Product Sales And Promotions Analysis:

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

**Spend:** It is the sales in dollars for all the transactions for various product categories.
**Unit Sales:** It is the number of units sold in all the product categories.
Household Sales: It is the number of purchasing households in all the product categories.

Steps for data cleaning, pre-processing and partitioning are as follows:
1. First, we need to subset the data for the analysis into three groups as rc_stores, rc_products and rc_transactions.
2. Then, we need to find if there are NA values and zeros in the data and remove those rows.
3. Next, we will drop oral hygiene products and all the associated transactions.
4. Lastly, we will need to extract the year, month and weeknum from the week_end_date column and factorize some of the predictors.

Plots: From the plots below, spend and log transformed units and hhs seems to follow a normal distribution.
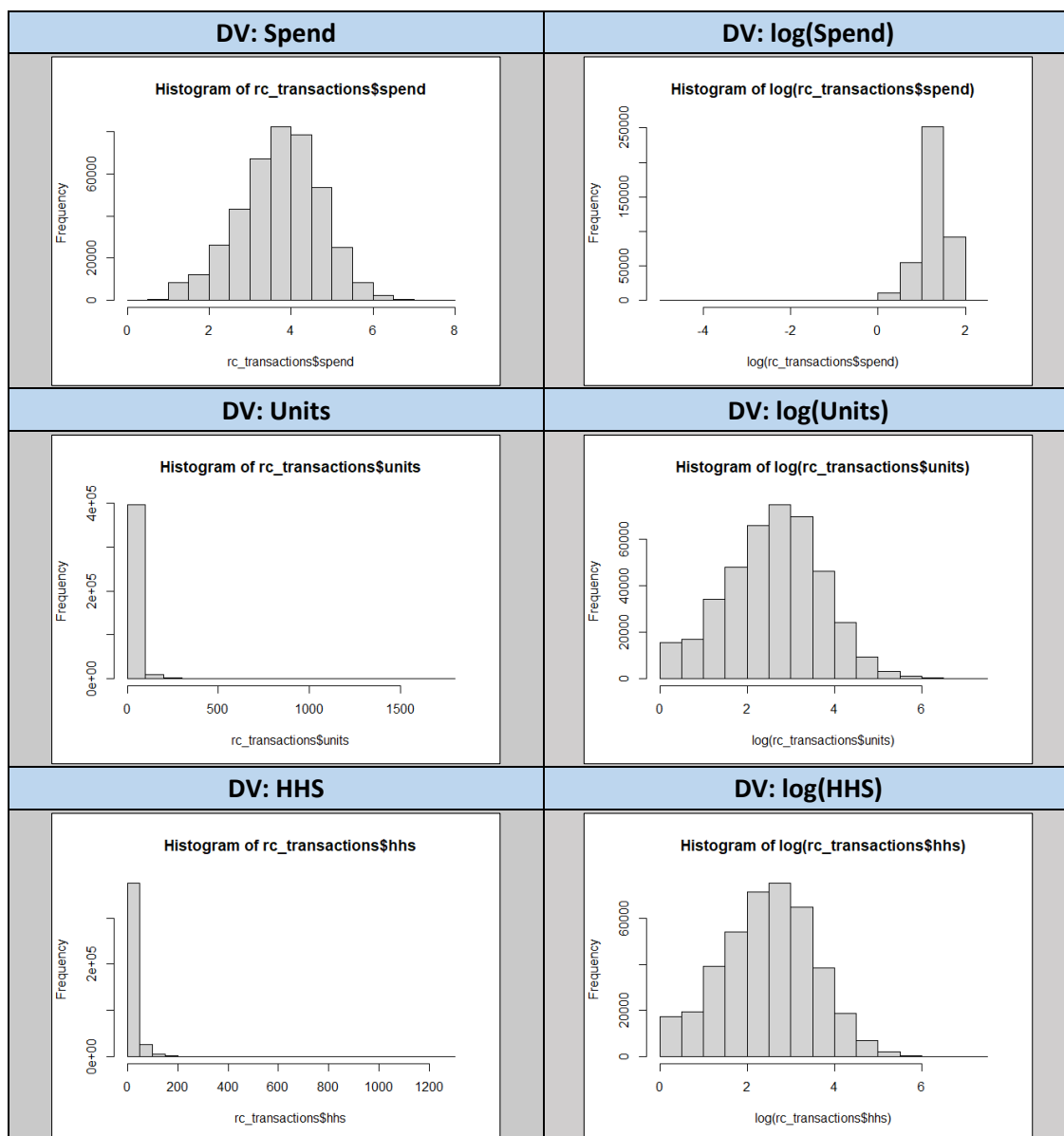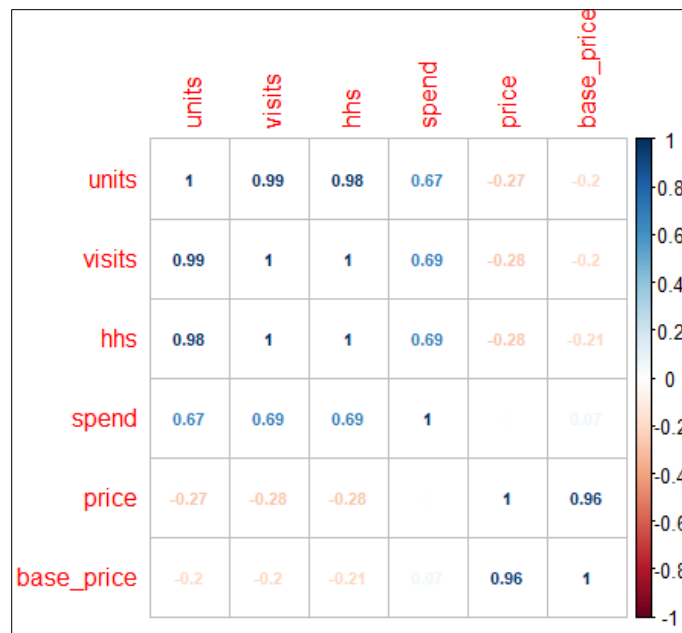
From the correlation plot above, we can see that price and base_price are highly correlated. Same goes for all our DV's as well. We can drop base_price from our analysis.

