Project 1: Analysis of Kroger Soft Drink Sales

Josh Allen

Sierra Funte

Kevin Gardner

Digvijay Sawant

The purpose of this analysis of Kroger's soft drink sales was to assess the impact of pricing promotion on sales. The data includes more than a million shopping trips taken over a 102 week timeframe. We excluded 13 weeks from the data set that did not include all of the households.

First, we analyzed the impact of pricing promotion on sales volume for 12/18 packs of 12 ounce cans, which is a subset of the larger soft drink sales table. This was executed by examining the number of days in a week that each of the four manufacturers (69,103, 1208, 2224) was running a discount on the product (see Fig. 2a). From this graph, we learned that for each of the selected manufacturers, a higher number of days with a retail discount was associated with higher sales volume and weekly market share. The sales impact of weekly number of days with a discount is most pronounced for the leading national brands, 103 and 1208. This is true for the 2 Liter as well as the 12/18 pack of 12 ounce cans category.

After examining the sales of soft drink products, we were able to make several recommendations for increasing Kroger's overall soft drink sales. The first recommendation is to eliminate some of the manufacturers that are not generating high sales. We studied over 30 different manufacturers and noticed that the top soft drink sales were coming from three specific manufacturers. This means that the other manufacturers, most of which had a market share value less that 1%, could have product on the shelf that is simply taking up space. In order to increase private brand sales, we recommend that Kroger consider reallocating some of its shelf space from the lowest selling national brands to expand its offering of private brands.

These 30+ manufacturers are selling their product at an average discount of $0.19 per product whereas Kroger is selling their product at an average discounted price of $0.06. Hence, these manufacturers are getting bought out more than private (See Table 5.a)

The plots of sales volume over week number suggests that sales of 2 Liter soft drinks are lower than 12/18 packs of 12 ounce cans. Manufacturers could potentially increase sales of 2 Liter products by offering greater discounts for this size category. Before recommending this we would would need to study whether or not increasing the quantity of 2 Liter sales by offering larger discounts would equate to higher profit.

Question 1

Table.1a. shows the proportion of sales by brand (National and Private). National brand has approximately 73% of total sales over the 102 weeks while private brand has 27%.

Table 1a.

	BRAND	N Rows	Sum(SALES_VALUE)	Proportion
1	National	1054974	2987981.47	72.98771678
2	Private	591102	1105832.67	27.01228322

Fig. 1a shows the proportion of total sales for national and private brands. The first 13 weeks of the data have been excluded because they contain far fewer households than the rest of the weeks in the sample. From the graph we see that national brands have the higher proportion of sales for each of the 89 weeks in our sample. The relation between national to private brand sales volume is generally about 70% national brand to 30% private brand across the 89 weeks. Furthermore, we see the proportion of sales fluctuating slightly between national brand and private brand, but do not notice a significant trend from plotting proportion of sales over time.
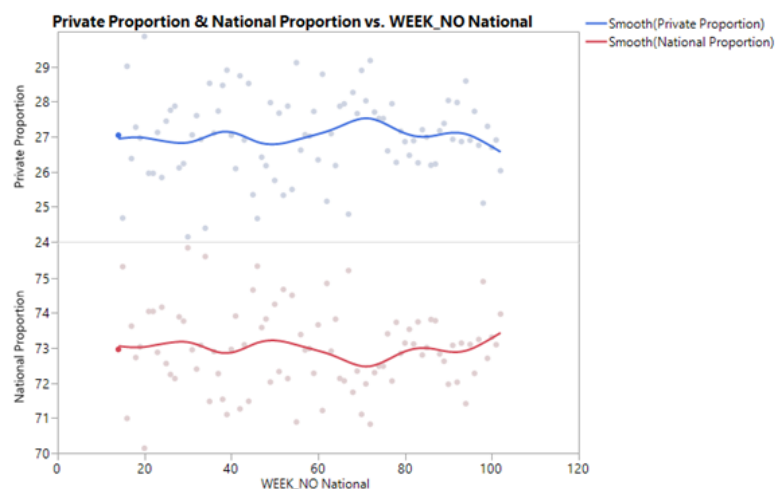
Figure 1a.

Table 1b shows the results from the Durbin-Watson test for autocorrelation. The autocorrelation coefficient is .06882 with a p-value of <.0001. This is statistically significant so we reject the null hypothesis and conclude that positive autocorrelation is present.

Table 1b.

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
0.5505623	102	0.6882	<.0001*

Table 1c shows the results from the Durbin-Watson test for autocorrelation with outlier weeks 1-13 excluded. From this test, we see that autocorrelation is present in the data.

Table 1c.

The autocorrelation reported in Table 1b and Table 1c most likely occurs because the data is time-ordered. A possible reason that we observe positive autocorrelation is that household preferences for national brands are similar to the preceding week.

Question 2

Table 2a lists the market share and average sales volume in dollars for each of the four manufacturers. These measures were obtained by selecting the subset of soft drinks with sub commodity = "12/18 packs of 12 oz. cans" and size="12 OZ", then grouping by manufacturer.

