Introduction

This report will work on the database from Bronnenberg, Bart J., Michael W. Kruger, Carl F. Mela. 2008. Database paper: The IRI marketing data set. Marketing Science. This purpose of IRI data set is to enable academic research for empirical validation of important theories, and to the creation of new theories. This database is a collection of several datafiles like Store data sets, Delivery Stores, Panel, Panel demographics, product attributes, IRI week translations and many more. In this report we will work on the store-level scanner data from two local markets - namely Eau Claire and Pittsfield - reporting sales, revenue measures and marketing activities of carbonated beverages in several different stores The data covers the period from January 2004 till December 2006.
The dataset consists of 21 variables showing the different brands or products of carbonated drinks being sold across stores in Eau Claire and Pittsfield over a period of 3 years from 2004-2006. The dataset also shows the potential marketing activities and strategies implemented to impact the sales and revenue of carbonated drinks. Additionally, the dataset also contains evolution of carbonated drink attributes over time, thereby impacting overall sales and revenue of different brands. Some of the product attributes of carbonated drinks discussed in this paper type of carbonated drink (for instance Soda, Club Soda, Tonic Water, Seltzer water, or bitter lemon), flavor of the drinks, brands, the volume of the drinks sold as well as the packaging like Cans, plastic bottles and glass bottles. The variables that show the marketing and promotional strategies are the display and feature variables, which have been operationalized as dummy variables. The display variable shows the placement of the drinks in the store and they have been categorized as major and minor. The feature variables show the advertising approaches taken by the stores and they have been categorized as small, medium and large, potentially showing the scope and extent of promotional activity.

Checking fit of data for analysis

The data set contains no missing values, which is advantageous for data preparation and analysis. There are outliers detected in the data such as in the volume equivalent, revenue, sales (units sold) and price (Figure 1). Although on further checking the data it can be realized that these extreme values are genuine data points, showing the high amount of revenue generated or sales done or high/low price being set. Therefore, we decide not to treat these extreme values in our data set but to incorporate them for further analysis.

However, there can be seen some inconsistency in the display and feature dummy variables. Apparently, these dummy variables have values between the range of 0 and 1, rather than 0 and 1. We handled this irregularity in the data by keeping a threshold of 0.5. The values which are less than 0.5 are assigned to 0 and values above and equal to the threshold are assigned to 1. This transforms all the display and feature dummy variables into binary variables with values 0 and 1.

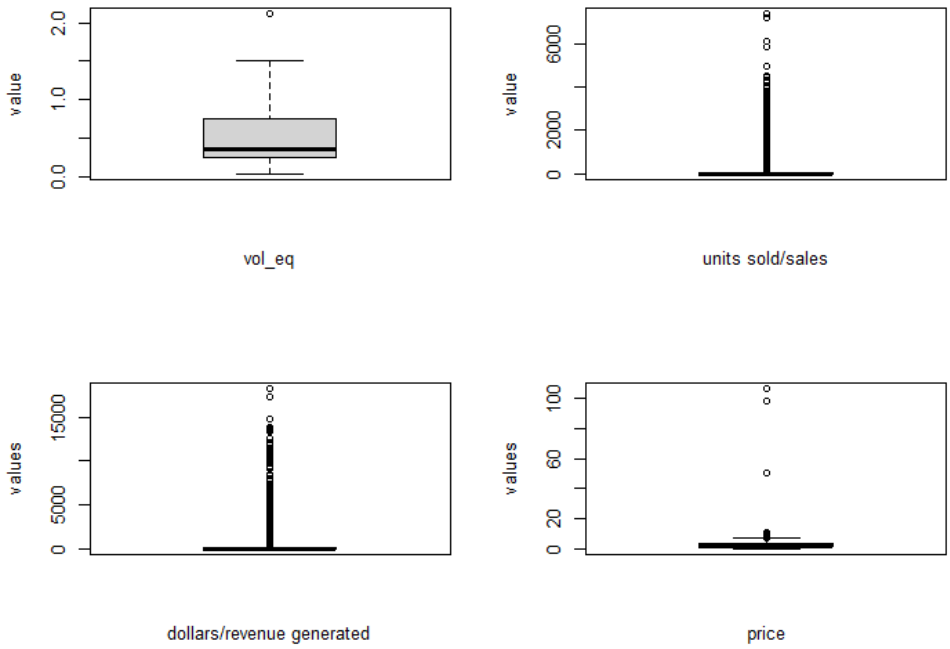

Figure 1: Outliers in the product attribute variable (vol_eq and price) and sales and

