**Final Exam – Retail Analytics Project**

This project is interesting because it allowed me to put together the entire curriculum and put it into practice in a real-life example. It also let me see where I had some shortcomings on this topic.

Looking at the different fields, I would think they would have different effects on spend (sales), units sold, and number of purchasing households:

|  |  |
| --- | --- |
| Display | This should increase all 3 measures since it would attract attention to the promotion |
| Feature | This should increase all 3 measures, proportionally to the number of people reading it |
| Price | The lower the price, the more people should buy, depending on the elasticity of the product |
| Base Price | This probably has no influence on customers’ behavior |
| Tpr\_Only | This should increase all 3 measures, esp. on the most elastic products |
| Visits | The more baskets including the product, the more the spend and the units sold should increase. It may not have an effect on purchasing households since they may just include more product in one basket |
| Week\_End\_Date | This may increase all 3 measures if some special event is happening that week (holiday, etc.) |
| Upc | This may have an impact on the measures, but it will probably depend on external factors (brand advertisement, popularity, etc.) |
| Store\_Num | This probably would affect all 3 measures, but each store may have a unique impact on these. Each store and its characteristics will bring heterogeneity to the data |
| Category | The category itself may not impact the measures, but this category will probably be a source of heterogeneity in the data. |
| Sub\_Category | Same as above |
| Description | N/A |
| Manufacturer | This will be tied to the UPC and the brand/manufacturer's appeal will have an impact on all 3 measures |
| Product\_Size | This may impact the 3 measures positively since larger-sized products tend to be cheaper per oz (all our products are sized in oz) |
| City/State/MSA | These are measures that will make the data vary based on the city's characteristics |
| Avg\_Weekly\_Baskets | I am not sure of the impact of that information on the 3 measures we are tracking |
| Parking | This info is probably tied to the size of the store and both are probably not necessary |
| Segment | This may influence the 3 measures depending on how many other stores in the area this company has to compete with in the same segment. |
| Size | The size of the store may influence the 3 measures positively since there may be more choices. |

1. Data Clean-Up

To be prepared for the analysis, I followed these steps:

Products:

* I removed the products classified as Oral Hygiene Products
* I removed the “OZ” from the size of the product to allow this to be a number

Stores:

* I removed the duplicate store names. I had to make an assumption on which segment to keep and which to remove. I used the average weekly basket information to help me.
* I removed the parking size information. This had too many empty values and the size of the store was already available and both were probably highly correlated with each other.

Transactions:

* A few records were missing the base price or the price. I removed these.
* I added columns for the year, month, quarter, and week id based on the “week end date”

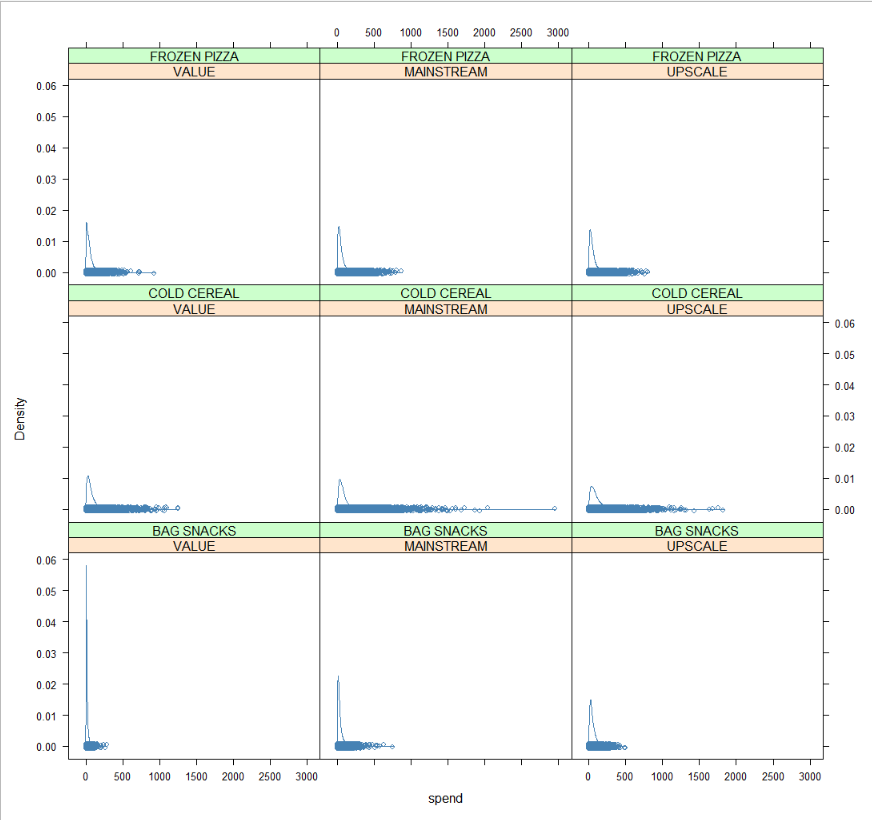
I then merged each data frame into one. I also then created a new variable “price per oz” in case this would have an impact later in the analysis.

