DS-853 Data Project
Author- Divya Raghunathan

1 Introduction - dataset

The dataset (3 rows shown in Table 1) contains details of the temporary price promotion of an item identified by the item number (SKUID). Each unique SKUID corresponds to two observations entry lift and exit lift. The sequence of the lifts and sales is represented below.

Pre-TPR baseline sales $\xrightarrow{\textit{Entry lift}}$ Promo sales $\xrightarrow{\textit{Exit lift}}$ Post- TPR sales

Entry lift : Represents ratio of promo sales to pre-TPR baseline sales. The corresponding discount percentage is mentioned when the items moves from Pre-TPR to promo sales.

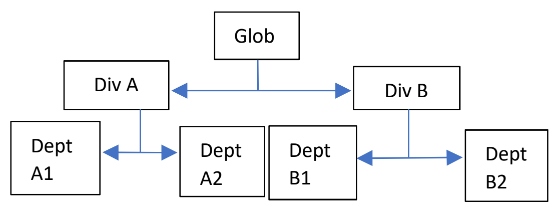
Exit lift : Represents ratio of Post-TPR sales to promo sales. The corresponding discount percentage is mentioned when the item moves back up from promo to post-TPR sales. Hence, the value is negative (moves from lower price to higher price).

Each observation is identified by division and department it belongs.

Table 1 : The first three rows of the dataset

obs	week	SKUID	Div	Dept	Sales$	Discount	Lift (Y)
1	5/5/18	100289	B	B2	$ 433	20%	1.19
2	3/10/18	100290	A	A1	$ 948	20%	1.48
3	3/10/18	100290	A	A1	$ 15,863	20%	1.26

Figure 1 : Hierarchy of store levels



2 List of Symbols, Abbreviations and Nomenclature

L = Lift (Response variable)
D= Discount (Predictor variable)
PE = Price elasticity (Parameter)
WSSE = Sales – Weighted Squared errors
$\hat{L}$ = Predicted lift
MAPE= Mean Absolute Percent Error
S=Sales
TPR= Temporary price promotion

3 Objective

The objective of the project is to find at which retail level (figure 1) – Department, Division , Global can we predict retail demand promotional lifts for each department the most accurately (least error).

Method : By calibrating the price elasticity model on the training data for different levels, we can evaluate the model on the department level using Sales weighted mean Absolute Percent error (MAPE) on the validation data.

4 Analysis

Splitting the dataset : Training and Validation

The dataset has been divided into two sets – Training set and Validation set. The sets are divided randomly using the “RAND” function in excel. Each observation is identified as “T” or “V” in the column “T/V”. The details of the dataset and the split is given in Table 2.

Table 2 : Training and validation split

Total number of observations	12915
Training set observations	6375
Validation set observations	6540

5 Price Elasticity Model (Training data)

We will be calibrating the model using the training dataset for this section. The objective of the price elasticity model is to obtain the price elasticities at 7 different levels -Global, Division- A, B, Department- A1, A2, B1, B2.The final results of WSSE and PE are mentioned in table 3 . The breakdown of the method to derive the results in table 3 is as below:

Step 1 : Predicted lift for each observation is calculated using :

$$\boxed{\hat{L} = e^{PE \cdot D}}$$

Step 2: Sales-weighted squared error for each observation is calculated using :

$$\boxed{\text{Sales-weighted squared error} = S_i \cdot (L_i - \hat{L}_i)^2}$$

Step 3: Sum of WSSE at different level is calculated as per chart 1 and is mentioned in table 3

$$\boxed{WSSE = \Sigma_i S_i \cdot (L_i - \hat{L}_i)^2}$$

Step 4: The PE for each level is obtained using Solver by minimizing the WSSE for the respective level.

Note: Predicted lift is calculated using a dummy PE value of 1 initially. Later, PE value changes when solver is applied.

I have used a single array formula to find WSSE, the formula satisfies steps 1-3. The results of WSSE is given in table 3. PE is obtained using step 4 and results are given in table 3.

Note: To calculate WSSE for global, we use all the training observations. To calculate WSSE for division and departments WSSE , the values are filtered based on the corresponding level. Only the training data is used for calculating WSSE and PE.

Table 3: WSSE and PE- Training data

	Weighted Sum of squared error	Price Elasticity
Global	5333442.544	2.042469039
Division A	1963050.73	1.959710502
Division B	3365643.741	2.085690792
Department A1	1218773.451	2.062893697
Department A2	736263.3495	1.785140884
Department B1	2416018.521	2.900167251
Department B2	713452.0063	1.755951060

6 Mean Absolute percent error (Validation data)

We will be validating the model using the validation set for this section. The purpose of this section is to calculate three MAPE values for each department (A1,A2,B1,B2) using the PE for the corresponding department, for the corresponding department's parent division and the global PE (as per table 3).

Step 1: For each department level, I will be calculating the MAPE value as per the formula below.

$$\mathrm{MAPE} = \frac{\Sigma i Si \cdot |Li - L\hat{i}|}{\Sigma i Si}$$

Step 2: Pick the lowest MAPE for each department. This is the one with the least Mean Absolute percentage error.

Similar to section 2, I have used an array formula to obtain the MAPE results as given in Table 4. I have made use of conditional formatting in excel to highlight the lowest MAPE value for each department. The highlighted green cell represents the lowest MAPE value for each department.

7 Results

Table 4: Validation MAPE using PE

Validation MAPE using PE			
Dept	**Dept**	**Division**	**Global**
A1	35.08%	34.50%	34.95%
A2	29.35%	30.50%	31.19%
B1	45.97%	44.42%	44.51%
B2	26.78%	29.02%	28.61%

8 Conclusion:

1. We can predict lifts for departments A1 and B1 using PE at the division level
2. We can predict lifts for departments A2 and B2 using PE at their respective department level