Data Description

The data preparation is followed by eyeballing the whole dataset. Table 1 shows the summary statistics of the entire dataset. The mean value of volume equivalent is 0.4779 which translates into 91.76 ounces or 2.75 liters of carbonated drinks. The average carbonated drinks sold and revenue generated from their sale in the stores was 22.88 and $50.4. This shows that the data has many extreme low values, meaning that many stores did not generate business from the sale of carbonated drinks. It can also be seen that the placement strategy was not widely implemented in the stores. The drinks were placed at the lobby end or aisle only about an average 12.59% (as can be seen from display_major variable) and only 5.77%(display_minor) times on other shelfs as compared to no placement strategy. Also, promotional activities seem to be low, with small advertisements on average being implemented 0.39%, medium ads 5.25% and large ads 4.42%. There seems to be a significant lack of marketing and promotional activities.

There are overall 5 types of products data available namely "soda", "seltzer water", "club soda", "tonic water" and "bitter lemon". Additionally, one can also that there are a variety of flavors of flavors of carbonated drinks available in the market, and the dataset managed to capture 161 such product flavors. There are 185 brands competing in the carbonated drinks market.

To further analyze the revenue generated by each flavor, we put a threshold of $0.5 million revenue and realize that the major contributor is the “Cola” flavor. From Figure 2, it is evident that the “Cola” Flavored carbonated drinks generate the highest revenue of $13.82 million in a period of 3 years, thereby contributing to 59% of the revenue among all the flavors which have a revenue of $0.5 million or more. Overall, Cola flavor contributes to 48% of all the revenue generated.

The Cola flavored drinks are followed by Lemon Lime, Dew, Root Beer with revenues $2.81 Million, $1.89 Million and $1.39 Million respectively.

	IRI_KEY	YEAR	WEEK	VOL_EQ	UNITS	DOLLARS	price
Min.:	213290	2004	1270	0.0365	1	0.1	0.01
1st Qu.:	233779	2004	1309	0.25	3	5.83	1.06
Median	259111	2005	1348	0.3521	8	13.5	1.59
Mean:	419464	2005	1348	0.4779	22.88	50.4	2.278
3rd Qu.:	652159	2006	1387	0.75	19	35	3.29
Max	1085053	2006	1426	2.1125	7375	18225.67	106.15

	display _minor	display _major	feature _small	feature _medium	feature _large	display _all	feature _all
Min.:	0	0	0	0	0	0	0
1st Qu.:	0	0	0	0	0	0	0
Median	0	0	0	0	0	0	0
Mean:	0.05767	0.1259	0.003869	0.05246	0.04418	0.1836	0.1005
3rd Qu.:	0	0	0	0	0	0	0
Max	1	1	1	1	1	1	1