Market share is calculated as sum(SALES_VALUE) divided by the combined sum(SALES_VALUE) of all 13 manufacturers. Average volume is calculated as mean(SALES VALUE).

Table 2a.

MANUFACTURER	MARKET_SHARE	AVG_VOLUME
1208	43.03%	$4.75
103	31.38%	$4.55
2224	13.26%	$3.67
69	10.76%	$2.25
436	0.67%	$3.82
1710	0.35%	$3.39
2152	0.22%	$3.55
1232	0.16%	$3.36
212	0.10%	$3.22
2434	0.06%	$4.31
1046	0.01%	$3.75
2733	0.00%	$3.59
2	0.00%	$0.00

Figure 2a shows the impact of how often items were discounted on weekly sales for each manufacturer.
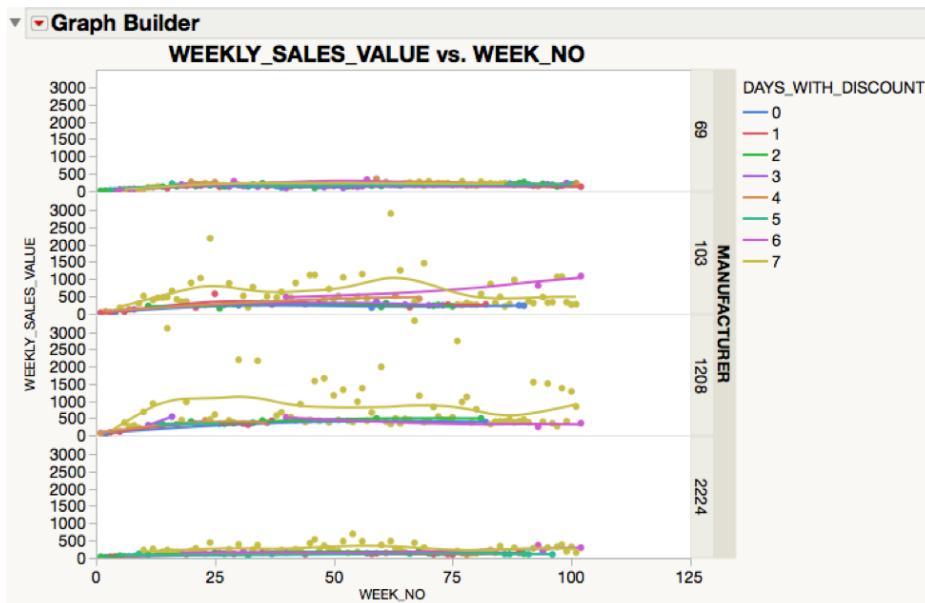
Figure 2a.

Figure 2b shows there may be a seasonal effect. This will be confirmed in part 2c.
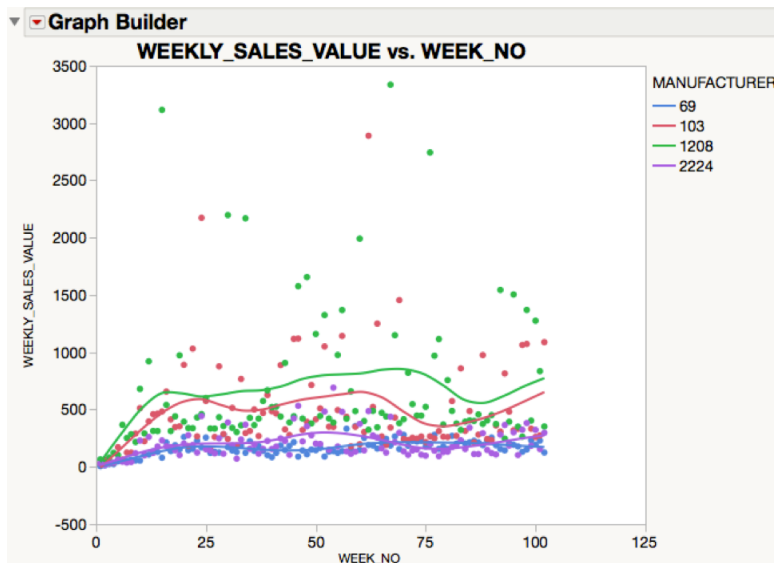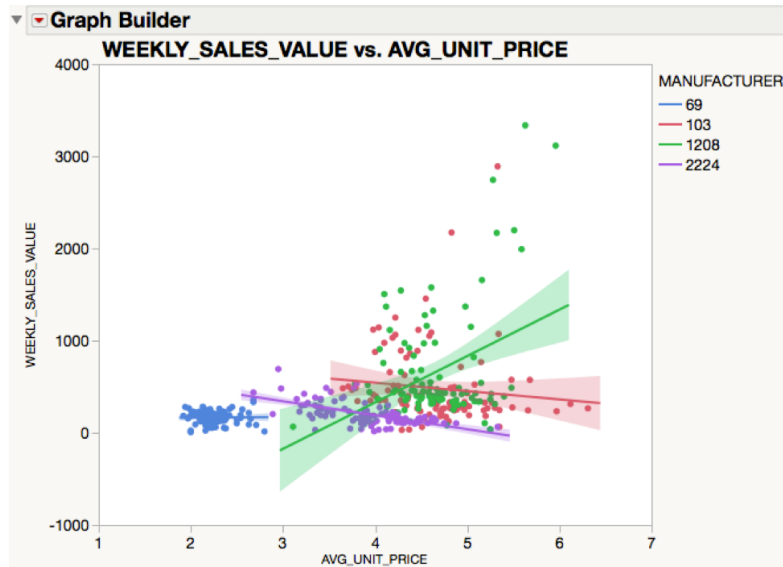
Figure 2b.

