**A7. SNACK CHAIN - Swathi Banna – U53837601**

**Data Preprocessing:**

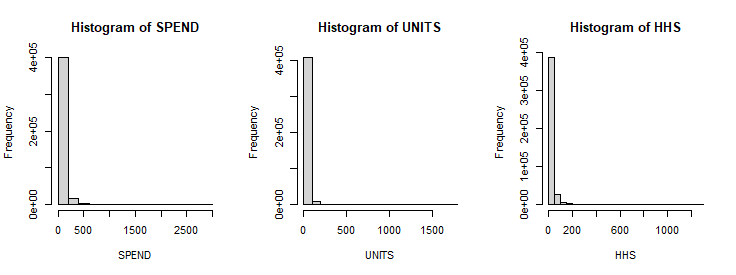
* 3 dataframes named stores, products and txns are created from the three sheets of SnackChain.
* Stores: There are 52 missing values in Parking. Since we don’t need this feature, it is dropped from the dataset. Two stores with IDs 4503 and 17627 have duplicates. Hence, duplicates are removed.
* Products: In the Description column, there are 5 duplicates. “GM CHEERIOS” has two quantities. So, named the product with smaller quantitiy as “GM CHEERIOS SMALL”. The remaining duplicates belong to oral hygiene products and hence not modified.
* Txns: In the transactions dataset, there are 185 missing values in Base Price and 23 in Price. These missing values have been removed.

***Stores****: Parking, City, State, MSA, Avg\_Weekly\_Basket, Store\_Name, Size*

***Products****: Manufacturer, Product\_Size, Sub\_Category*

* Dropped Columns:
* A new dataframe, total, is created by merging txns and stores datasets with common fields Store\_Num and Store\_ID respectively. After that, it is merged with products dataset with the common field UPC.
* From the total dataframe, the category Oral hygiene products is deleted.
* Now, there are 0.4 million observations with Spend, Units, HHS as the Dependent Variables and 12 feature variables.

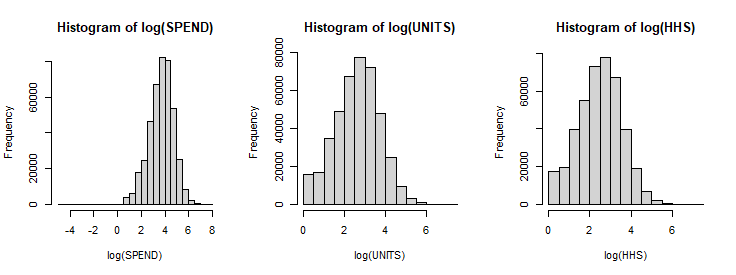
**Data Exploration:**

****

SPEND, UNITS AND HHS are not following a normal distribution.

-------------------------------------

The logs of SPEND, UNITS AND HHS variables are near to normal distribution.

****

> Hmisc::rcorr(as.matrix(temp))

**Correlation**:

There is no correlation between the numeric feature variables Visits and Price.

UNITS VISITS HHS SPEND PRICE

UNITS 1.00 0.99 0.98 0.90 -0.28

VISITS 0.99 1.00 1.00 0.90 -0.28

HHS 0.98 1.00 1.00 0.89 -0.28

SPEND 0.90 0.90 0.89 1.00 -0.03

PRICE -0.28 -0.28 -0.28 -0.03 1.00

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor Table** | |  |  |  |
| *DV: Sales, Units, HHS* | |  |  |  |
| **Predictor** | **Rationale** | **Sales** | **Units** | **HHS** |
| Visits | When the number of visits increase, sales increase. | + | + | + |
| Price | Sales increase with increase in price due the broader profit margin. However, number of units will reduce. | + | - | - |
| Feature | Sales increase when a product is featured in the circular. | + | + | + |
| Display | Sales increase when a product is displayed in-store | + | + | + |
| TPR | Sales increase when a product’s price is reduced in shelf-tag | + | + | + |
| Segment | Required for analyzing the effects on the store category | +/- | +/- | +/- |
| Description | Required for analyzing the price elasticity | +/- | +/- | +/- |
| Category | Required for analyzing the effects on the product category | +/- | +/- | +/- |

**1. Effects of product display, being featured in in-store circular, and temporary price reduction on product sales (spend), unit sales, and number of household purchasers.**

The data is multi-level data (Category and Segment). As the Independent Variables (Units and HHS) have a Poisson distribution, Poisson Random Effects model is used. But the models are over-dispersed. Since their log transformations are near to normal, LMER model is considered.

