

Big Mart Sales – Promotion, TPR, Display Data Analysis

1. Action Plan

Data importing and merging

- After importing all 3 spreadsheets my plan is to inner join transaction and product data-frame on 'upc' variable storing it into a temporary data-frame and then Inner join this temporary data-frame with store data-frame on column 'store_num'.

Data Pre-processing

- I dropped missing values from 'price' and 'base_price' column and dropped rows with product category as oral hygiene
- Looking at the dataset I can see that 'week_end_date' has transactions per weeks. Hence I have split week_end_date into 4 columns year, month, day, week using lubridate library. Here week is the seasonality variable which is cycled per year.
- Converted store_num, upc, feature, display, tpr_only, manufacturer, category, sub-category, segment, city, week and upc into factor.

Statistical Modelling

- I built a one **fixed effect model** with DV as spend and two **Quasipoisson** models with DV's Units and HHS as both are count variables. Also I built two **mixed level random effect** model with DV spend and units to analyse products with lowest spend and units sold.
- I included predictors along with interaction terms between category*feature, category *display, category*tpr_only, segment*feature, segment*display, and segment*tpr_only. To gain insights about varying marginal effects as per product category and store segment.

Assumption Testing.

- Conducted **Dispersion test** and checked **VIF values** for quasi-poisson model. Both models are over dispersed but Quasipoisson works well with over dispersed model.

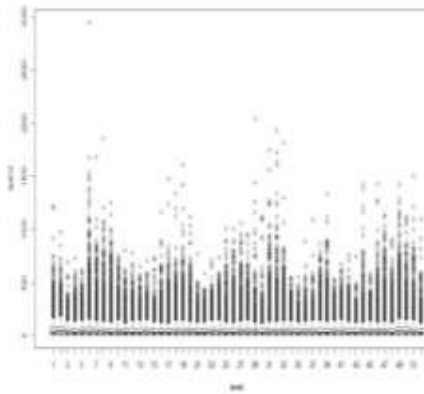
2. Relevant Predictor Table

The table below display the predictors used in the analysis along with its impact, explanation and relevance in determining spend, units and hhs.

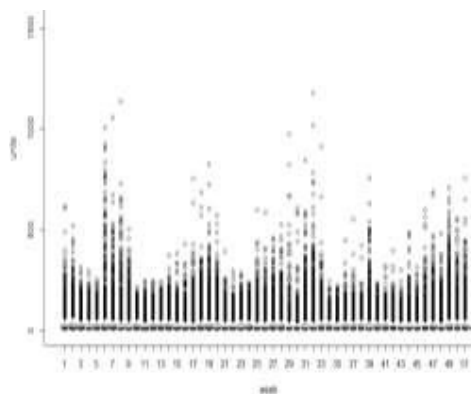
Predictors	Impact	Used for DV's	Explanation/Rationale
visits	+	spend, units, hhs	Increase in store visits can increase product spends, units sold and hhs.
hhs	+	spend, units	Increase in hhs can result in more people buying the products thus increasing spend and units sold.
price	+/-	spend, units	Price can affect spend and units in both ways as some customers may prefer low price products whereas some may prefer specific products irrespective of its price.
feature	+/-	spend, units, hhs	If a product is featured in in-store circular then it could grab attention of the people leading to higher spend, hhs and units.
display	+/-	spend, units, hhs	If a product is a part of promotion display then it will get sold more in higher quantity
tpr_only	+/-	spend, units, hhs	If a store has high number of visits then product with temporary reduction might get sold more with higher units.
category	+/-	spend, units, hhs	Depending upon store segment and manufacturer customers may prefer different product categories and hence can have both side effects on all three DV's
segment	+/-	spend, units, hhs	Depending upon store segment and store location people may prefer different products and hence can have both side effects on all three DV's
units	+/-	Spend, hhs	Higher number of units will result into more people buying it but this effects depend on the price as well.
Spend	+/-	hhs	More the product spend, can result in more customers buying it depending upon the product price and category.