Factor Effects:
Positive Effect: If there is a direct proportionality (+X then +y / -X then -y) between the predictor variable (X) and the response variable (y), we can say there is a positive effect.

Negative Effect: If there is an inverse proportionality (+X then -y / -X then +y) between the predictor variable (X) and the response variable (y), we can say there is a negative effect.

Relevant Factors for [y <- Spend, Units, HHS]			
Predictor	**Spend**	**Units**	**HHS**	**Rationale**
State	+	+	+	Helps us understand how all factors vary based on location which could help us better understand which locations need better promotions.
Segment	+	+	+	Store segment helps us understand the impacts of store type on spend, sales etc.
Avg_Weekly_Basket	+	+	+	It is number of items sold in a single purchase which could help us better understand the spend of the customers.
Category	+	+	+	Product category is vital to look at as it could help us understand which type of products sell more and which less. Helps with better procurement and reduced wastage.
Visits	+	+	+	Partial dependency on basket size but helps us understand each products unique purchase history.
Feature	+	+	+	Key predictor which shows if the product was an in-store circular feature.
Display	+	+	+	Key predictor which shows if the product was an in-store promotion.
Tpr_Only	+	+	+	Key predictor which shows if there was a price drop in the product before the transaction.
Year	+	+	+	Feature engineered variable to see how key predictors change with year.

Month	+	+	+	Feature engineered variable to see how key predictors change with month.
Weeknum	+	+	+	Engineered variable to see the impact of predictors on a weekly continuous basis.

```r
#Multi-Level Analysis
spend_model <- lmer(spend ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + (1 | store_num), data = rc_transactions)
units_model <- lmer(log(units) ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + (1 | store_num), data = rc_transactions)
hhs_model <- lmer(log(hhs) ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + (1 | store_num), data = rc_transactions)
```

```r
#Stargazer
stargazer(spend_model, units_model, hhs_model, type="text", single.row=TRUE)
```

```
## 
## ============================================================================
##                                 Dependent variable:
##                     ------------------------------------------------------
##                          spend           log(units)          log(hhs)
##                           (1)                (2)                (3)
## ----------------------------------------------------------------------------
## log(visits)         0.956*** (0.0003)  1.000*** (0.0003)  0.985*** (0.0001)
## feature1            0.094*** (0.001)   0.052*** (0.001)   0.014*** (0.001)
## display1            -0.023*** (0.001)  0.008*** (0.001)   0.013*** (0.0005)
## tpr_only1           0.019*** (0.001)   0.063*** (0.001)   0.011*** (0.0004)
## price               0.257*** (0.0002)  -0.010*** (0.0002) -0.005*** (0.0001)
## year2010            -0.005*** (0.001)  -0.006*** (0.001)  0.001 (0.001)
## year2011            -0.017*** (0.002)  -0.015*** (0.002)  0.004*** (0.001)
## year2012            -0.0001 (0.004)    -0.006* (0.003)    0.001 (0.002)
## weeknum             0.0002*** (0.00002) 0.0001*** (0.00002) -0.00002* (0.00001)
## Constant            0.484*** (0.002)   0.130*** (0.002)   0.024*** (0.003)
## 
## ----------------------------------------------------------------------------
## Observations              408,238            408,238            408,238
## Log Likelihood          138,848.600        213,635.200        446,500.400
## Akaike Inf. Crit.       -277,673.200       -427,246.400       -892,976.900
## Bayesian Inf. Crit.     -277,542.100       -427,115.300       -892,845.800
## ============================================================================
## Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

Quantitative Insights:

	Spend	Units Sold	Household Sales
Products that are part of in-store promotional displays contribute to	**2.3% decrease** in sales	**0.8% increase** in units sold	**1.3% increase** in household sales
Products that are part of in-store circulars contribute to	**9.4% increase** in sales	**5.2% increase** in units sold	**1.4% increase** in household sales

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

