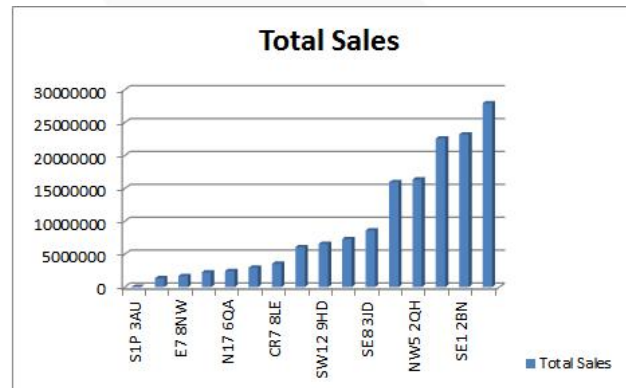
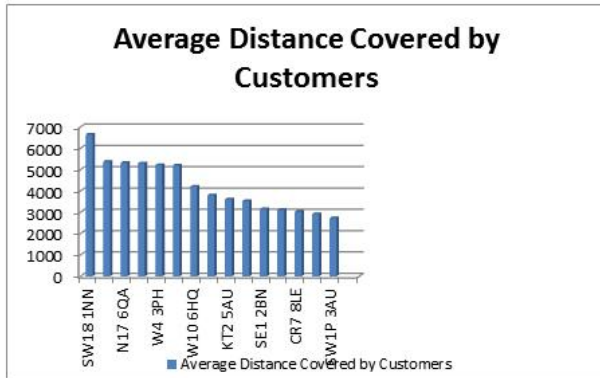
- Use of the above mentioned data to perform regression modelling will assist in understanding the key drivers of price and sales.
- The regression modelling can be used to predict sales in 2010.
- With further information gathered, price elasticity modelling can be developed to identify optimal pricing points

**[Note: You may have limitations with answering business problems with available data. What other data can be helpful to answer other business problems]**

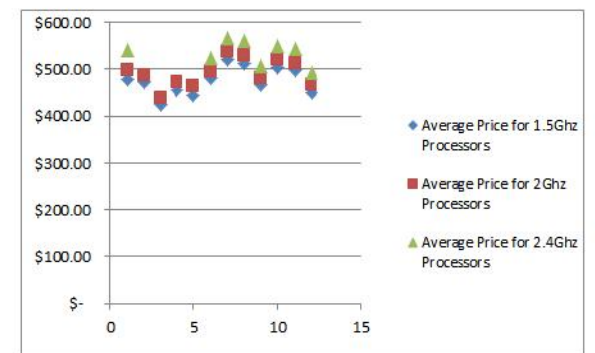
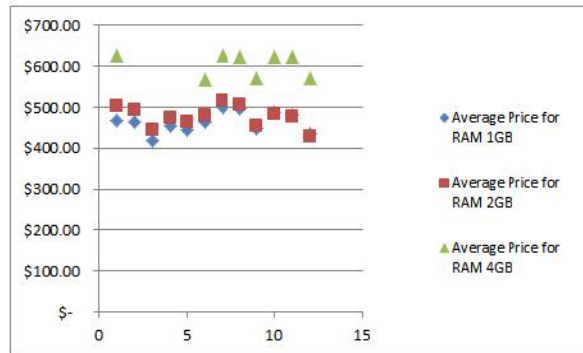
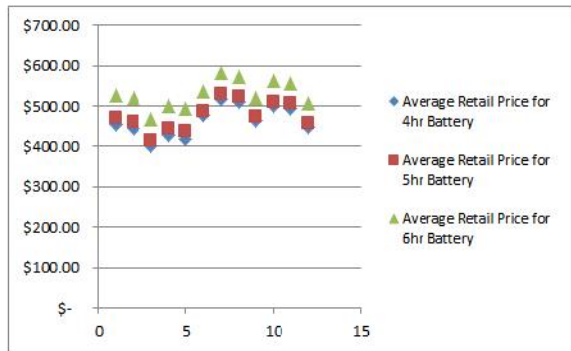
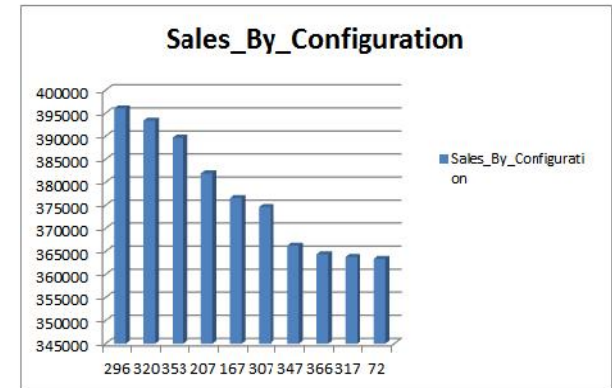
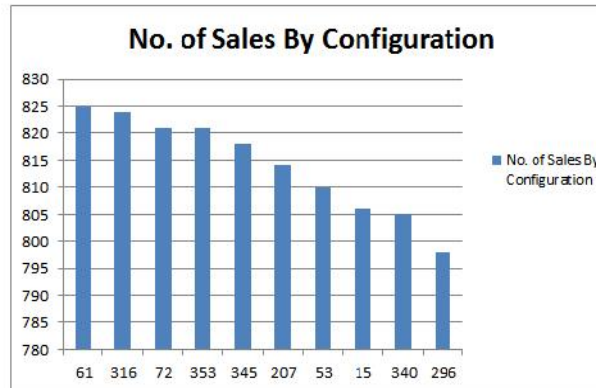
# Sample graphs