All other feature variables are not directly related to the questions which we are interested in and hence I have excluded them from my analysis.

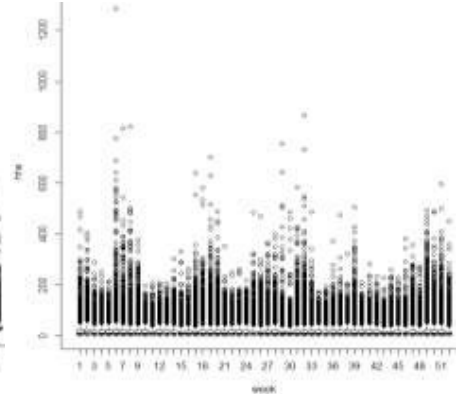
3. Exploratory Data Analysis.



Spend vs Weeks



Units vs Week



HHS vs Week

4. Models. Fixed Effect Model

```
model_spend_1 <- lm(spend ~ week*feature + week*tpr_only + week*display + price + visits + hhs + units + category*feature +  
category*display + category*tpr_only + segment*feature + segment*display + segment*tpr_only + price*upc, data=train)
```

Quasi-poisson model

```
qpoisson <- glm(units ~ feature + tpr_only + week*display + visits + log(hhs) + price + category*feature + category*display +  
category*tpr_only+segment*feature+segment*display+segment*tpr_only,family=quasipoisson(link=log),data=master.dataset)
```

```
qpoisson_hhs <- glm(hhs ~ feature + tpr_only + display + visits + spend + category*feature + category*display +  
category*tpr_only + segment*feature + segment*display + segment*tpr_only, family=quasipoisson (link=log),  
data=master.dataset)
```

Multilevel Random Effect Model

```
model_spend_2 <- lmer(spend ~ feature + tpr_only + display + price + visits + hhs + units+ category + segment+ (1 | upc) ,  
data=df_filtered)
```

```
model_units_2 <- lmer(units ~ feature + tpr_only + display + price + visits + hhs + price+ category + segment+ (1 | upc) ,  
data=df_filtered)
```

Stargazer snapshot: Please note, I have not included entire stargazer output because it was too big.

	spend OLS (1)	units glm: quasipoisson link = log (2)	hhs glm: quasipoisson link = log (3)
featured1	-13.221*** (0.402)	-0.022*** (0.003)	0.178*** (0.010)
tpr_only1	-3.936*** (0.205)	0.095*** (0.002)	-0.071*** (0.007)
display1	-12.672*** (0.197)	-0.015*** (0.001)	0.824*** (0.005)
price	-25.044*** (0.893)	-0.030*** (0.0004)	
visits	2.110*** (0.025)	0.0002*** (0.00001)	0.001*** (0.00004)
hhs	-1.075*** (0.023)		
units	1.299*** (0.006)		
log(hhs)		0.991*** (0.0005)	
spend			0.002*** (0.00002)
categoryCOLD CEREAL	-121.712*** (1.879)	-0.010*** (0.001)	0.372*** (0.003)

1. What is the effect of promotions, displays, or being featured in the circular on product sales (spend), unit sales, and number of household purchasers? (3 points)

DV	Feature	Display	TPR_only
Spend	If a product is featured in a in store circular then spend decreases by \$13.22	If a product is displayed in store promotion display then spend decreases by \$12.67 units	If a product has temporary price reduction then spend decreases by \$3.935 units
Units	If a product is featured then its units sold is low by 2.198% than if it is not featured	If a product is displayed then its units sold is low by 1.467% than if it is not displayed	If a product is under TPR then its units sold is higher by 9.48% than if it is not under TPR
HHS	If a product is featured then its purchasing households is higher by 17.78% than if it is not featured	If a product is displayed then its purchasing households is higher by 82.44% if it is not displayed	If a product is under TPR then its purchasing households is lower by 7.07% than if it is not under TPR

2. How do the above effects vary by product categories (cold cereals, frozen pizza, bag snacks) and store segments (mainstream, upscale, value)? (2 points)