Table 1: Summary of the numerical variables in the dataset

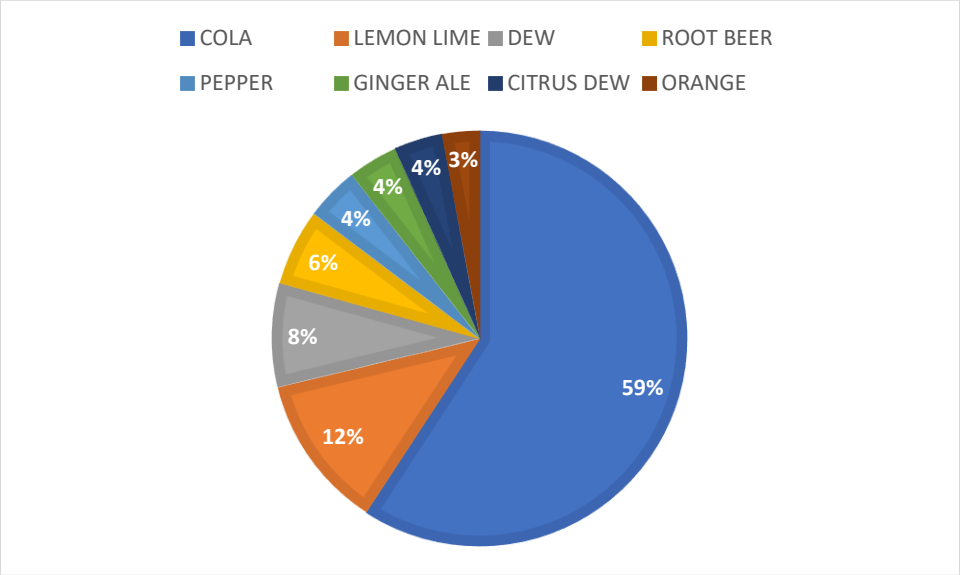

Figure 2: Percentage revenue generated by different Flavors

Eyeballing the COLA Flavored Carbonated Drinks

Going forward, the paper will look into the Cola Flavored Carbonated Drinks since they have the highest market share in terms of revenue generated. We subset the data to generate dataset containing information only of Cola flavored carbonated drinks. We looked into the correlations among the variables in the cola dataset (Figure 3) and realized there is correlation of package with

price (0.59), volume equivalent(0.53),vendor(0.37) and brand(0.41). Also, price and package have very high correlation of 0.87. The display and feature variables seem to be correlated with the vendor, and brand. Also, the amount of sales and revenue generated seem to be correlated with feature_large(0.25,0.19) and display_major variables(0.22 ,0.21) respectively. This also makes sense that taking aggressive promoting approach improves the sales and revenue.

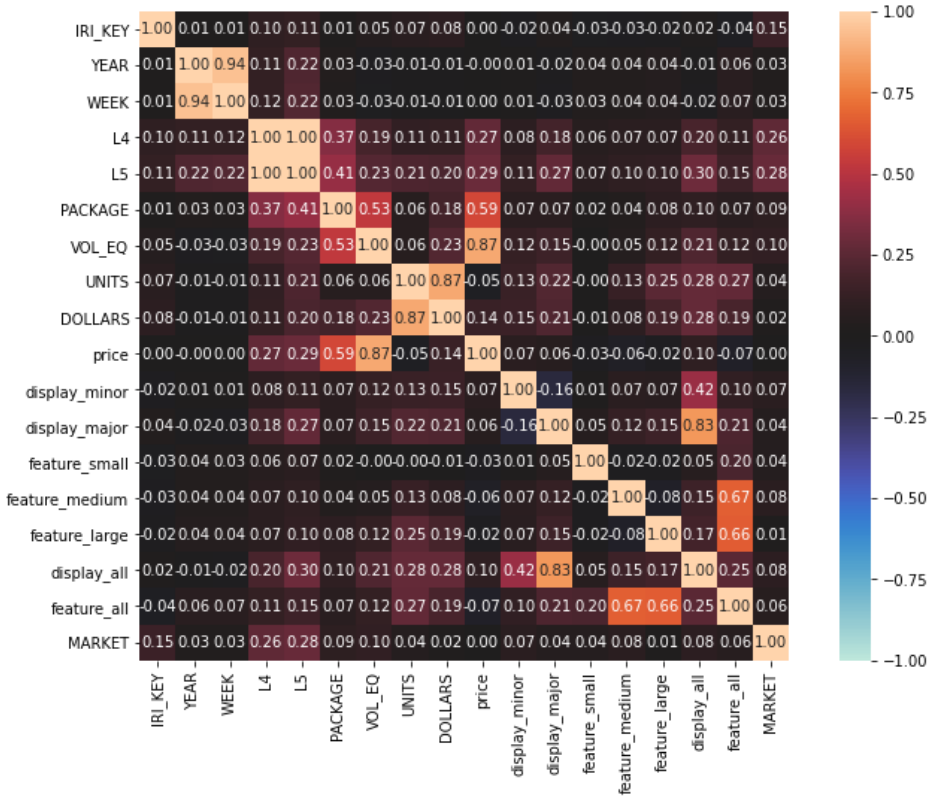

Figure 3:Correlations/Covariance Heatmap for variables in the dataset

From figure 4a it can be seen that placing the Cola drinks on lobby or end-aisle resulted in higher revenue as compared to minor display strategy. Additionally, figure 4b shows that more the scale of the advertising campaigns higher would be the revenue because feature_large has the highest revenue and the revenue decreases as the extent of advertising activity decreases. This is indicative that aggressive promotional activities are capable of higher revenue generation. Additionally, looking at the other variables which are the highest contributors in overall sales and revenue, we kept a threshold of $0.3 million revenue. We realized that the volume equivalent of 0.75 of 52% has the largest share in revenue generation followed by 1.5(23%),0.3521(14%) and 0.1042(5%) among all other options of volume equivalent, those which generated revenue more than 0.3 million (figure 5a). Meaning that 144 ounces (4 liters) of cola was dominating the market.