Figure 2c shows the impact of average unit price on weekly sales for each manufacturer. It is worth noting that for 1208, higher average unit price is associated with higher weekly sales, whereas for the other manufacturers the association is either negative or relatively flat.

Figure 2c.

WEEKLY_SALES_VALUE vs. AVG_UNIT_PRICE

From the Durbin Watson test we do find significant autocorrelation among weekly sales observations for manufacturers 69 and 2224. For manufacturer 69 the autocorrelation is .4169 with a p-value of <.0001. For manufacturer 2224 the autocorrelation is .1561 and has p-value of .0208. In both cases we reject the null hypothesis and conclude there is positive autocorrelation.

Response WEEKLY_SALES_VALUE

MANUFACTURER=69
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.1071276	102	0.4169	<.0001*

MANUFACTURER=103
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.8837161	102	0.0413	0.2746

MANUFACTURER=1208
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
2.1224392	102	-0.0634	0.7265

MANUFACTURER=2224
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.6006756	102	0.1561	0.0208*

Figure 2d shows the impact of average unit price on weekly market share for each manufacturer. Similar to the previous chart, we see that for 1208 higher average unit price is associated with higher weekly market share while for the other manufacturers the association is either negative or relatively flat. This seems to maybe indicate that 1208 needs to raise its average prices.

Figure 2d.
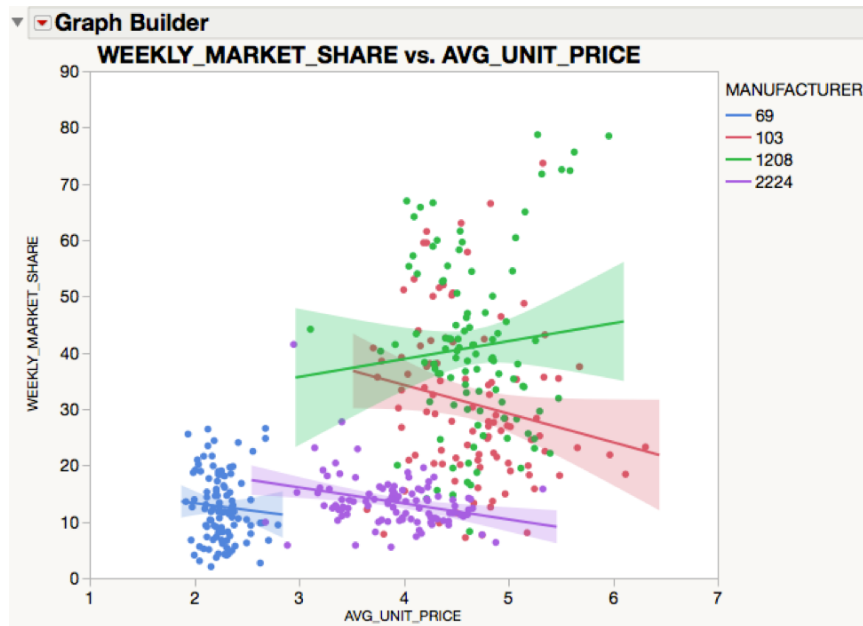
Graph Builder
WEEKLY_MARKET_SHARE vs. AVG_UNIT_PRICE

From the Durbin Watson test we do we do not find significant autocorrelation among weekly market share observations for any of the four manufacturers. In each case the autocorrelation coefficient is negative but significantly different.

Response WEEKLY_MARKET_SHARE

MANUFACTURER=69
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
2.2128579	102	-0.1225	0.8536

MANUFACTURER=103
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
2.2128579	102	-0.1225	0.8536

MANUFACTURER=1208
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
2.1511138	102	-0.0975	0.7756

MANUFACTURER=2224
Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
2.2293629	102	-0.1262	0.8744

Question 3

Table 3a lists the market share and average sales volume in dollars for each of the four manufacturers. These measures were obtained by selecting the subset of soft drinks with sub commodity = "2 LITER BTL CARB INCL" and size="2 LTR", then grouping by manufacturer.

Market share is calculated as sum(SALES_VALUE) divided by the combined sum(SALES_VALUE) of all 10 manufacturers. Average volume is calculated as mean(SALES VALUE).

Table 3a.