1. Data Visualization and Explorations

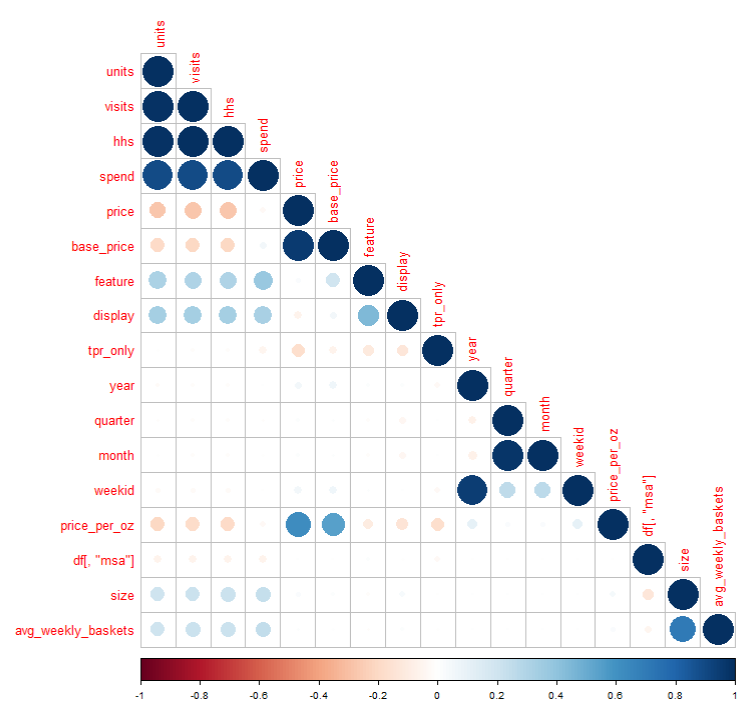
I did check whether there was autocorrelation of each dependent variable and did not find any. It doesn’t seem that there is a clear seasonality trend in consumptions at these stores. Some weeks do seem to be higher than others, but it doesn’t seem to be related to the previous week’s consumption.

Looking at the data per product, per week, per store, we can see that it skewed right with the majority of sales below $100 $(64 on average, with a median of $43), the number of units sold below 40 (median of 15 items), and the number of purchasing households averaging 20.

The heterogeneity that we expected certain data to introduce was confirmed by plots. For instance, looking at sales (spend), we can see that bag snacks have a very high probability of being a small spend at value stores, while the probability for higher spend increases at mainstream stores, only to decrease a little at upscale stores. For cold cereal, however, we can see that the probability of spend is higher at value store than bag snacks, but still lower than at mainstream stores. Here again, the upscale store has a probability of spend higher than at value stores, but lower than at mainstream stores. This changes for frozen pizzas. For that category, the segment of the store doesn’t seem to matter much. We see a similar behavior for the units sold and the number of purchasing households. Other plot techniques confirm this finding.



As expected, a few of the fields are highly correlated. This includes our three dependent variables as well as visits. This is logical since the amount spent, the number of unique baskets containing the product (visits), the number of purchasing households, and the number of units sold are different ways to assess a product’s sales performance. In addition, year and week id as well as quarter and month, are highly correlated. The store size and its average weekly baskets seem to be highly correlated. This makes sense since a larger store would also provide more choices to its customers and larger stores were probably planned in proportion to the area’s spending power and need. We do see a small positive relationship between feature and display with our dependent variables.



Questions:

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

When doing the analysis with just focusing on these three predictors, we can see that all three promotions methods are significant. Picking the best of the PLM models (fixed effects) for spends, we can see that only about 20% of the variation in sales is tied to these promotions at an average store. The same method is best to help explain variations in units and purchasing households, but these variables only explain 16% of the variation in units sold, and just about the same for purchasing households. We will have to look at more models to be able to provide more specific recommendations.

Interestingly, TPR-only seems to have a negative effect ($1.49) while displays ($42.4) and feature (75.9) both improve the weekly, per product, per store average sales. On the other hand, all three promotion methods improve the number of units sold and purchasing households. The impact is still most for both feature and display (an improvement of about 24 units and 19 purchasing households for both), but TPR-only does improve the number of units sold per product, per week, per store on average by 4.32 and by an average of 2.79 for purchasing households.

This may indicate that the feature and display may have a bigger impact than TPR-only promotions. One possible explanation is that customers don’t realize as much the savings in TPR-only as they do for features and display promotions.

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

As noted above, it seems that category and segments are some of the variables that we need to consider to handle the heterogeneity of this chains’ transactions data and provide more specific information.

Once we consider the additional levels, we can see that all promotions have a positive effect on spend, units, and purchasing households. The only negative impacts are mainstream stores which seem to decrease spend by about $1.8 per product per week as compared to value stores and price increases that decrease unit sold on average by 4.5 and purchasing household by 3.5. Since spend is positively impacted in our fixed-effects model, we can imagine that this drop is not enough to impact the weekly spend for products. This figure was also validated using lmer with coefficients rounding to the same value.

For mainstream stores, they, on the other hand, improve units sold (as compared to value stores) by 0.43 and purchasing households by 0.74.