Effects varying by Product Category

DV	Feature	Display	TPR_only
Spend	Product of category Frozen Pizza if featured has \$27.22 units sales more than Bag Snacks. Product of category Cold Cereal if featured has \$23.58 units sales more than Bag Snacks.	Product of category Frozen Pizza if displayed has \$30.61 units sales more than Bag Snacks. Product of category Cold Cereal if displayed has \$16.54 units sales more than Bag Snacks.	Product of category Frozen Pizza if under TPR has \$2.35 units sales more than Bag Snacks. Product of category Cold Cereal if featured has \$7.55 units sales more than Bag Snacks.
Units	Frozen Pizza category products if featured have 1.763% higher units sold than bag snacks Cold Cereal category products if featured have 10.19% higher units sold than bag snacks	Frozen Pizza category products if displayed have 1.61% higher units sold than bag snacks Cold Cereal category products if displayed have 2.27% higher units sold than bag snacks	Frozen Pizza category products if under TPR have 11.08% lower units sold than bag snacks Cold Cereal category products if under TPR have 7.388% lower units sold than bag snacks

HHS	Frozen Pizza category products if featured have 33.17% higher purchasing households than bag snacks	Frozen Pizza category products if displayed have 23.07% lower purchasing households than bag snacks	Frozen Pizza category products if under TPR have 44.13% higher purchasing households than bag snacks
	Cold Cereal category products if featured have 16.56% higher purchasing households than bag snacks	Cold Cereal category products if displayed have 46.96% lower purchasing households than bag snacks	Cold Cereal category products if under TPR have 35.93% higher purchasing households than bag snacks

Effects varying by Store Segment

DV	Feature	Display	TPR_only
Spend	Upscale segment stores has \$0.45 more spend than Mainstream stores if their products are featured. Value segment stores has \$9.15 less spend than Mainstream stores if their products are featured.	Upscale segment stores has \$1.16 less spend than Mainstream stores if their products are displayed. Value segment stores has \$4.8 more spend than Mainstream stores if their products are displayed.	Upscale segment stores has \$3.22 less spend than Mainstream stores if their products are under TPR. Value segment stores has \$1.44 more spend than Mainstream stores if their products are under TPR.
Units	Upscale stores if featured have 1.402% higher units sold than Mainstream stores Value stores if featured have 4.47% lower units sold than Mainstream stores	Upscale stores if displayed have 0.116% lower units sold than Mainstream stores Value stores if displayed have 3.895% lower units sold than Mainstream stores	Upscale stores if under TPR have 0.718% higher units sold than Mainstream stores Value stores if under TPR have 4.395% lower units sold than Mainstream stores
HHS	Upscale stores if featured have 6.5% lower purchasing households than Mainstream stores Value stores if featured have 26.45% higher purchasing households than Mainstream stores	Upscale stores if displayed have 6.15% lower purchasing households than Mainstream stores Value stores if displayed have 10.05% higher purchasing households than Mainstream stores	Upscale stores if under TPR have 0.36% higher purchasing households than Mainstream stores Value stores if under TPR have 7.47% lower purchasing households than Mainstream stores

3. What are the five most price elastic and five least price elastic products? Price elasticity is the change in sales for unit change in product price? (3 points)

Products with UPC code **1600027527, 3000006560, 88491212971, 3800031838 and 3000006610** are **most price elastic products**.

Products with UPC code **7192100339, 1111009507, 7192100337, 7192100336 and 7218063979** are **least price elastic products**.

4. As the retailer, which products would you lower the price to maximize (a) product sales and (b) unit sales, and why? (2 points)

Interpreted following results using two mixed level random effect model mentioned above in model section.

- Product with UPC code **1111009477, 1111009497, 1111085319** have lowest spend and hence we can either consider promotional offers for them or we can decrease their price in order to generate demand and increase sales.
- Product with UPC code **1111009497, 3800031838, 2840004768** have lowest units sold and hence we can either consider promotional offers for them or we can decrease their price in order to generate demand and increase number of units sold.