MANUFACTURER	MARKET_SHARE	AVG_VOLUME
103	35.53%	$1.87
1208	31.59%	$1.88
69	17.01%	$0.77
2224	14.35%	$1.43
1647	0.68%	$1.67
212	0.36%	$1.25
2152	0.18%	$1.12
1710	0.17%	$1.28
436	0.11%	$1.31
1918	0.02%	$0.90

Figure 3a shows the impact of how often items were discounted on weekly sales for each manufacturer (we may or may not want to include this).
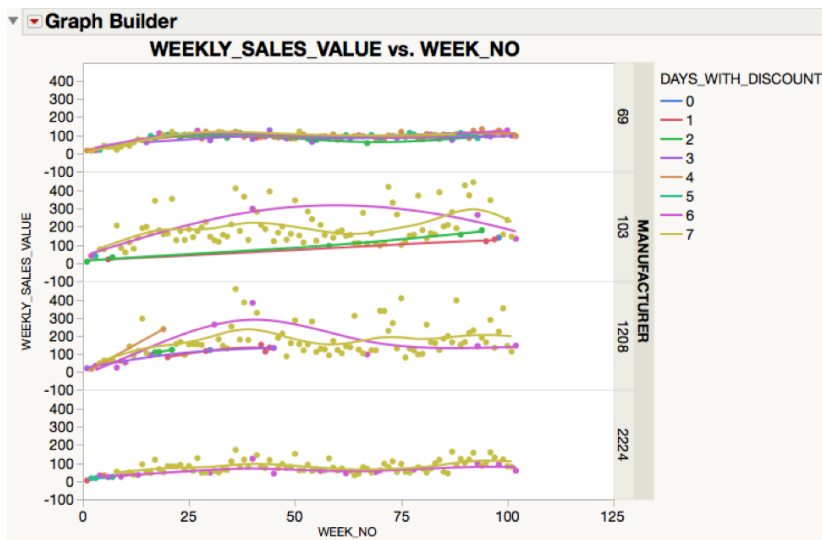
Figure 3a.

Figure 3b shows there may be a seasonal effect. We will need to use the Durbin Watson test to check for autocorrelation.
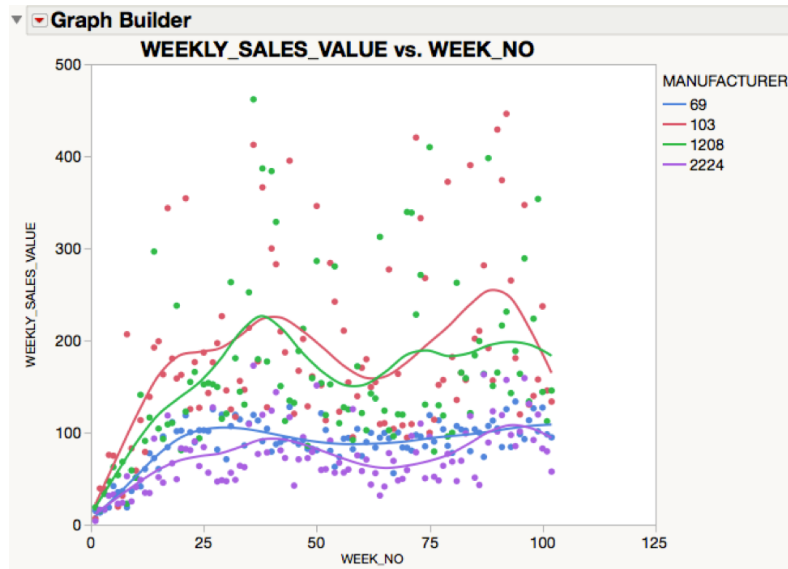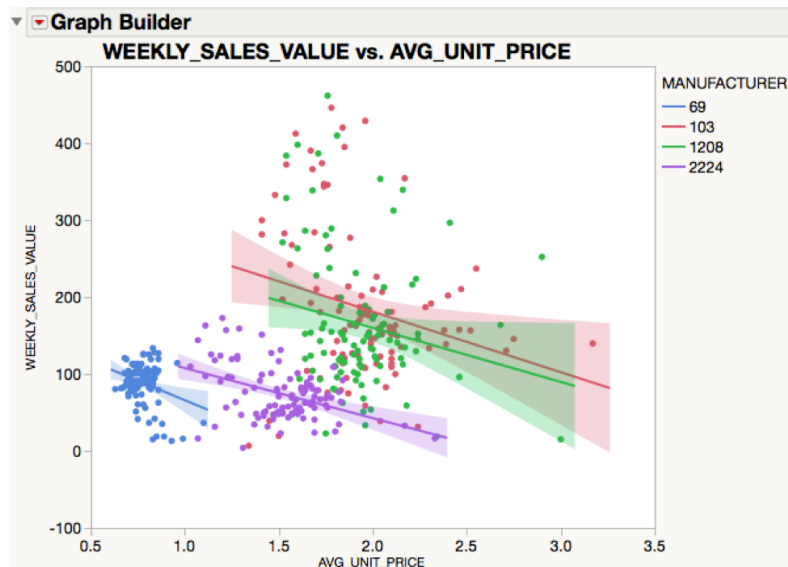
Figure 3b.

Figure 3c shows the impact of average unit price on weekly sales for each manufacturer. For all four of the selected manufacturers, there is a negative association between average unit price and weekly sales volume.