Furthermore, we selected the top 4 revenue generating volume options i.e. 0.75, 1.5, 0.3521 and 0.1042 to analyze which packaging option in combination with these volumes was generating highest market share. From Figure 6a, it can be seen that 0.75 volume equivalent CANS were the

largest revenue generator overall. In, plastic bottle packaging the dominant volume with respect to revenue is 0.3521. Apparently, the Glass bottles of Cola were had a fringe market share for all volumes. Furthermore, the price variation of different combinations of volume and packaging corresponds with revenue generated(figure 6b). The 0.75 volume CAN has low price as compared to 1.5 vol_eq CAN, which explains higher sale of the former. Also the 0.3521 vol_eq plastic bottle has lowest price and high revenue. This seemingly indicates that lower price can be a driving a factor in for higher sales and revenue. The Diet Coke and Coke Classic both contribute approximately 23% share each in the revenue (with 2.23 and 2.20 million respectively) generated by Cans followed by Pepsi (18%) and Diet Pepsi (14%). However, the scenario is different in case of plastic bottles. Coke classic and Pepsi have approximately 21% share in revenue generation with $ 0.84 million, followed by Diet Coke (18%) and Diet Pepsi (14%). Seems like people prefer to buy sugar Cola drinks in plastic bottles and non-sugar Cola in Cans.

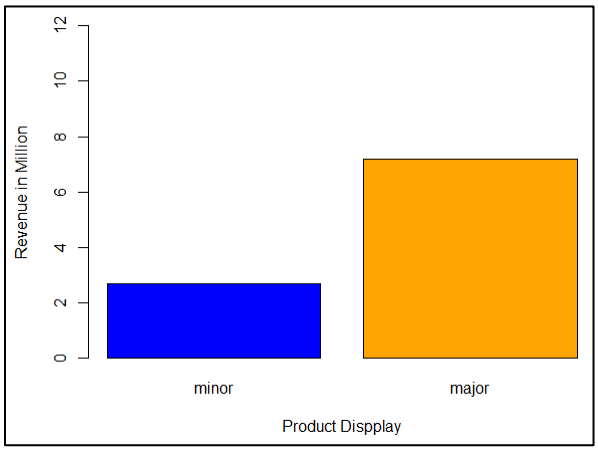

Figure 4a: Revenue generated by placement strategy for Cola flavored carbonated drinks

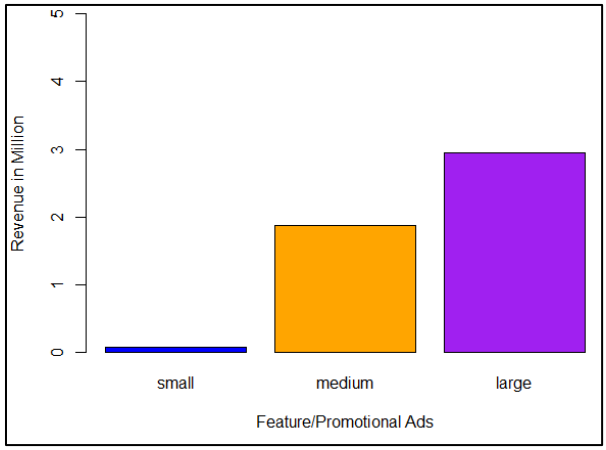

Figure 4b: Revenue generated by advertisements for Cola flavored carbonated drinks

