Snack Chain

Predictor Table:

DV: Spend, Unit, hhs (same expected effect for all three dv's)

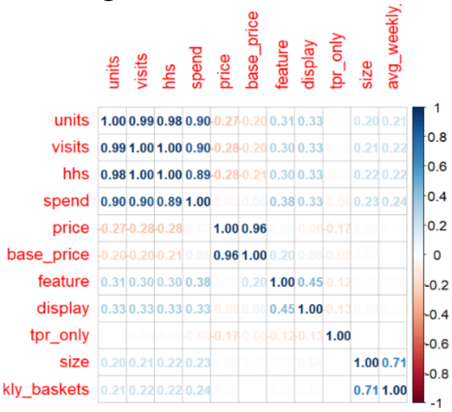
Predictors	Expected Effect	Rationale
Base price	-	Since it's the original price of the product without any type of promotions, this might lower the spend as it gets increased.
Category	?	Consumption might vary based on type of product as some are required more frequently which might impact spend.
Product size	+	It's the equivalent for quantity that may help determine the amount spent.
Segment	?	For analyzing effect of display, feature, promotion by store segment.
Feature	+	To analyze variation of presence in store circular by segment and categories.
Display	+	To analyze effect of in store promotional display by segment and categories.
TPR only	+	To analyze effect of temporary price reduction by segment and categories.
Store id	NA	Used as a common column for joining the data tables.
UPC	NA	Unique product identifier used for merging.
Excluded Factors		
Parking	NA	Can neither be imputed nor be dropped as missing values are extremely large. Also, snack shopping might not be affected by availability to park.
Visits/ Price/ Units/ hhs	NA	High correlations suggest if taken they will drive the model and be capable of fully explaining the variation in spend, without making a room for other predictors. They are considered as DV's. for various models.
Manufacturer	NA	Might be correlated with description/category/subcategory.
Week_end_date	NA	The promotions/ display/ feature change independent of date therefore not considering for analysis.
Subcategory	NA	Since we need to answer questions of variability by categories, therefore not considering subcategory.
City	NA	Considered state for analysis to capture random effects on state level.
Description	NA	Taken store id for analysis.

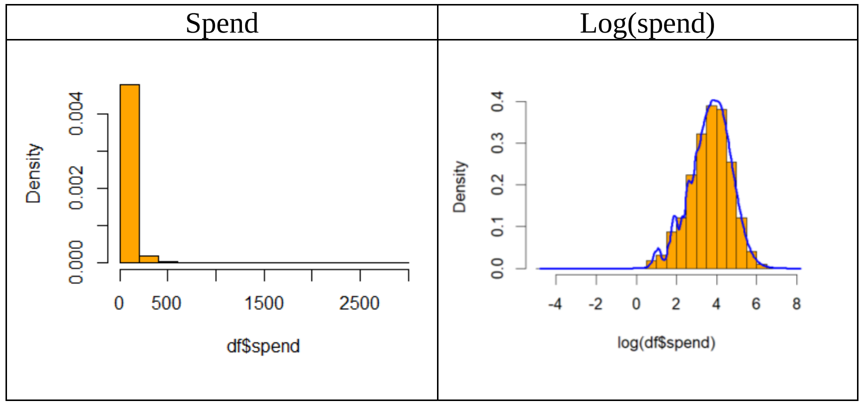
Cleaning and Processing

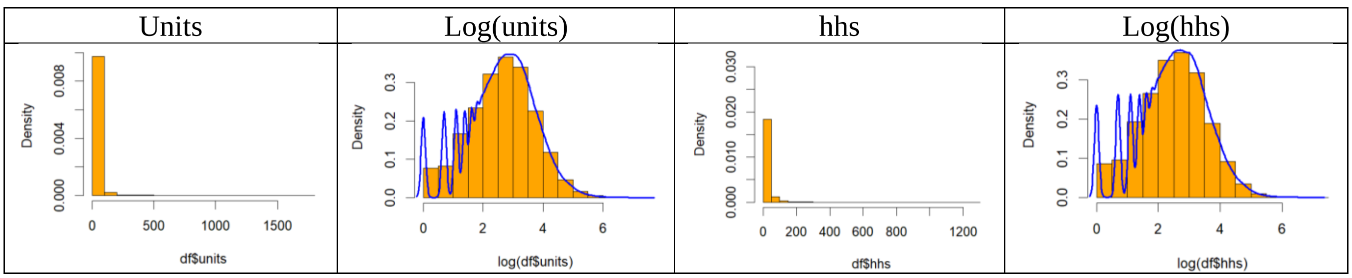
>Considering mainstream segment stores for storeid's of flower mound and rockwall as they have duplicated values and discarding upscale segment of these stores as one store cannot be segmented twice.
>Joined products and transactions by upc, then joined this with stores using store id.
>found missing values for price and base_price columns.
>for price replaced na's with spend/units.
>for using column product size, dropped the unit 'oz' from the values.
>filled base_price with price as no promotional discount was found particularly with rows that had na's.
>converted upc, store id, category, description, manufacturer, subcategory, city, state, msa, segment, store name to factor variables.
>dropped 11 rows as spend was 0 (justified during modeling)
> The data is at two levels: Level 1: Products (upc, description), and Level 2: Stores (store id, segment). Hence, using multi-level models with random effects for this analysis.