*re.spend=lmer(log(SPEND) ~ PRICE + DISPLAY + FEATURE + TPR\_ONLY + (1|CATEGORY) +*

*(1|SEGMENT),data=total, REML=FALSE)*

*re.units=lmer(log(UNITS) ~ PRICE + DISPLAY + FEATURE + TPR\_ONLY + (1|CATEGORY) + (1|SEGMENT),*

*data=total, REML=FALSE)*

*re.hhs=lmer(log(HHS) ~ PRICE + DISPLAY + FEATURE + TPR\_ONLY + (1|CATEGORY) + (1|SEGMENT),*

*data=total, REML=FALSE)*

stargazer::stargazer(nb.spend,nb.units,nb.hhs,type="text",single.row = TRUE)

========================================================================

Dependent variable:

----------------------------------------------------

log(SPEND) log(UNITS) log(HHS)

(1) (2) (3)

------------------------------------------------------------------------

PRICE 0.001 (0.001) -0.288\*\*\* (0.001) -0.257\*\*\* (0.001)

DISPLAY1 0.746\*\*\* (0.005) 0.766\*\*\* (0.005) 0.771\*\*\* (0.005)

FEATURE1 0.532\*\*\* (0.005) 0.511\*\*\* (0.005) 0.482\*\*\* (0.005)

TPR\_ONLY1 0.020\*\*\* (0.004) 0.051\*\*\* (0.004) 0.010\*\* (0.004)

Constant 3.481\*\*\* (0.292) 3.354\*\*\* (0.274) 3.113\*\*\* (0.281)

------------------------------------------------------------------------

Observations 408,201 408,201 408,201

Log Likelihood -517,636.100 -520,639.500 -508,888.200

Akaike Inf. Crit. 1,035,288.000 1,041,295.000 1,017,792.000

Bayesian Inf. Crit. 1,035,376.000 1,041,382.000 1,017,880.000

========================================================================

|  |  |  |  |
| --- | --- | --- | --- |
| **Marginal Effects** | | | |
|  | **Spend** | **Units** | **HHS** |
| Display  (If a product is part of in-store display) | The spend increases by 74.6% | The unit sales increase by 76.6% | The number of purchasing households increases by 77.1% |
| Feature  (If a product is featured in in-store circular) | The spend increases by 53.2% | The unit sales increase by 51.1% | The number of purchasing households increases by 48.2% |
| Temporary Price Reduction  (When a product’s price is reduced on the shelf-tag) | The spend increases by 2% | The unit sales increase by 5.1% | The number of purchasing households increases by 1% |

From the above data, we can infer that reducing the price doesn’t have an impact on the sales, unless the information reaches the customers through in-store display or in-store circular.

**2. Effects of display, feature, and TPR on SPEND vary by product categories and store segments.**

*re.segment=lmer(log(SPEND) ~ PRICE + DISPLAY\*SEGMENT+ FEATURE\*SEGMENT + TPR\_ONLY\*SEGMENT + (1|CATEGORY),data=total, REML=FALSE)*

Fixed effects:

Estimate Std. Error t value

(Intercept) 3.5112873 0.2302580 15.249

PRICE 0.0008378 0.0014837 0.565

DISPLAY1 0.7451307 0.0061255 121.643

SEGMENTUPSCALE 0.2669001 0.0040136 66.498

SEGMENTVALUE -0.3471958 0.0039009 -89.004

FEATURE1 0.5569547 0.0070086 79.468

TPR\_ONLY1 0.0569052 0.0056307 10.106

DISPLAY1:SEGMENTUPSCALE -0.0062692 0.0122862 -0.510

DISPLAY1:SEGMENTVALUE 0.0276848 0.0121937 2.270

SEGMENTUPSCALE:FEATURE1 -0.2366220 0.0136600 -17.322

SEGMENTVALUE:FEATURE1 0.0859543 0.0127333 6.750

SEGMENTUPSCALE:TPR\_ONLY1 -0.0036491 0.0110527 -0.330

SEGMENTVALUE:TPR\_ONLY1 -0.1478238 0.0102294 -14.451

*re.category=lmer(log(SPEND) ~ PRICE + DISPLAY\*CATEGORY+ FEATURE\*CATEGORY + TPR\_ONLY\*CATEGORY + (1|SEGMENT),data=total, REML=FALSE)*

Estimate Std. Error t value

(Intercept) 3.022698 0.141508 21.361

PRICE 0.004888 0.001486 3.289

DISPLAY1 0.842227 0.007423 113.456

CATEGORYCOLD CEREAL 0.961966 0.003812 252.340

CATEGORYFROZEN PIZZA 0.356307 0.006378 55.868

FEATURE1 0.173928 0.015693 11.083

TPR\_ONLY1 -0.091934 0.006647 -13.830

DISPLAY1:CATEGORYCOLD CEREAL -0.225424 0.011657 -19.337

DISPLAY1:CATEGORYFROZEN PIZZA -0.129810 0.011322 -11.465

CATEGORYCOLD CEREAL:FEATURE1 0.388602 0.017927 21.677

CATEGORYFROZEN PIZZA:FEATURE1 0.503394 0.017677 28.478

CATEGORYCOLD CEREAL:TPR\_ONLY1 0.140482 0.009434 14.891

CATEGORYFROZEN PIZZA:TPR\_ONLY1 0.303514 0.011415 26.589

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Marginal Effects on Segment** | | **Marginal Effects on Category** | |
|  | Upscale | Value | Cold Cereal | Frozen Pizza |
| Display  (If a product is part of in-store display) | The spend decreases by 0.62% than Mainstream | The spend increases by 2.7% than Mainstream | The spend decreases by 22.54% than Bag Snacks | The spend decreases by 12.98% than Bag Snacks |
| Feature  (If a product is featured in in-store circular) | The spend decreases by 23.66% than Mainstream | The spend increases by 8.5% than Mainstream | The spend increases by 38.86% than Bag Snacks | The spend increases by 50.33% than Bag Snacks |
| Temporary Price Reduction  (When a product’s price is reduced on the shelf-tag) | The spend decreases by 0.36% than Mainstream | The spend decreases by 14.78% than Mainstream | The spend increases by 14.04% than Bag Snacks | The spend decreases by 30.35% than Bag Snacks |

* When products are on display, Bag Snacks have the highest sales. This is because of the instantaneous trigger that arises when a snack is on display. Customers might not go specifically to the shelf in search of a snack. Hence, this category should be given priority while placing products on display.
* Cold Cereal and Frozen Pizza are like staple foods. Hence people tend to purchase more when they find an interesting deal in the in-store circular. The same goes true with the price reduction on the shelf-tag (TPR).
* The sales are not much influenced by the store segment. Mainstream has better sales than Upscale, when a product is featured and better than Value stores when there is a temporary price reduction.

**5 most price elastic and 5 least price elastic products.**

*Text

Description automatically generated Text

Description automatically generated*

|  |  |  |  |
| --- | --- | --- | --- |
| **Top Price Elastic Products** | **Price Elasticity** | **Least Price Elastic Products** | **Price Elasticity** |
| KELL FROOT LOOPS | -5.51 | RLDGLD BRAIDED HONEY WHT | -0.06 |
| PL SR CRUST 3 MEAT PIZZA | -4.01 | SNYDR FF MINI PRETZELS | 0.17 |
| FRSC PEPPERONI PIZZA | -3.67 | PL RAISIN BRAN | 0.3 |
| FRSC 4 CHEESE PIZZA | -3.15 | MKSL DUTCH PRETZELS | -0.32 |
| FRSC BRCK OVN ITL PEP PZ | -3.13 | SHURGD PRETZEL RODS | -0.46 |

