Humboldt University Berlin

Institute of Marketing

Prof. Dr. Daniel Klapper

Advanced Marketing Modeling

SS 2021

Special Work Performance 1: Describing Data

This is non-graded individual work.

Your answers including all tables and graphs must not exceed 10 pages. Please start a new page when providing your report to a new subtask. Please use typeface Times Roman in 12pt with 1.15 line spacing (in tables and graphs you may use 10pt and 1.0 line spacing) and 1 inch space on all sides. Do not forget to report your name and student number and a page number on each page starting with number one on the first answering page.
Do not include a title page or content page.
Send your report as pdf to my email address daniel.klapper@hu-berlin.de not later than May, 07, 2019, 4:00pm. Please report in the subject line "AMM SWP1 and your name".

Download the dataset "data.iri_key.prepared.csv" from the Moodle course page. The data are store-level scanner data from two local markets, reporting sales, revenue measures and marketing activities of carbonated beverages in several different stores The data cover the period from January 2004 till December 2006. Your dataset has 57786 (first row reports the variable names) data rows and 20 columns. The columns provide information about the following criteria:

1. IRI_KEY
2. YEAR
3. WEEK
4. PRODUCT.TYPE
5. FLAVOR.SCENT
6. L4
7. L5
8. VOL_EQ
9. PACKAGE
10. UNITS
11. DOLLARS
12. price
13. display_minor
14. display_major
15. feature_small
16. feature_medium
17. feature_large
18. display_all
19. feature_all
20. MARKET

Additional information about the data is found on Moodle and is discussed in class.

This special work performance is designed that you get familiar with your data, to learn to explore data structures and to estimate the effect of price and promotion on sales.

Describe the data in a meaningful way, e.g. report sales and marketing activities of products or brands within and across stores. Use tables and graphs to support your description and explain in words the key facts of the data set. Also document your estimation strategy in some detail. Do not report R-codes and edit the estimation results you obtained with R.

AMM SWP 01 – Describing the Data

In this first special work performance a specific part of a panel data, collected by the International Market Research Organization IRI, containing information about the stores in US and the products over time. More specifically, mostly to simplify and to have more compact data, the data examination will be kept limited to carbonated beverages product category and two markets namely that, Eau Claire and Pittsfield.

Data Explanation

The data set has 579785 observations and is based on store data, 7 in Pittsfield 6 in Eau Claire, over the years of 2004, 2005 and 2006. The data set has also product specific information such as type, flavour, manufacturer, brand, price, packaging, volume and promotional activites, all of which will be examined and elaborated on in detail below.

BITTER LEMON	CLUB SODA	SELTZER WATER	SODA	TONIC WATER
20	8549	42495	514986	13735
(0.00003%)	(0.01%)	(0.07%)	(0.89%)	(0.02%)

Figure1. Beverage types

Above, in Figure 1, one can get a grasp of information such as which carbonated beverage type is demanded/produced more and how they compare with each other. Obviously, soda has the biggest share and is the most important one. The other three that adds up to 10% of the market likely are less demanded and less profitable. And lastly, bitter lemon beverage is most probably a new product test by Dr. Pepper but it is still important to report that to get the full picture of the market.