Figure 3c.

From the Durbin Watson test we find significant autocorrelation among weekly sales observations for all four of the selected manufacturers. For each manufacturer the autocorrelation is positive with a with a p-value near .0001. Thus we reject the null hypothesis and conclude there is positive autocorrelation.

Response WEEKLY_SALES_VALUE

MANUFACTURER=69

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.0164321	102	0.4543	<.0001*

MANUFACTURER=103

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.3174636	102	0.3123	0.0002*

MANUFACTURER=1208

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.461719	102	0.2606	0.0027*

MANUFACTURER=2224

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.2000708	102	0.3644	<.0001*

Figure 3d shows the impact of average unit price on weekly market share for each manufacturer. For all four of the selected manufacturers, there is a negative association between average unit price and weekly market share percentage.

Figure 3d.
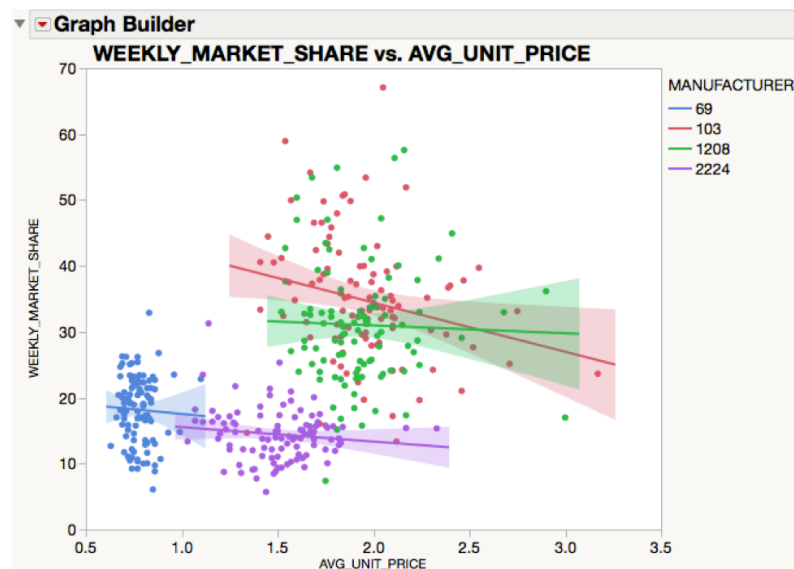
From the Durbin Watson test we do not find significant autocorrelation among weekly market share observations for any of the four manufacturers. In each case the autocorrelation coefficient is negative but significantly different different 0.

Response WEEKLY_MARKET_SHARE

MANUFACTURER=69

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.819766	102	0.0451	0.1728

MANUFACTURER=103

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.8248389	102	0.0533	0.1782

MANUFACTURER=1208

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.8894552	102	0.0485	0.2858

MANUFACTURER=2224

▼ ▾ **Durbin-Watson**

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.9892854	102	-0.0082	0.4870

Question 4:

Show the distribution of frequency of trips including soft drink sales and average amount spent on soft drinks for households.

Fig.4a shows the percentage of baskets which contain soda by week. As we can see there is a generally season trend towards higher summer sales but declining sales as well. The percentage is around 5% of baskets containing soda.

Figure 4a.

PercentSoda vs. WEEK_NO of SODABASKETS
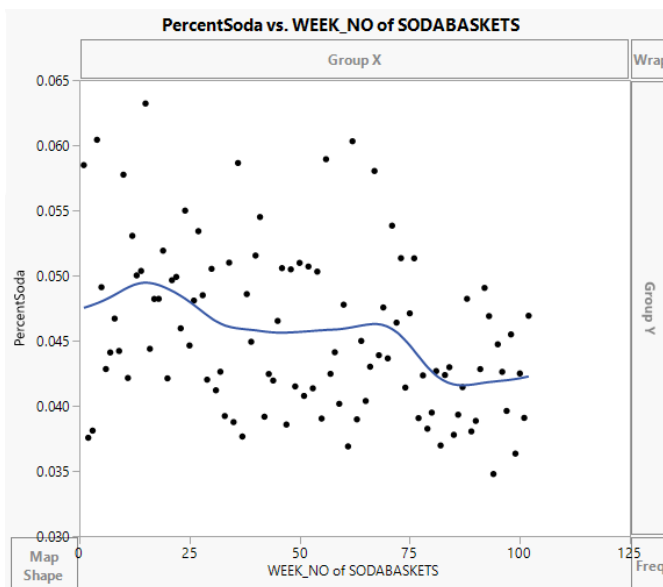
Below is the distribution of average sales of soda per trip. From what we see here mean dollar sales per trip is around $2.8. Although there are quite a few outliers indicating high amounts spent by certain households per trip on soda.