```r
#Multi-Level Analysis Based on Category & Segment
spend_model2 <- lmer(spend ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + segment + category + (1 | store_num), data = merged_rc)
units_model2 <- lmer(log(units) ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + segment + category + (1 | store_num), data = merged_rc)
hhs_model2 <- lmer(log(hhs) ~ log(visits) + feature + display + tpr_only + price + year
 + weeknum + segment + category + (1 | store_num), data = merged_rc)
```

```
#Stargazer
stargazer(spend_model2, units_model2, hhs_model2, type="text", single.row=TRUE)
```

```
## 
## ================================================================================
##                                   Dependent variable:
##                        ---------------------------------------------------------
##                            spend            log(units)           log(hhs)
##                             (1)                (2)                  (3)
## --------------------------------------------------------------------------------
## log(visits)          0.937*** (0.001)   1.012*** (0.0005)   0.983*** (0.0004)
## feature1             0.059*** (0.002)   0.016*** (0.002)    0.013*** (0.001)
## display1             0.006*** (0.002)  -0.013*** (0.001)    0.008*** (0.001)
## tpr_only1            0.012*** (0.001)   0.042*** (0.001)    0.008*** (0.001)
## price                0.236*** (0.001)  -0.037*** (0.0005)  -0.020*** (0.0004)
## year2010            -0.011*** (0.002)  -0.009*** (0.002)    0.0003 (0.001)
## year2011            -0.021*** (0.003)  -0.017*** (0.003)    0.007*** (0.002)
## year2012            -0.002 (0.007)     -0.012* (0.006)      0.001 (0.005)
## weeknum              0.0002*** (0.00003) 0.0001*** (0.00003) -0.00004* (0.00002)
## segmentUPSCALE       0.001 (0.003)      -0.001 (0.003)      0.0002 (0.002)
## segmentVALUE        -0.006 (0.009)       0.004 (0.003)     -0.049*** (0.011)
## categoryCOLD CEREAL  0.085*** (0.001)   -0.025*** (0.001)  -0.015*** (0.001)
## categoryFROZEN PIZZA 0.120*** (0.002)   0.125*** (0.002)    0.055*** (0.001)
## Constant             0.533*** (0.007)   0.172*** (0.003)    0.081*** (0.007)
## --------------------------------------------------------------------------------
## Observations             136,093            136,093            136,093
## Log Likelihood          45,471.510         71,580.540        105,196.600
## Akaike Inf. Crit.      -90,911.030       -143,129.100       -210,361.300
## Bayesian Inf. Crit.    -90,753.890       -142,971.900       -210,204.100
## ================================================================================
## Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

Assumption Testing

1. Multicollinearity - Passed

Spend	Units Sold	Household Sales

```   GVIF Df GVIF^(1/(2*Df))```
 ```## log(visits) 1.682765 1   1.297214```
 ```## feature   1.380449 1   1.174925```
 ```## display   1.368067 1   1.169644```
 ```## tpr_only   1.084037 1   1.041171```
 ```## price   3.288506 1   1.813424```
 ```## year   8.875698 3   1.438910```
 ```## weeknum   8.840835 1   2.973354```
 ```## segment   1.001002 2   1.000250```
 ```## category   3.868661 2   1.402459```
	``` GVIF Df GVIF^(1/(2*Df))```
	 ```## log(visits) 1.678309 1 1.295496```
	 ```## feature 1.379963 1 1.174718```
	 ```## display 1.368181 1 1.169693```
	 ```## tpr_only 1.083945 1 1.041127```
	 ```## price 3.282864 1 1.811867```
	 ```## year 8.876861 3 1.438942```
	 ```## weeknum 8.841984 1 2.973547```
	 ```## segment 1.005622 2 1.001403```
	 ```## category 3.857445 2 1.401441```
		``` GVIF Df GVIF^(1/(2*Df))```
		 ```## log(visits) 1.683657 1 1.297558```
		 ```## feature 1.380537 1 1.174962```
		 ```## display 1.368049 1 1.169636```
		 ```## tpr_only 1.084053 1 1.041179```
		 ```## price 3.289515 1 1.813702```
		 ```## year 8.875503 3 1.438905```
		 ```## weeknum 8.840641 1 2.973322```
		 ```## segment 1.000283 2 1.000071```
		 ```## category 3.870682 2 1.402642```