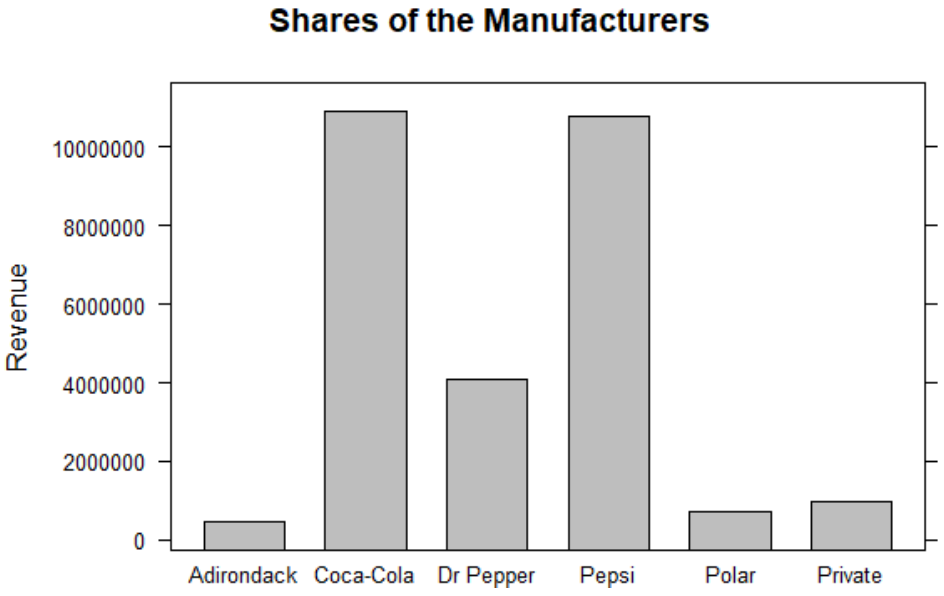


Figure 2. Shares of the Manufacturers based on Total Revenue

In the data set there are 57 manufacturers of different sizes. In figure 2, the market shares that are based on the total revenue of the 6 biggest players in the market are shown, which, when added all together, equals to 98.34% of the whole market. This leaves only 01.66% for the remaining 51 manufacturers. Based on the bar chart above one can conclude that Coca-Cola and Pepsi are playing for the market leadership together although Coca-Cola has slightly higher revenues. 3rd rank is occupied by Dr. Pepper and the rest are sort of competing for whatever is left from the big corporations.

	BITTER	LEMON	CLUB SODA	SELTZER WATER	SODA TONIC WATER
COCA COLA	0	0	0	123366	0
PEPSICO	0	0	0	127593	0
DR PEPPER/SEVEN-UP	20	5429	36	122189	7846
POLAR CORP	0	1448	16939	23551	0
ADIRONDACK	0	0	14304	25387	813
PRIVATE LABEL	0	1336	6550	43025	1151

Figure 3. Product Distribution Based on Manufacturers

Figure 3 illustrates which and how many carbonated beverage types the companies that almost corresponds to whole market sell. One can see that the predominant product type that is sold is Soda, which corresponds to 96.19% of the all products sold while the rest corresponds only to 3.81%.