1. What are the five most prices elastic and five least price-elastic products? Price elasticity is the change in sales for a unit change in product price.

To evaluate elasticity, we ran regression analysis at the UPC level, taking into accounts each store. This had a few items that came up as benefiting from price increases. These items may be the same ones that caused the overall results to look like price increases were not as detrimental to sales as expected.

The most price elastic articles in regards to spend are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **upc** | **description** | **manufacturer** | **size\_in\_oz** | **elasticity** |
| 3800039118 | KELL FROOT LOOPS | KELLOGG | 12.2 | -3.426 |
| 7192100336 | DIGIORNO THREE MEAT | TOMBSTONE | 29.8 | -3.136 |
| 7218063979 | FRSC PEPPERONI PIZZA | TONYS | 27.35 | -3.116 |
| 7218063052 | FRSC BRCK OVN ITL PEP PZ | TONYS | 22.7 | -2.997 |
| 7192100337 | DIGRN SUPREME PIZZA | TOMBSTONE | 32.7 | -2.743 |

The least price elastic articles in regards to spend are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **upc** | **description** | **manufacturer** | **size\_in\_oz** | **elasticity** |
| 1111085350 | PL BT SZ FRSTD SHRD WHT | PRIVATE LABEL | 18 | -0.635 |
| 1111009497 | PL PRETZEL STICKS | PRIVATE LABEL | 15 | -0.278 |
| 1111009507 | PL TWIST PRETZELS | PRIVATE LABEL | 15 | -0.253 |
| 1111085319 | PL HONEY NUT TOASTD OATS | PRIVATE LABEL | 12.25 | -0.236 |
| 1111085345 | PL RAISIN BRAN | PRIVATE LABEL | 20 | -0.149 |
| *1111009477* | *PL MINI TWIST PRETZELS* | *PRIVATE LABEL* | *15* | *0.089* |
| *1600027564* | *GM CHEERIOS* | *GENERAL MI* | *12* | *0.267* |
| *7027312504* | *SHURGD PRETZEL RODS* | *SHULTZ* | *12* | *0.583* |
| *7027316204* | *SHURGD MINI PRETZELS* | *SHULTZ* | *16* | *0.647* |
| *7027316404* | *SHURGD PRETZEL STICKS* | *SHULTZ* | *16* | *0.754* |

I left the suspicious products in italics because while a positive elasticity is probably not accurate, they are still probably be the least elastic.

For instance, Kellogg’s Fruit Loops is the most elastic. A 1% increase in price, sale of the product will decrease by 3.42%. On the other hand, a 1% increase in the price of Private Label Raisin Bran will only impact its sale by 0,09%.

The most price elastic articles in regards to units sold are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **upc** | **description** | **manufacturer** | **size\_in\_oz** | **elasticity** |
| 3800039118 | KELL FROOT LOOPS | KELLOGG | 12.2 | -4.426 |
| 7192100336 | DIGIORNO THREE MEAT | TOMBSTONE | 29.8 | -4.136 |
| 7218063979 | FRSC PEPPERONI PIZZA | TONYS | 27.35 | -4.116 |
| 7218063052 | FRSC BRCK OVN ITL PEP PZ | TONYS | 22.7 | -3.997 |
| 7192100337 | DIGRN SUPREME PIZZA | TOMBSTONE | 32.7 | -3.743 |

The least price elastic articles in regards to units sold are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **upc** | **description** | **manufacturer** | **size\_in\_oz** | **elasticity** |
| 1111009477 | PL MINI TWIST PRETZELS | PRIVATE LABEL | 15 | -0.911 |
| 1600027564 | GM CHEERIOS | GENERAL MI | 12 | -0.733 |
| 7027312504 | SHURGD PRETZEL RODS | SHULTZ | 12 | -0.417 |
| 7027316204 | SHURGD MINI PRETZELS | SHULTZ | 16 | -0.353 |
| 7027316404 | SHURGD PRETZEL STICKS | SHULTZ | 16 | -0.246 |

The impact on unit sold mimic the one on spend. In this case, the bottom 5 match the one that appeared to have positive price elasticity. Here, we can see that an increase in price by 1% for Shultz’s Shurgd Pretzel Sticks, units sold of that product would only drop by 0.25%. Similarly, the decrease in units sold of Kellogg’s Fruit Loops would be 4.426% for each percentage increase in price.

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

To maximize product and unit sales, dropping the price of the most elastic articles should have the most returns since a 1% drop in price should result an increase of about 4% on average in units sold and of about 3.5% in revenue for these 5 products.

Doing promotions on the least elastic products will most likely not change demand or revenue by much. Their sales would be more resilient to price increases.

I really enjoyed working on this project. In all, I probably spent about 25-28 hours on it. Some items were more painful than others, but figuring out the elasticity was actually easier than I thought when I started breaking up the process and building on it. I do think this is the most challenging class in the curriculum so far. In the other classes, we were more spoon-fed the homework so we really just had to rehash what was presented. This made us feel we had a good command of what was covered. This class was a wake-up call. Here, we were faced with whether or not we truly understood it (or at least thought we did) and could implement it. It wasn’t easy, but I am glad to have gone through it.