Data Exploration

Checking correlations for the numeric variables:







The plots suggest that the distribution is not gaussian, so we can try log models.

Modelling

>spend: m2_s = lmer(log(spend) ~ display*category + feature*category + tpr_only*category +
display*segment + feature*segment + tpr_only*segment +
base_price + product_size + (1 | state) + (1 | store_id), data = df)

>Dropped manufacturer as it showed 322 for multicollinearity test as stated in the predictor table.
>Dropped 11 rows for spend=0 as we are using lmer models which cannot handle 0.
>Considered interactions between different promotions (display/feature/tpr) -categories and promotions-segments for answering questions to analyze variations by segments and categories.

>unit: m1_u = lmer(log(units) ~ display + feature + tpr_only + base_price+ category +
product_size + segment + (1 | store_id) + (1 | state), data = df)

>hhs: m1_h = lmer(log(hhs) ~ display + feature + tpr_only + base_price+ category +
product_size + segment + (1 | store_id) + (1 | state), data = df)

Stargazer

```
DV's
================================================================================
                                        Dependent variable:
                              --------------------------------------------------
                                 log(spend)        log(units)         log(hhs)
                                    (1)               (2)               (3)
--------------------------------------------------------------------------------
display                       0.786*** (0.008)  0.812*** (0.004)  0.808*** (0.004)
categoryCOLD CEREAL           0.988*** (0.004)  0.918*** (0.003)  0.912*** (0.003)
categoryFROZEN PIZZA          -0.097*** (0.008) -0.022*** (0.008) -0.082*** (0.008)
feature                       0.295*** (0.015)  0.780*** (0.005)  0.725*** (0.005)
tpr_only                      -0.085*** (0.007) 0.214*** (0.004)  0.155*** (0.004)
segmentUPSCALE                0.273*** (0.090)  0.241*** (0.089)  0.242*** (0.090)
segmentVALUE                  -0.375*** (0.082) -0.377*** (0.082) -0.408*** (0.082)
base_price                    0.024*** (0.001)  -0.246*** (0.001) -0.217*** (0.001)
product_size                  0.035*** (0.0003) 0.032*** (0.0004) 0.029*** (0.0003)
display:categoryCOLD CEREAL   -0.162*** (0.011)
display:categoryFROZEN PIZZA  -0.153*** (0.010)
categoryCOLD CEREAL:feature   0.320*** (0.016)
categoryFROZEN PIZZA:feature  0.382*** (0.016)
categoryCOLD CEREAL:tpr_only  0.162*** (0.009)
categoryFROZEN PIZZA:tpr_only 0.271*** (0.010)
display:segmentUPSCALE        -0.057*** (0.011)
display:segmentVALUE           0.025** (0.011)
feature:segmentUPSCALE        -0.201*** (0.012)
feature:segmentVALUE          0.082*** (0.012)
tpr_only:segmentUPSCALE       -0.053*** (0.010)
tpr_only:segmentVALUE         -0.140*** (0.009)
Constant                      2.527*** (0.115)  2.396*** (0.113)  2.248*** (0.106)
--------------------------------------------------------------------------------
Observations                      408,410           408,410           408,410
Log Likelihood                  -479,093.500      -489,823.000      -477,373.300
Akaike Inf. Crit.                958,237.100       979,672.000       954,772.500
Bayesian Inf. Crit.              958,510.100       979,814.000       954,914.500
================================================================================
```

Assumptions

Test for autocorrelation is not suggested for multilevel data, therefore tested for multicollinearity.

<table>
<tr><th>Spend</th><th>Units</th><th>Hhs</th></tr>
<tr><td><pre>
> vif(m2_s)
                       GVIF Df GVIF^(1/(2*Df))
display   3.857293 1   1.963999
category   9.217570 2   1.742425
feature   12.317161 1   3.509581
tpr_only   3.309039 1   1.819076
segment   1.000905 2   1.000226
base_price   3.405904 1   1.845509
product_size   3.584164 1   1.893189
display:category   6.132631 2   1.573663
category:feature 20.400190 2   2.125243
category:tpr_only 2.964847 2   1.312202
display:segment   2.728481 2   1.285228
feature:segment   2.858026 2   1.300219
tpr_only:segment   1.831116 2   1.163266
</pre></td><td><pre>
> vif(m1_u)
                 GVIF Df GVIF^(1/(2*Df))
display   1.286192 1   1.134104
feature   1.328112 1   1.152437
tpr_only   1.028504 1   1.014152
base_price   3.347989 1   1.829751
category   6.891537 2   1.620239
product_size 3.516469 1   1.875225
segment   1.000011 2   1.000003
</pre></td><td><pre>
> vif(m1_h)
                 GVIF Df GVIF^(1/(2*Df))
display   1.286192 1   1.134104
feature   1.328112 1   1.152438
tpr_only   1.028504 1   1.014152
base_price   3.347988 1   1.829751
category   6.891536 2   1.620239
product_size 3.516470 1   1.875225
segment   1.000010 2   1.000002
</pre></td></tr>
<tr><td>Pass</td><td>Pass</td><td>Pass</td></tr>
</table>