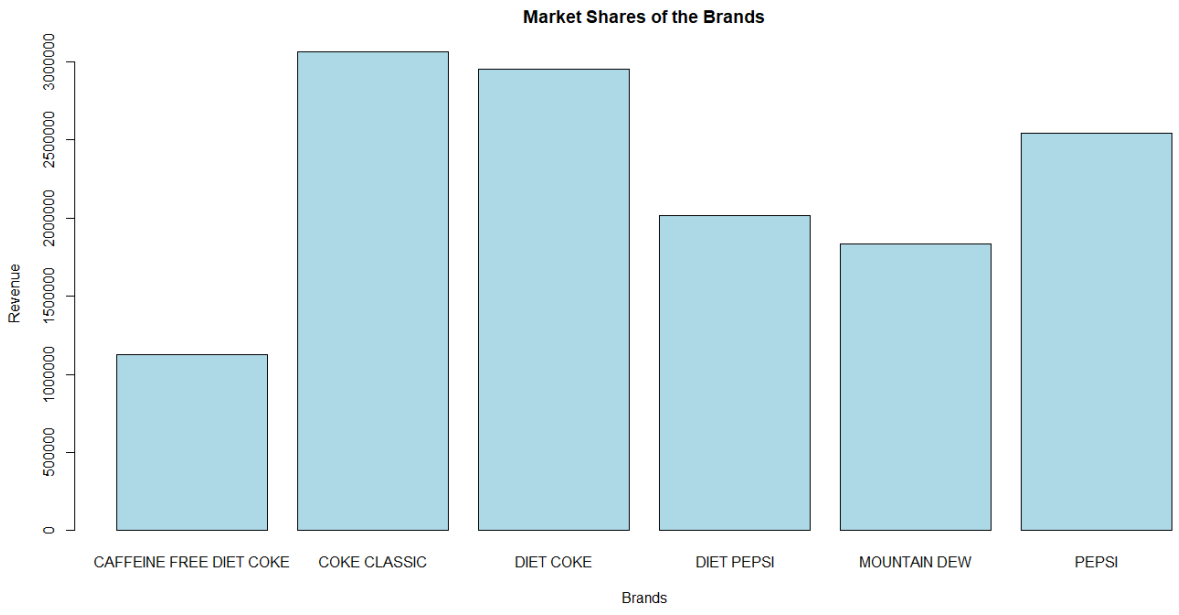


Figure. 4 Market Shares of the Brands (Based on Revenue)

In the data set there were 187 different brands produced by different manufacturers. In figure 4, the revenue based market shares of the brands that have at least 1 million dollars of revenue are displayed. These 6 brands corresponds to the 47.58% of all of the brands with regards to the revenue they make. Products can be categorized in many ways such as based on flavours, manufacturers or ingredients like caffeine, sugar etc. If we base our analysis on the last categorization mentioned, we can conclude that in diet coke category Coca-Cola shows a better performance; whereas, in regular coke Pepsi seem to do have higher revenues as Mountain Dew and Pepsi both belongs to PepsiCo. Since the first three brands from left belongs to Coca-Cola and first three brands from the left to Pepsi, one can conclude that the competition between the two corporations is maintained at the brand level.

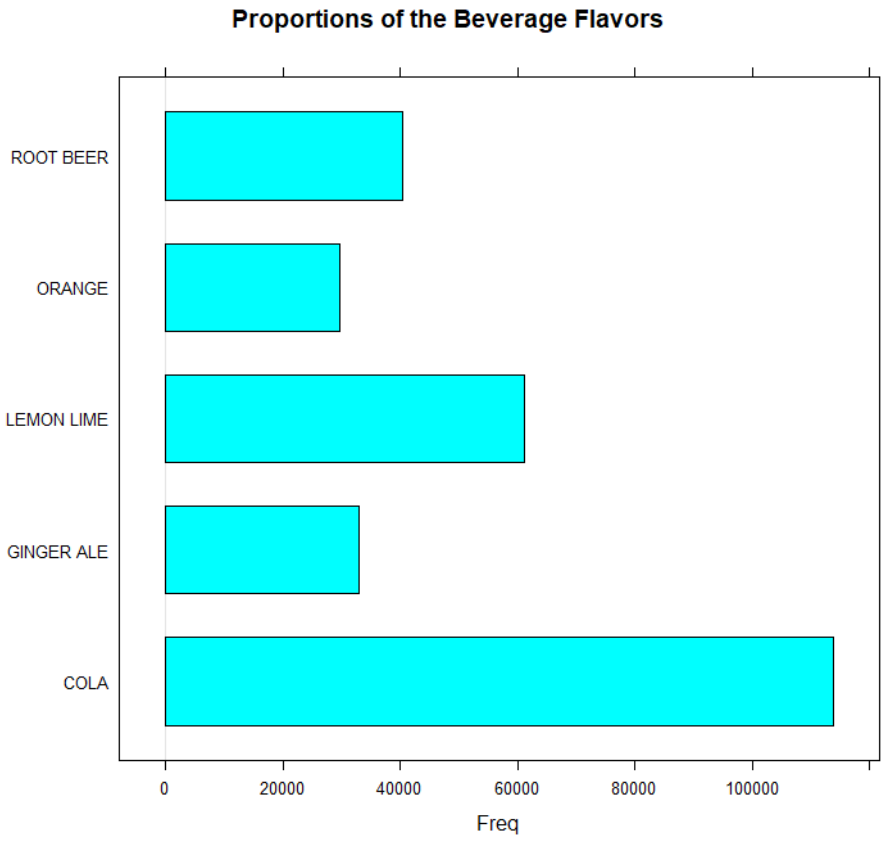


Figure 5. Proportions of the Beverage Flavors (Based on Revenue)