[Note: All graphs created using dummy data and provided them to help you with different visualizations for answering questions]

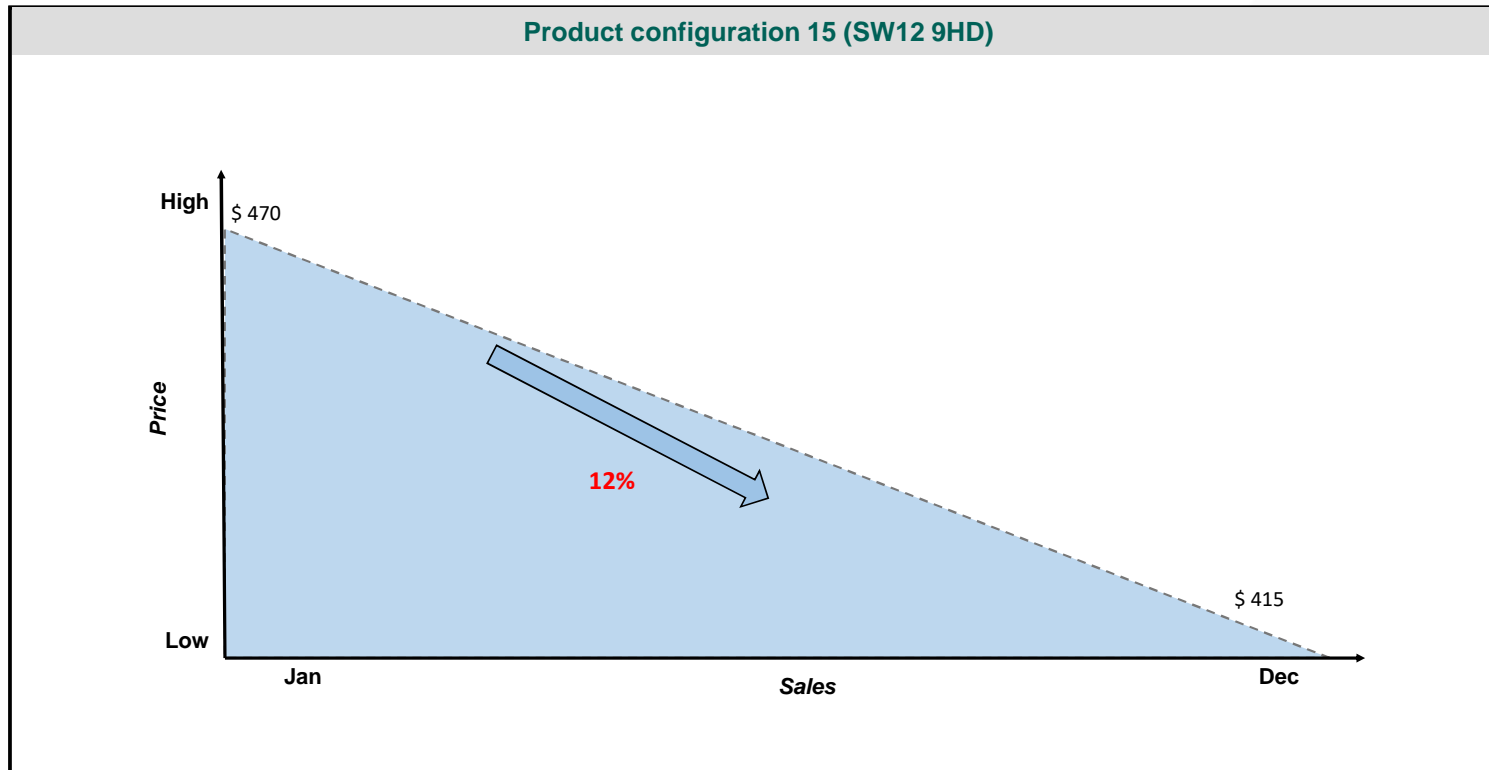
# Sample Graphs-1



# Sample Graphs-2



## Sample Graph-3

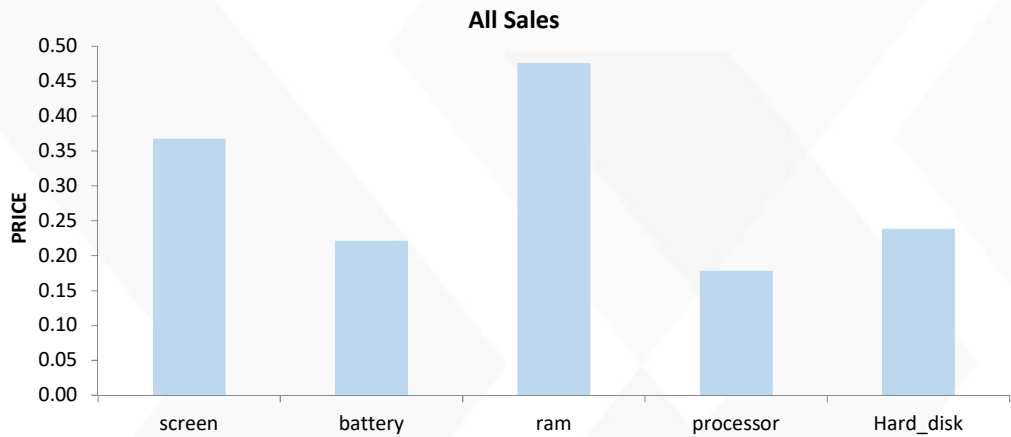
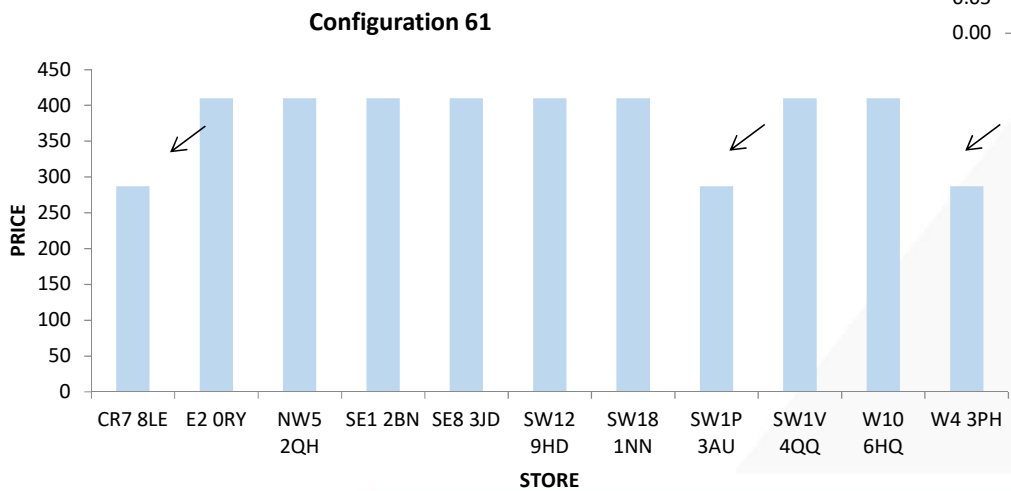


Price decreases over time and the theme is consistent across stores suggesting declining value of existing configuration.



# Sample Graphs-4

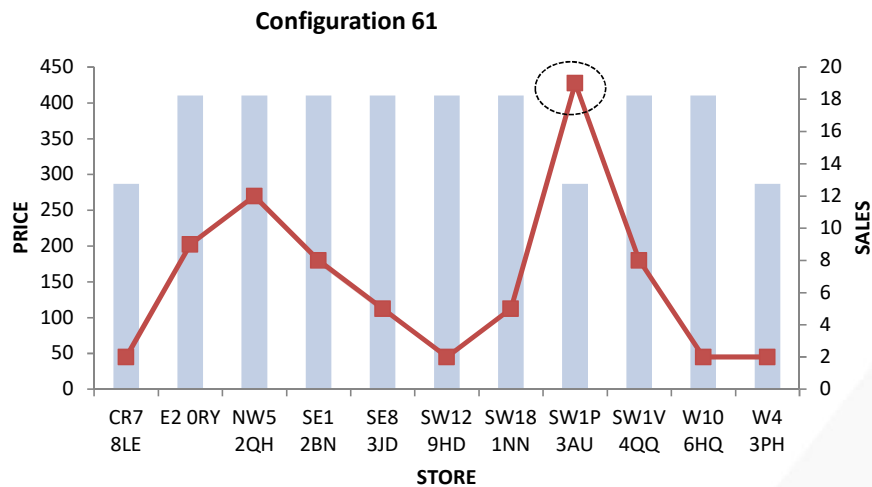
- Increased configuration has a positive influence on pricing



- Pricing varies across stores

# Sample Graphs-5

- Sales are concentrated in a few stores. Sales vary even with similar pricing.



- Sales increase with lower pricing at certain stores

Sales across stores doesn't comprise of all product configurations suggesting that every product is not available in every store.

# Contact Us

Visit us on: <http://www.analytixlabs.in/>

For more information, please contact us: <http://www.analytixlabs.co.in/contact-us/>

Or email: [info@analytixlabs.co.in](mailto:info@analytixlabs.co.in)

Call us we would love to speak with you: (+91) 9910509849

Join us on:

Twitter - <http://twitter.com/#!/AnalytixLabs>

Facebook - <http://www.facebook.com/analytixlabs>

LinkedIn - <http://www.linkedin.com/in/analytixlabs>

Blog - <http://www.analytixlabs.co.in/category/blog/>