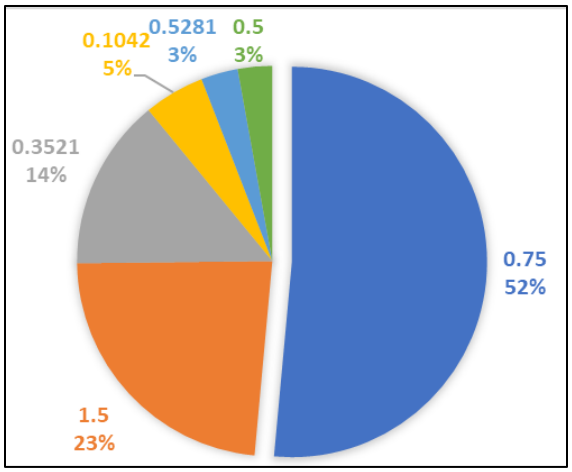

Figure 5a: Percentage Revenue Generated by volume equivalent

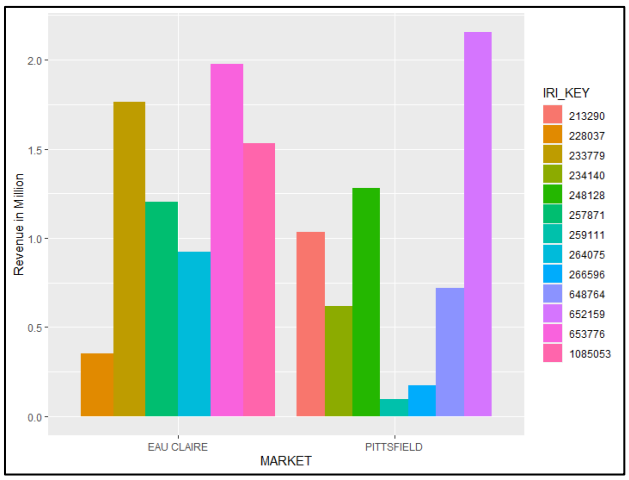

Figure 5b: Revenue generated by stores (given in IRI key) across both the market (Eau Claire and Pittsfield)

From figure 5b, we see that store with IRI keys 653776 and 228037, generate highest revenue for Eau Claire market and Store 652159 generate highest revenue for Pittsfield market. The store with IRI key 228037, generates highest revenue from Diet Coke of $ 93865.15 (27%) followed by Coke Classic (24%), Pepsi(15%) and diet Pepsi(14%). Exact same trend is followed by the store 653776 in Eau Claire, with Diet Coke having a share of 26% in revenue ($ 513,271.51) followed by Coke Classic, Pepsi and Diet Pepsi. However, for the store 652159, which is the highest contributor in business revenue, the trend is different. Pepsi is the highest contributor in revenue of $ 504028.55, 23%, followed By Coke Classic, Diet Pepsi and Diet Coke. It seems that the non-sugar carbonated were not able to penetrate the Pittsfield market and people still prefer conventional drinks like Pepsi and Coke Classic. Even though Pepsi has the highest mean price ($ 3.11) among Coke Classic, Diet Coke and Diet Pepsi, yet its revenue is higher than other three. This can be an indication of peoples' preference towards Pepsi in Pittsfield.