In the data set there is 156 different beverage flavors. The reason why these specific flavors were chosen is that all of the flavors that are not displayed above had a market share smaller than 0.05% and the ones that are displayed were the only ones that had a higher share of 0.05%. Infact, the above flavors corresponds to 47.95% of all of the flavors with regards to revenue they bring when compared to remaining 151 ones. Examining the figure above, one can observe that Cola is the predominant flavour in the market, which constitutes 48.56% of the all revenue by itself.

Market Comparisons

	MARKET	UNITS	DOLLARS
1	EAU CLAIRE	6320431	16597521
2	PITTSFIELD	6594595	11852440

Figure 6. Market Shares

Firstly, a sensible thing to check is the market sizes. Looking at the figure 6, although at first glance Pittsfield might seem bigger because of the number of the units sold there are higher than Eau Claire, evaluating the market size in terms of total revenue is more practical and can be helpful in making decisions because the firms care about the total revenue more than they do units sold.

	YEAR	UNITS	DOLLARS
1	2004	4354575	9586254
2	2005	4378504	9531541
3	2006	4181947	9332166

Figure 7. Yearly Overall Revenues

: 2004	: 2004
: EAU CLAIRE	: PITTSFIELD
[1] 5546877	[1] 4039378
: 2005	: 2005
: EAU CLAIRE	: PITTSFIELD
[1] 5503450	[1] 4028091
: 2006	: 2006
: EAU CLAIRE	: PITTSFIELD
[1] 5547194	[1] 3784972

Figure 8. Total Revenues Based on the Market

Figure 7, on the other hand, shows a slight decrease in the whole revenue in total by the two market together in the year of 2006. To understand why that is the case, figure 7 sheds some light on the issue. Carefully observing the table in figure 8, one can conclude the decrease in total revenues comes from not both of the markets but rather only Pittsfield.

Non-Coca-Cola	Coca-Cola	Non-Pepsi	Pepsi
: 2004	: 2004	: 2004	: 2004
: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE
: FALSE	: TRUE	: FALSE	: TRUE
[1] 3458985	[1] 2087892	[1] 3166621	[1] 2380256
: 2005	: 2005	: 2005	: 2005
: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE
: FALSE	: TRUE	: FALSE	: TRUE
[1] 3278804	[1] 2224646	[1] 3317624	[1] 2185826
: 2006	: 2006	: 2006	: 2006
: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE	: EAU CLAIRE
: FALSE	: TRUE	: FALSE	: TRUE
[1] 3242056	[1] 2305138	[1] 3418933	[1] 2128261
: 2004	: 2004	: 2004	: 2004
: PITTSFIELD	: PITTSFIELD	: PITTSFIELD	: PITTSFIELD
: FALSE	: TRUE	: FALSE	: TRUE
[1] 2626301	[1] 1413077	[1] 2603418	[1] 1435960
: 2005	: 2005	: 2005	: 2005
: PITTSFIELD	: PITTSFIELD	: PITTSFIELD	: PITTSFIELD
: FALSE	: TRUE	: FALSE	: TRUE
[1] 2530763	[1] 1497327	[1] 2667971	[1] 1360120
: 2006	: 2006	: 2006	: 2006
: PITTSFIELD	: PITTSFIELD	: PITTSFIELD	: PITTSFIELD
: FALSE	: TRUE	: FALSE	: TRUE
[1] 2394700	[1] 1390272	[1] 2483530	[1] 1301442

Figure 9. Revenue Changes of Pepsi & Coca-Cola in Different Markets Over Time

In Figure 9, looking at Coca-Cola's revenue change in different markets over time, we observe that while its competitors are losing revenue in the Eau Claire market, Coca-Cola's revenue is consistently increasing in all of the years, which is a good sign for the company. For Coca-Cola, both its competitors and it is losing revenue in the Pittsfield market as the market is shrinking for everyone. For the Pepsi side and the Pittsfield market, basically the shrinkage of the market is also observed; however, there is an important remark to make regarding the Eau Claire market for Pepsi. There are many manufacturers in the market but since Pepsi's rivals' revenues are increasing and its revenue is decreasing and since Coca-Cola and Pepsi are competing directly with each other, one may interpret the situation as Coca-Cola is stealing some of the market share of Pepsi through the years of 2004 to 2006 in the market of Eau Claire.