Distribution of Trips and Sales

Figure 4b.

sales_per_trip

Quantiles		
100.0%	maximum	30
99.5%		11.2416754
97.5%		7.014125
90.0%		4.6567104
75.0%	quartile	3.4504555
50.0%	median	2.569024
25.0%	quartile	1.8147025
10.0%		1.3346
2.5%		0.95408175
0.5%		0.67200234
0.0%	minimum	0.435

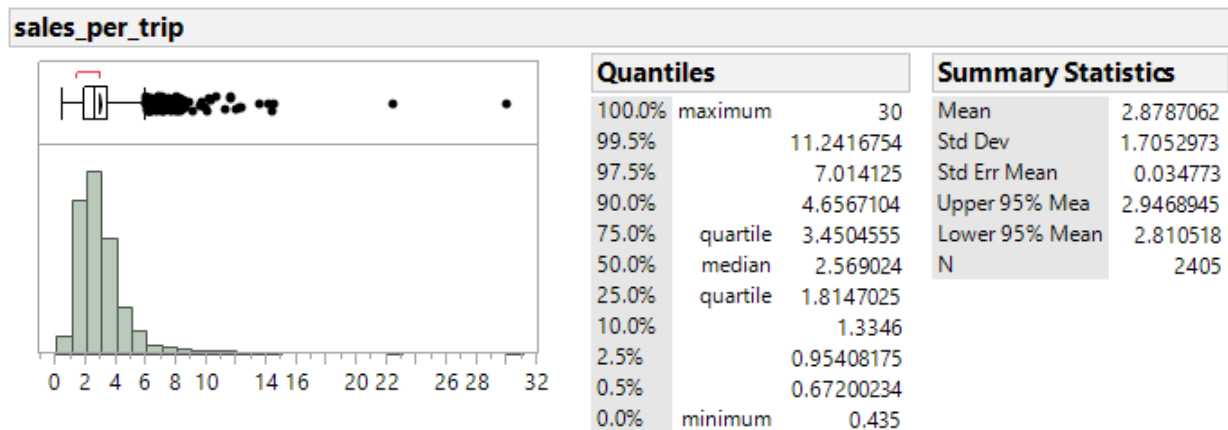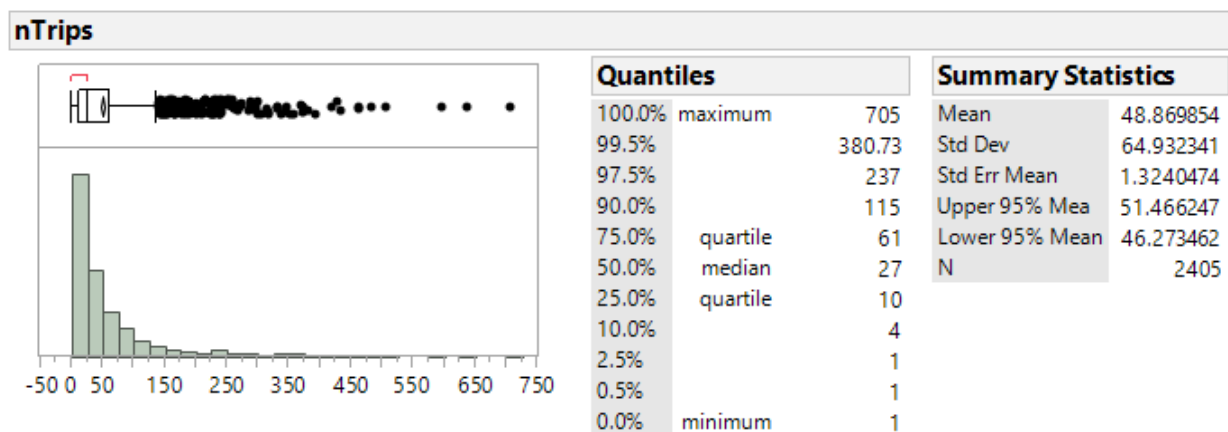| Summary Statistics	
Mean	2.8787062
Std Dev	1.7052973
Std Err Mean	0.034773
Upper 95% Mea	2.9468945
Lower 95% Mean	2.810518
N	2405

Fig.4b shows the distribution of number of trips made by each household that bought soda or had soda in their basket. Since this is a period of 2 years, we can conclude in a year around 24 trips were made by each household each year. After looking at both of these graphs we can conclude that with a mean spent of $2.8 per trip and total 48.8 number of trips, each household has spent over $136 on soda over a period of 2 years.

**Figure 4c.**

nTrips

Quantiles		
100.0%	maximum	705
99.5%		380.73
97.5%		237
90.0%		115
75.0%	quartile	61
50.0%	median	27
25.0%	quartile	10
10.0%		4
2.5%		1
0.5%		1
0.0%	minimum	1

Summary Statistics	
Mean	48.869854
Std Dev	64.932341
Std Err Mean	1.3240474
Upper 95% Mea	51.466247
Lower 95% Mean	46.273462
N	2405

Brand Loyalty and Price Sensitivity based on household:

This graph shows percentage of brand loyalty. On the Y axis we have highest percentage of soda brought by a household from one manufacturer. For example, if a household buys 25%, 10% and 65% of soda from three different manufacturers, this means he was 65% loyal to one manufacturer, which is reported on Y axis. There appears to be moderate amount of loyalty to one manufacturer here. On and average households are loyal 60% towards one manufacturer, but they prefer soda from other manufacturers may be due to higher discounts. Most of the brand loyalty comes due to discounts. We observed that according to this graph, households that show a full 100% brand loyalty, that buy products from just 1 or 2 manufacturers aren't the ones that buy soda often. They show a high brand loyalty in the graph because they bought the product probably just once or twice.