2. Independence - Passed

Spend	Units Sold	Household Sales
DW = 2.0009, p-value = 0.8653	DW = 2.0008, p-value = 0.8892	DW = 2.0001, p-value = 0.9888

Quantitative Insights:

	Spend	Units Sold	Household Sales
Product Categories			
Cold cereal contributes to	**8.5% increase in sales than bag snacks** for products that are part of in-store promotional displays	**2.5% decrease in units sold than bag snacks** for products that are part of in-store promotional displays	**1.5% decrease in household sales than bag snacks** for products that are part of in-store promotional displays
Frozen pizza contributes to	**12% increase in sales than bag snacks** for products that are part of in-store promotional displays	**12.5% increase in units sold than bag snacks** for products that are part of in-store promotional displays	**5.5% increase in household sales than bag snacks** for products that are part of in-store promotional displays

Store Segments			
Upscale segment contributes to	**0.1% increase in sales than mainstream segment** for products that are part of in-store promotional displays	**0.1% decrease in units sold than mainstream segment** for products that are part of in-store promotional displays	**0.02% increase in household sales than mainstream segment** for products that are part of in-store promotional displays
Value segment contributes to	**0.6% decrease in sales than mainstream segment** for products that are part of in-store promotional displays	**0.4% increase in units sold than mainstream segment** for products that are part of in-store promotional displays	**4.9% decrease in household sales than mainstream segment** for products that are part of in-store promotional displays

III. 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 (if you have never heard this term before, you can Google it).

The Five Most Price Elastic Products:

```
##         UPC standing                description manufacturer      category
## 1 7192100339 2.797298      DIGRN PEPP PIZZA    TOMBSTONE FROZEN PIZZA
## 2 7192100337 2.493018   DIGRN SUPREME PIZZA    TOMBSTONE FROZEN PIZZA
## 3 7192100336 2.184538    DIGIORNO THREE MEAT   TOMBSTONE FROZEN PIZZA
## 4 7218063052 1.945852 FRSC BRCK OVN ITL PEP PZ     TONYS FROZEN PIZZA
## 5 7218063979 1.864758    FRSC PEPPERONI PIZZA      TONYS FROZEN PIZZA
```

From our analysis, products that are in frozen pizza category contributes to a significant chunk of the sales. The same is reinforced from our analysis in the table above. The five most price elastic products are:

1. DIGRN PEPP PIZZA.
2. DIGRN SUPREME PIZZA.
3. DIGIORNO THREE MEAT.
4. FRSC BRCK OVN ITL PEP PZ.
5. FRSC PEPPERONI PIZZA.

The Five Least Price Elastic Products:

```
##          UPC    standing               description  manufacturer    category
## 37 7797502248 -0.8069794      SNYDR PRETZEL RODS       SNYDER S BAG SNACKS
## 38 1111009507 -0.9067250       PL TWIST PRETZELS PRIVATE LABEL BAG SNACKS
## 39 7110410455 -1.2643368 MKSL MINI TWIST PRETZELS         MKSL BAG SNACKS
## 40 7110410470 -1.3273755     MKSL DUTCH PRETZELS         MKSL BAG SNACKS
## 41 7110410471 -1.5562879     MKSL PRETZEL STICKS         MKSL BAG SNACKS
```

From our analysis, products that are in bag snacks category does not seem to contribute much to the sales. The same is reinforced from our analysis in the table above. The five least price elastic products are:

1. SNYDR PRETZEL RODS.
2. PL TWIST PRETZELS.
3. MKSL MINI TWIST PRETZELS.
4. MKSL DUTCH PRETZELS.
5. MKSL PRETZEL STICKS.

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

From the previous question, we estimated that frozen pizza category contributes to a significant chunk of the sales. Even if we increase the unit price, the demand will be there, and the unit sales can be maximized. That is because the highest selling items will always be in demand so raising prices to get higher margins per unit is something that will not change much in terms of a products appeal.

But products that are in bag snacks category does not seem to contribute much to the sales. Here, if we decrease the unit price the product sales can be maximized. That is because maximizing unit sales highly depends on the product category. In most cases, a single product from a category underperforming is a rare thing as most of the food items are categorized based on the key ingredient that is present in them, in this case pretzels. So, maximizing product sales by lowering the price will eventually lead to an uptick in unit sales.