	CAN	GLAS BOTTLE	PLASTIC BOTTLE
EAU CLAIRE	127822	20584	129823
PITTSFIELD	90318	17810	193428

Figure 10. Packaging Proportions Across Markets

CAN	GLAS BOTTLE	PLASTIC BOTTLE
218140	38394	323251
(0.37%)	(0.06%)	(0.56%)

Figure 11. Packaging Proportions Overall

In figure 10, one can observe how the packaging of the products are distributed among different markets. Figure 11 shows that two primary packaging formats are plastic bottle and can, with the former being more used and the latter less. Additionally, figure 10 points out one of the peculiarities we have confronted before, namely that although in Pittsfield the number of products sold was higher, in terms of revenue it is smaller than Eau Claire (Figure 5). So, there must be a product that costs more. In Eau Claire market glass bottles and cans are higher in number. Probably both of them contributes to this fact in figure 5 but also in further analysis glass bottles were found to be more expensive per 1 litter of carbonated beverage (This might stem from high production cost of glasses or low recyclable value of it compared to cans or plastic bottles). So, this might explain the difference between the markets' units sold and the revenue differences between them.

: 2004	: 2004	: 2004
: CAN	: GLAS BOTTLE	: PLASTIC BOTTLE
[1] 6467375	[1] 119478.6	[1] 2999400
: 2005	: 2005	: 2005
: CAN	: GLAS BOTTLE	: PLASTIC BOTTLE
[1] 6439903	[1] 130621.9	[1] 2961016
: 2006	: 2006	: 2006
: CAN	: GLAS BOTTLE	: PLASTIC BOTTLE
[1] 6343158	[1] 134797.2	[1] 2854210

Figure 12. Packaging Distributions Over Time

Figure 12. shows how the packaging distributions change over the span of 3 years. Examining the table above, one can notice the mutual decrease of number of cans and plastic bottles while the number of glass bottles produced are increasing. This might be because of environmentalist consciousness increasing over the years as well as the information that glass bottles are healthier are spreading.