Figure 4d.
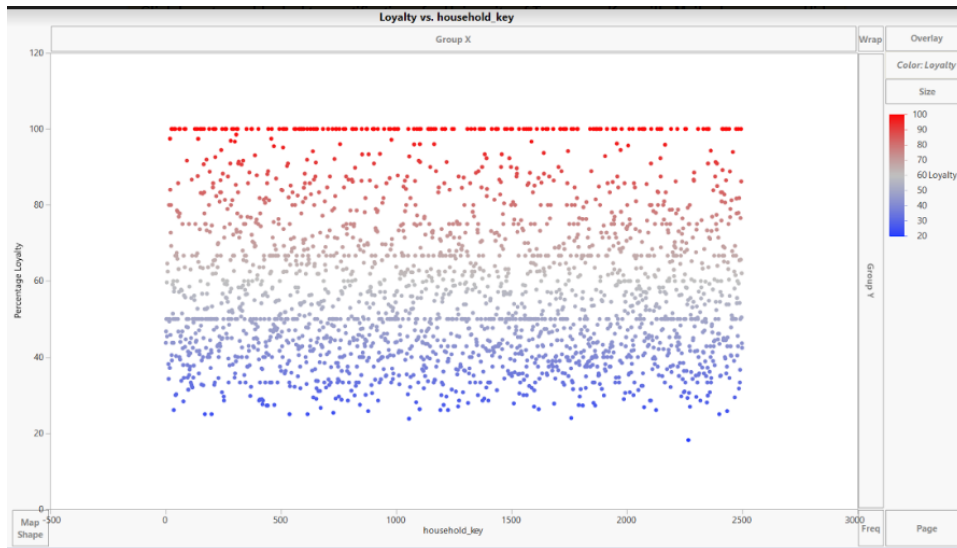
Price sensitivity per household

After conducting a Lack of Fit test on the average discount on quantity, we found that households are price sensitive and would buy more if more discounts were offered. The negative slope of -661.62 indicates that for every increase of $1 in discount, the quantity sold increases by 661.62. The P value of the ANOVA indicates that the test is significant and the model is a good fit.
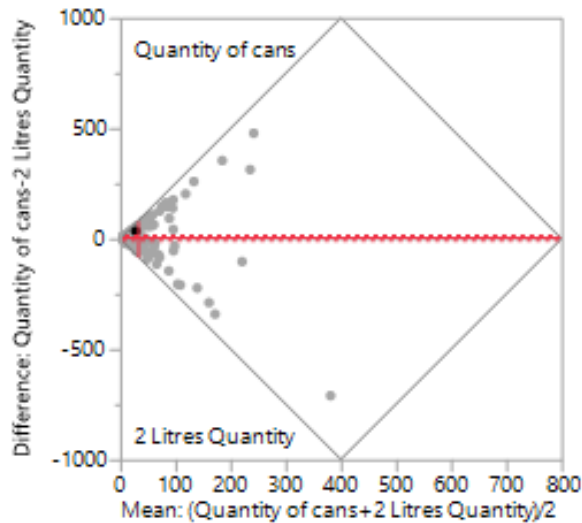
Figure 4e.
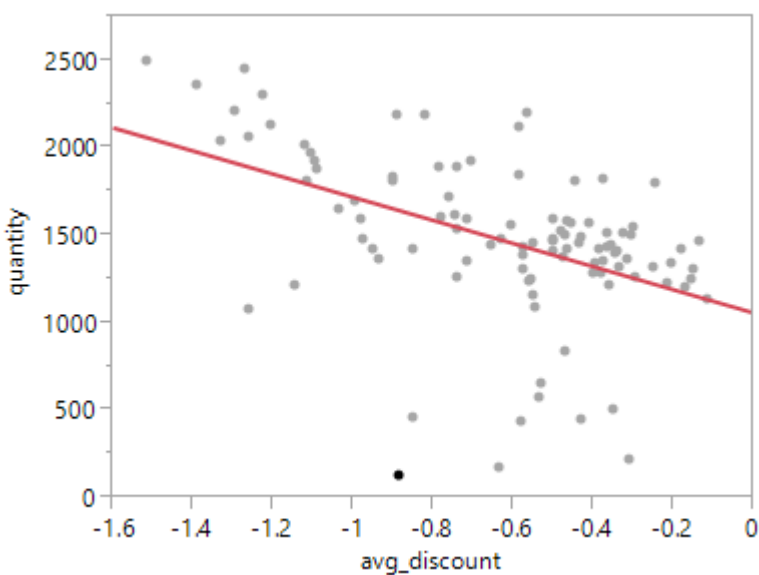
quantity = 1058.0995 - 661.62663*avg_discount

Analysis of Variance

| Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	4822046	4822046	27.9733
Error	100	17238033	172380	Prob > F
C. Total	101	22060079		<.0001*

2 Litres vs Cans:

Difference: Quantity of cans-2 Litres Quantity

Figure 4f.