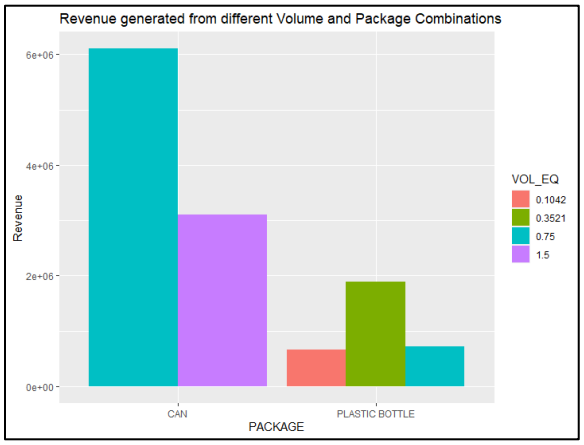

Figure 6a: Revenue from different combination of volume and package

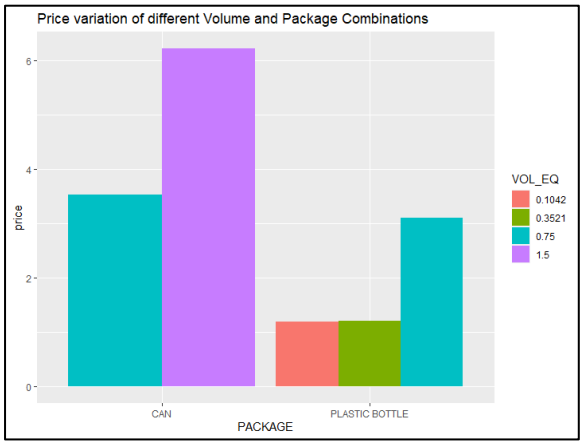

Figure 6b: Price variation from different combination of volume and package

Estimating the impact of marketing

To identify drivers in Sales, we started with a look at the development of sales over time. This helps identify special points in time through local maxima later as well. Firstly, there are different trends for Coca Cola- and Pepsi-owned brands.

While Coke Classic and Diet Coke are steadily increasing their revenue, Pepsi and Diet Pepsi follow a downward trend. It also appears that stronger fluctuations appear more often for Coke products, which indicates that they receive more noticeable events regarding marketing actions.
Prices and sold units are what drives Sales revenue. Therefore, looking at the patterns of price development over the same period as sales is the next step towards discovering their connection. Apart from Pepsi in the middle of the third year, the courses of all product prices look extremely similar, with Coke products being more expensive than Pepsi products.