A1: Effects:

	Display	Feature	Temp price reduction
Spend	78% more than no display	29% more than not being in-store circular	8.5% less than not reducing price
Unit	~81% more than no display	78% more than not being in-store circular	~21% more than not reducing price
Hhs	80.8% more than no display	72.5% more than not being in-store circular	15.5% more than not reducing price

(With reference to stargazer)

Effect variation by category and segment:
spend:
log(spend) = 2.52 + 0.78(display) + 0.98(categorycold cereal) - 0.09(categoryfrozen pizza) + 0.29(feature) - 0.08(tpr_only) + 0.27(segmentupscale) - 0.37(segmentvalue) + 0.02(base_price) + 0.03(product_size) - 0.16(display:categorycold cereal) - 0.15(display:categoryfrozen pizza) + 0.32(categorycold cereal:feature) + 0.38(categoryfrozen pizza:feature) + 0.16(categorycold cereal:tpr_only) + 0.27(categoryfrozen pizza:tpr_only) - 0.05(display:segmentupscale) + 0.02(display:segmentvalue) - 0.2(feature:segmentupscale) + 0.08(feature:segmentvalue) - 0.05(tpr_only:segmentupscale) - 0.14(tpr_only:segmentvalue)

units:
log(units) = 2.39 + 0.81(display) + 0.91(categorycold cereal) - 0.02(categoryfrozen pizza) + 0.78(feature) + 0.21(tpr_only) + 0.24(segmentupscale) - 0.37(segmentvalue) - 0.24(base_price) + 0.03(product_size)

hhs:
log(hhs) = 2.24 + 0.80(display) + 0.91(categorycold cereal) - 0.08(categoryfrozen pizza) + 0.72(feature) + 0.15(tpr_only) + 0.24(segmentupscale) - 0.4(segmentvalue) - 0.21(base_price) + 0.02(product_size)

Category by display: Spend variation of display by category is ~62% more for cold cereals than bag snacks. Spend variation of display by category is ~63% more for frozen pizza than bag snacks.

Category by feature: Spend variation of feature by category is ~61% more for cold cereals than bag snacks. Spend variation of feature by category is ~67% more for frozen pizza than bag snacks.

Category by tpr: Spend variation of tpr by category is ~24% more for cold cereals than bag snacks. Spend variation of tpr by category is ~35% more for frozen pizza than bag snacks.

Segment by display: Spend variation of display by segment is 83% more for upscale than mainstream. Spend variation of display by segment is 80% more for value than mainstream.

Segment by feature: Spend variation of feature by segment is 9% more for upscale than mainstream. Spend variation of feature by segment is 37% more for value than mainstream.

Segment by tpr: Spend variation of tpr by segment is 3% more for upscale than mainstream. Spend variation of tpr by segment is 6% less for value than mainstream.

Price Elasticity refers to how much the quantity demanded of a product will increase or decrease as the price of the product changes. Here we have considered absolute values as the sign defines the direction of relationship, but elasticity is signified by magnitude.

Top 5 elastic products:

UPC	Elasticity	Description	Price (mean)
7110410471	0.404154557	MKSL PRETZEL STICKS	2.34
7218063979	0.320722047	FRSC PEPPERONI PIZZA	6.87
7218063983	0.304485143	FRSC 4 CHEESE PIZZA	6.88
7110410455	0.262493678	MKSL MINI TWIST PRETZELS	2.34
7218063052	0.232834260	FRSC BRCK OVNITL PEP PZ	6.88

Bottom 5 elastic products:

UPC	Elasticity	Description
3800039118	0.002270669	KELL FROOT LOOPS
1111087395	0.003471602	PL SR CRUST SUPRM PIZZA
7797508006	0.006181680	SNYDR FF MINI PRETZELS
88491212971	0.007380526	POST FRUITY PEBBLES
1600027527	0.018973280	GM HONEY NUT CHEERIOS

Reducing the base price of a product generally increases the no. of units purchases but also depends on the limit of reduction. The trade-off between lowering the price and the profit due to its sale should be optimum. Therefore, for increasing product sales the stores should consider lowering the cost of products highlighted in red in the table above. Products with high elasticity and high price should be considered for giving discount as they tend to determine most change in product sale.

For increasing unit sale, the product with highest elasticity should be considered which are MKSL PRETZEL STICKS, FRSC PEPPERONI PIZZA, FRSC 4 CHEESE PIZZA (top 3).

As we see that 2 products are common in both the unit and product sale, we can recommend reducing the price of those products which are:

FRSC PEPPERONI PIZZA
FRSC 4 CHEESE PIZZA