This is matched pair test between the volume of cans sold and volume of 2 Liter bottles sold on a weekly basis.

On an average 35.09 cans of soda were sold vs 28.6 bottles of soda. On an average, cans reported 6.49% more in volume sale than bottles. The reason behind this is highlighted in the next matched pair test which summarizes the differences between the discounts offered by cans vs 2 Liter bottles on a weekly basis.
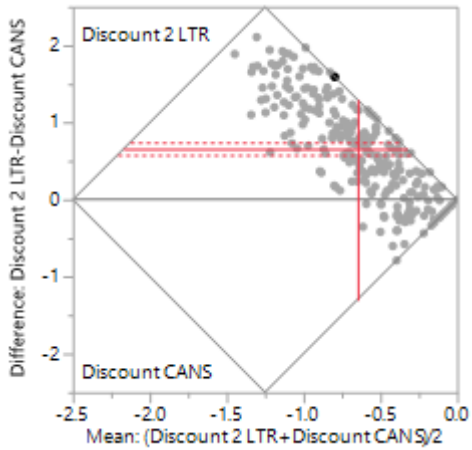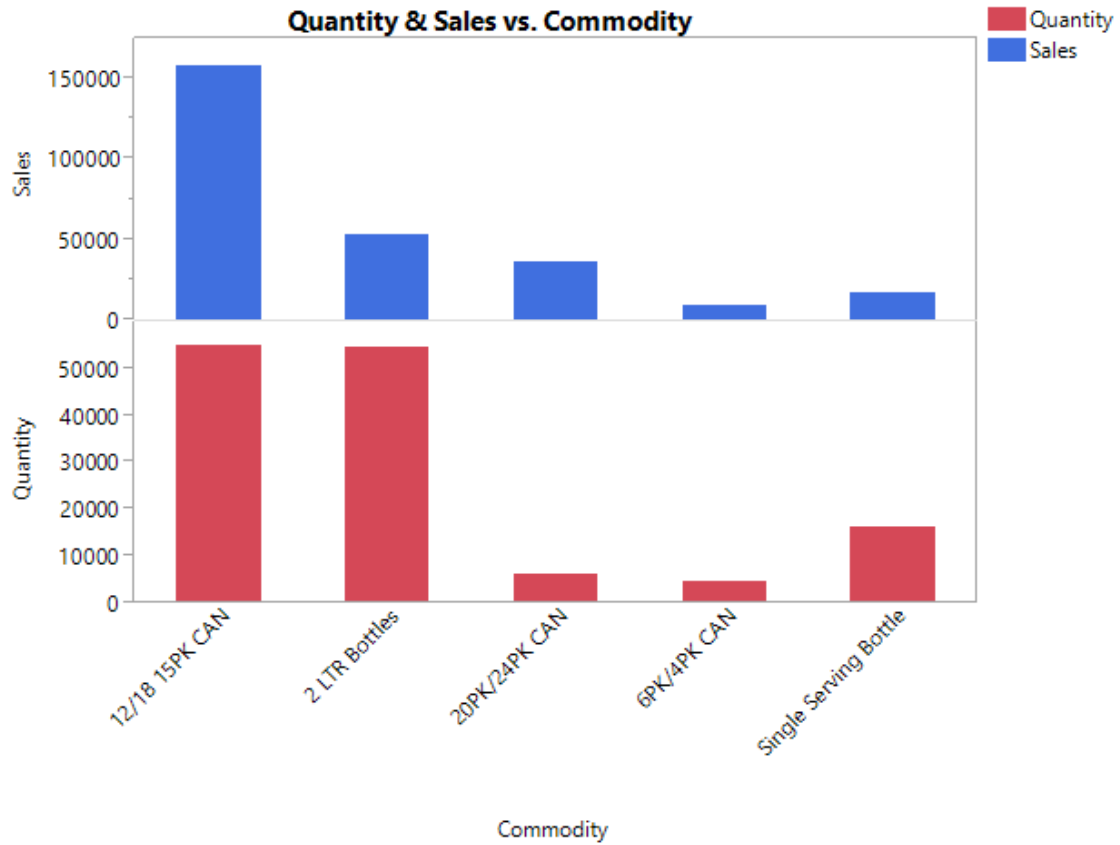
Figure 4g.

Difference: Discount 2 LTR-Discount CANS

Mean discount on 2 Ltr bottles is \$0.3175 and the mean discount for cans is \$0.9782. This explains why cans were sold in higher volumes than 2 Liters bottles.

Figure 4e.

Which categories are more common?

Quantity & Sales vs. Commodity

From **Figure 4e.** it is evident that 12/18 15PK cans is the most common commodity in terms of quantity sold but in terms of sales amount 12/18 15PK cans and 2Ltr bottles generate equal amounts of revenue.

Table 5a.

	Sales(Quantity)	Discount	Average Discount
Other Nationals	28390.18	5426.19	0.191
Top 3	262135.26	100044.18	0.38165
Private	37121.86	2307.91	0.062