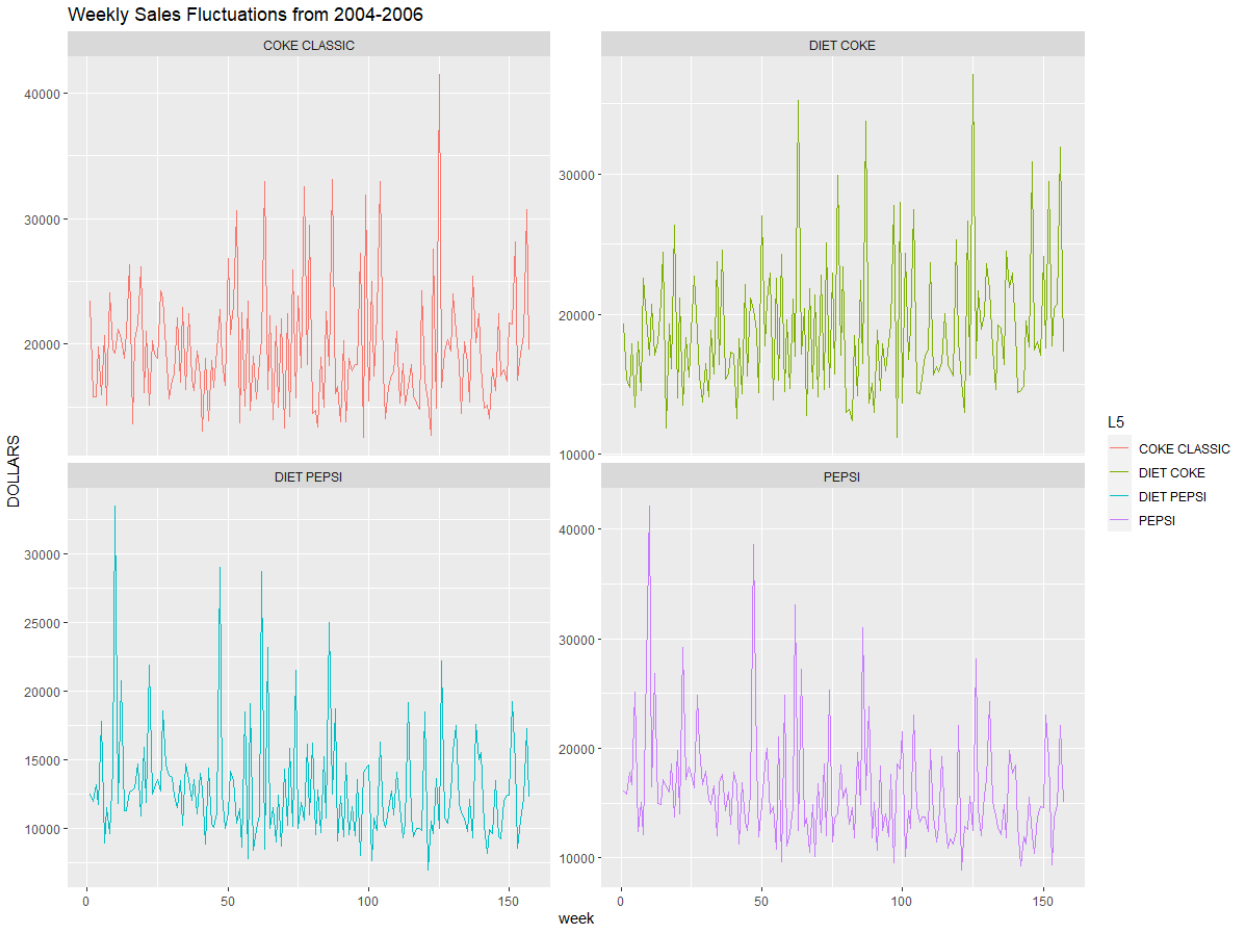

Figure 7a: Weekly Sales revenue fluctuations by product

Prices and sold units are what drives Sales revenue. Therefore, looking at the patterns of price development over the same period as sales is the next step towards discovering their connection. Apart from Pepsi in the middle of the third year, the courses of all product prices look extremely similar, with Coke products being more expensive than Pepsi products.

The trends in Sales revenue and price commute positively for Coke products but indicate different directions for Pepsi and Diet Pepsi respectively. At this point, a unidirectional relation between price and revenue is questionable at best, given that our goal is to estimate the effects of pricing and promotion on Sales and their revenue. To perform a proper estimation, we compared the three best weeks of each product with their previous week. In this comparison we included revenue, units sold and price.

Surprisingly, there was no noticeable difference in prices, even across different package variations. Still, units sold increased between 23% at the lowest (Coke) and 204% at most (Pepsi). Therefore, the prevalent conclusion is that pricing did not impact the changes in Sales revenue significantly, but promotions did. In addition to this comparison, we also ran two regressions for units sold. The first linear model postulates our hypothesis for minimal impact of prices on sales, whereas the second model is used to prove the validity of our claim.

	Intercept	Large feature	Medium feature	Small feature	Major display	Minor display	price
LM 1	105	213	134	-229	324	214	
LM 2	-10	215	157	-216	318	206	41

Table 2: Comparison of linear models regarding sales

It can be seen that the inclusion of pricing does not change the impact of promotional activities, and that the sales fluctuate far less with price. Furthermore, if discounts increased sales, the coefficient for prices should be negative. The positive, but small coefficient of 41 again summarizes the tale of steadily increasing prices with slightly increasing sales. We consider our hypothesis proven: The vast majority of unnatural fluctuations in sales are attributed to promotions.

Figure 7b: Weekly price fluctuations by product

Going Niche: Coke Classic Data

Finally, we take a look at a single product to eliminate influence from other brands in the field.