|  |  |
| --- | --- |
|  |  |

**Assumptions:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **re.spend** | **re.units** | **re.hhs** |
| **Independence** | DurbinWatson test is not functioning for lmer models. Hence couldn’t test | | |
| **Multi Collinearity** | PRICE DISPLAY FEATURE TPR\_ONLY  1.102159 1.293968 1.289269 1.074688  **PASS** | PRICE DISPLAY FEATURE TPR\_ONLY  1.102157 1.293968 1.289268 1.074688  **PASS** | PRICE DISPLAY FEATURE TPR\_ONLY  1.102158 1.293968 1.289269 1.074688  **PASS** |

|  |  |  |
| --- | --- | --- |
|  | **re.segment** | **re.category** |
| **Independence** | DurbinWatson test is not functioning for lmer models. Hence couldn’t test | |
| **Multi Collinearity** | GVIF Df GVIF^(1/(2\*Df))  PRICE 1.103660 1 1.050552  DISPLAY 2.121521 1 1.456544  SEGMENT 1.806288 2 1.159302  FEATURE 2.210167 1 1.486663  TPR\_ONLY 1.834817 1 1.354554  DISPLAY:SEGMENT 3.182902 2 1.335690  SEGMENT:FEATURE 3.251161 2 1.342795  SEGMENT:TPR\_ONLY 2.228493 2 1.221808  **PASS** | GVIF Df GVIF^(1/(2\*Df))  PRICE 3.083019 1 1.755853  DISPLAY 3.139731 1 1.771928  CATEGORY 4.777837 2 1.478454  FEATURE 11.508514 1 3.392420  TPR\_ONLY 2.585387 1 1.607914  DISPLAY:CATEGORY 5.987093 2 1.564242  CATEGORY:FEATURE 20.240784 2 2.121079  CATEGORY:TPR\_ONLY 2.965045 2 1.312223  FEATURE has multicollinearity |

**Products for which the price has to be lowered to maximize (a) product sales and (b) unit sales, and why:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Products for maximizing Product Sales** | **Sales Elasticity** | **Products for maximizing Unit Sales** | **Units Elasticity** |
| KELL FROOT LOOPS | -4.81 | KELL FROOT LOOPS | -5.51 |
| PL SR CRUST 3 MEAT PIZZA | -3.24 | PL SR CRUST 3 MEAT PIZZA | -4.01 |
| FRSC PEPPERONI PIZZA | -2.99 | FRSC PEPPERONI PIZZA | -3.67 |
| PL BT SZ FRSTD SHRD WHT | 2.9 | FRSC 4 CHEESE PIZZA | -3.15 |
| PL HONEY NUT TOASTD OATS | 2.71 | FRSC BRCK OVN ITL PEP PZ | -3.13 |
| FRSC 4 CHEESE PIZZA | -2.45 | MKSL PRETZEL STICKS | -2.96 |
| FRSC BRCK OVN ITL PEP PZ | -2.44 | PL SR CRUST SUPRM PIZZA | -2.95 |
| MKSL PRETZEL STICKS | -2.35 | PL SR CRUST PEPPRN PIZZA | -2.73 |
| PL MINI TWIST PRETZELS | 2.33 | QKER CAP N CRUNCH | -2.52 |
| PL SR CRUST SUPRM PIZZA | -2.17 | QKER CAP N CRUNCH BERRIES | -2.4 |

From the above data, we can see that KELL FROOT LOOPS, PL SR CRUST 3 MEAT PIZZA and FRSC PEPPERONI PIZZA have the highest impact on both Product sales and Unit sales. So, if the price of these products is lowered and featured in the in-store circular (as they belong to the Cold Cereal and Frozen Pizza categories where Feature has the highest effect), the chances of maximizing product sales and unit sales are high.