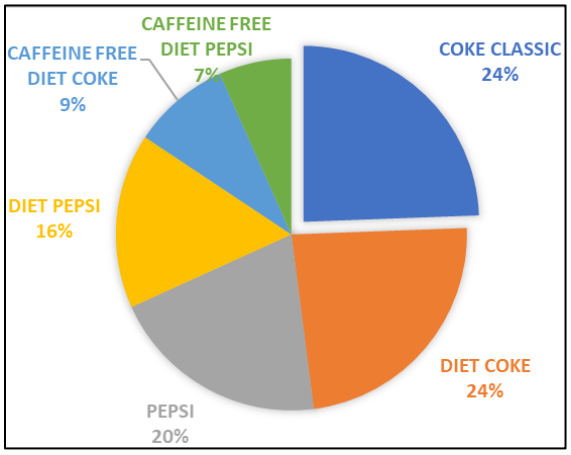

Figure 8a: Revenue generated per Cola brand

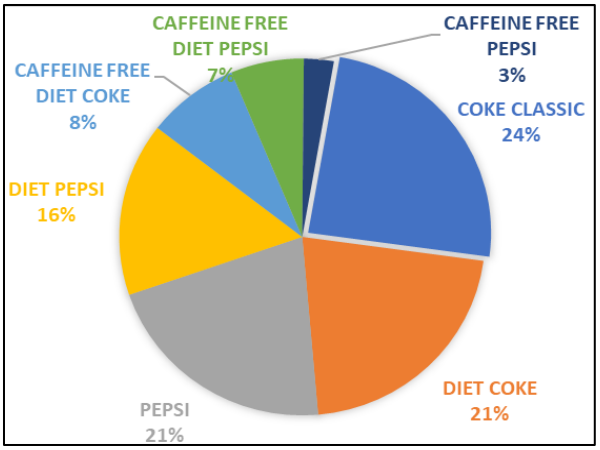

Figure 8b: Units sold per Cola brand

As shown in figures 8a and 8b, Coca Cola is the biggest competitor in the market. It is because of the sold items that we choose Coke Classic over Diet Coke. Once again, the volume equivalents of 0.75 and 1.5 dominate sales, with 0.3521 following in third. Sales revenue for Coke Classic also stayed consistent at $1.01m, $1.04m, and $1.01m in over the three years. A look at the price per volume equivalent also shows that the three previously mentioned quantities result in the lowest price per liter of Cola. This proposes that there is indeed an influence of price on the choice of bought product. So, while price changes are not necessary to increase sales, they do build a decision basis for the customer to even before marketing influences. This intel can be used to find out which products, packaging, and volume equivalents to promote stronger.

Interestingly, the 0.3521 volume is the only Coke bottle (2 liters) that is preferred by customers. Its pricing makes it advantageous in comparison to 1 liter, and its size makes it more convenient than the 4-liter bottle, which is also slightly worse in terms of price/performance ratio.

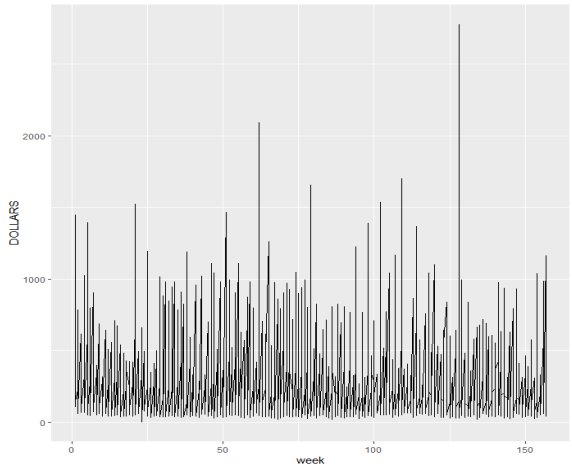

Figure 9a: revenue Coke classic over time

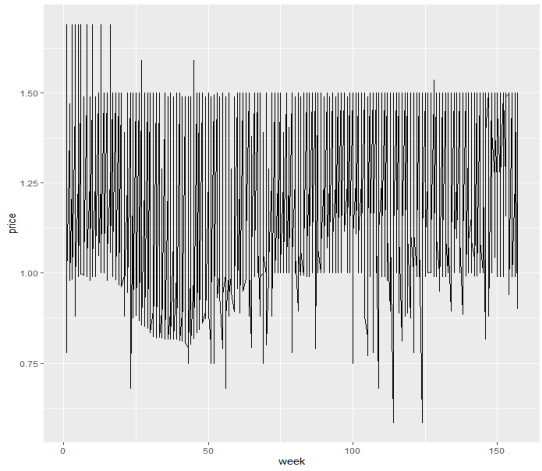

Figure 9b: price Coke classic over time

Once again, our hypothesis is reinforced by the example of a single product. The other products, cans worth 0.75 and 1.5 volume equivalents, show behavior similar to figures 9a and 9b. The spikes in sales revenue came seemingly independent of price fluctuations. However, it is noteworthy that canned Coke Classic experienced slight increases in price towards the end of the observed timeframe, whereas the 2-liter bottle did not.

In conclusion, we estimate the impact of price changes as negligible in comparison to marketing tools like ads and displays when it comes to evaluating extreme fluctuations in sales. Following this rule of thumb, the story of this dataset promotes more aggressive, and more marketing in general. The decline of Pepsi in comparison to Coke products is a result of worse and less marketing decisions.

Another point specifically for the soft drink industry could be that consumers are ignorant towards prices once they have chosen their favorite product. This also marks a caveat for further testing from Coke and Pepsi. The lack of a drastic policy for price changes makes marketing the primary driving force in our observed timeframe, where pricing has either been tested already or is yet to be researched for today's soft drink market.

As a reference for who wrote which text: Everything up to the Diagrams on page 6 was written by Akanksha, and everything following this was written by Max.