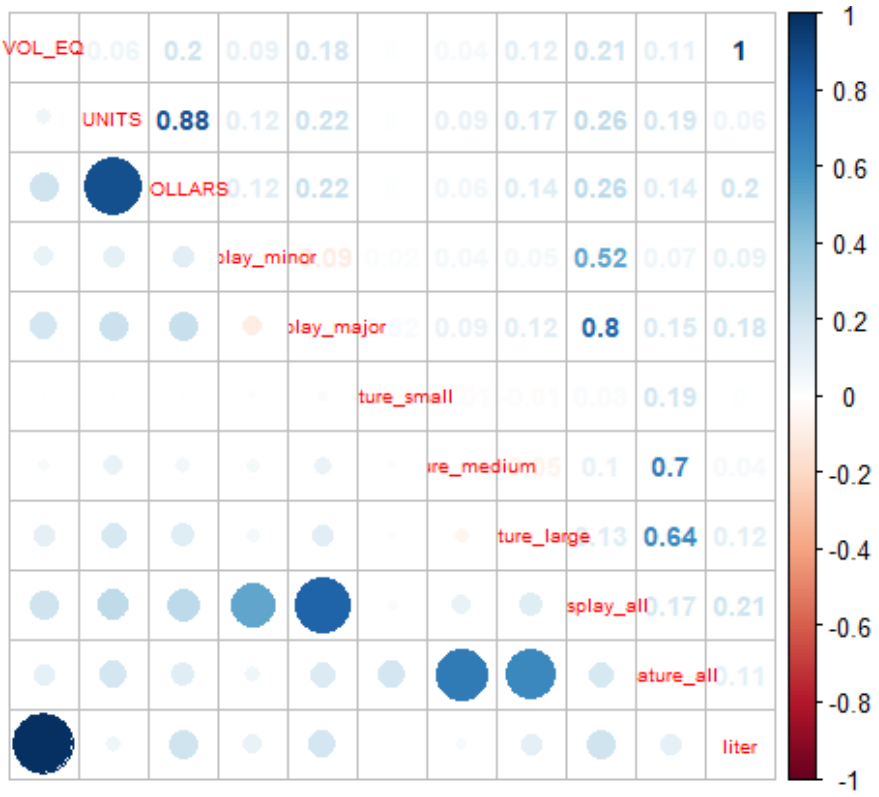


Figure 13. Relationship between promotional activites and sales

The correlation plot in figure 13 shows us to what extent the promotional activities such as increased display area in the retail store or some other advertisement strategy implemented by the retail store such as booklets etc. affect the sales thus also the revenues. Examining the figure above one can see positive relationship between display area and sales, namely that 0.12 of positive correlation with revenue and display minor and 0.22 with revenue and display major. As one would expect, larger display area gets the consumers' attention more, increases the chance of purchasing the item and results in higher sales of the displayed product. One another point that the correlation plot suggests is that small advertisements aren't necessarily effective in increasing the sales. Figure above suggest that while the largest advertisement seems to have more positive impact on the sales with positive correlation of 0.14 with the revenues, a big display area in the retail store end up increasing the sales more.

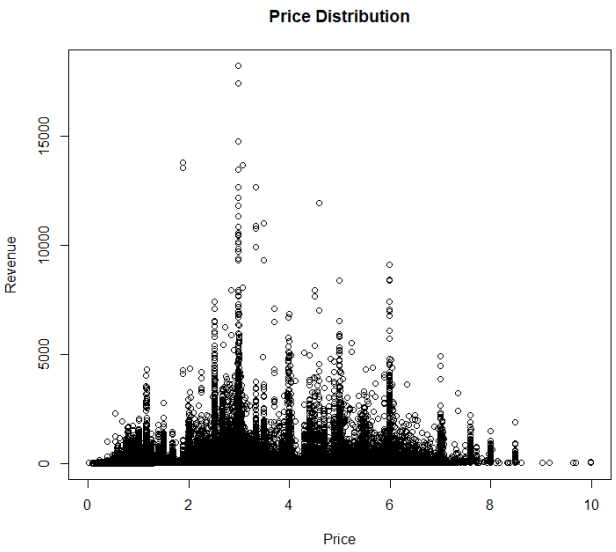


Figure 14. Price vs. Revenue Distribution

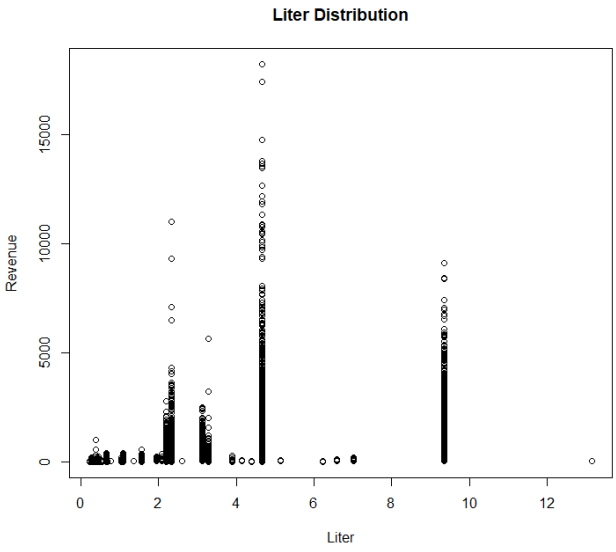
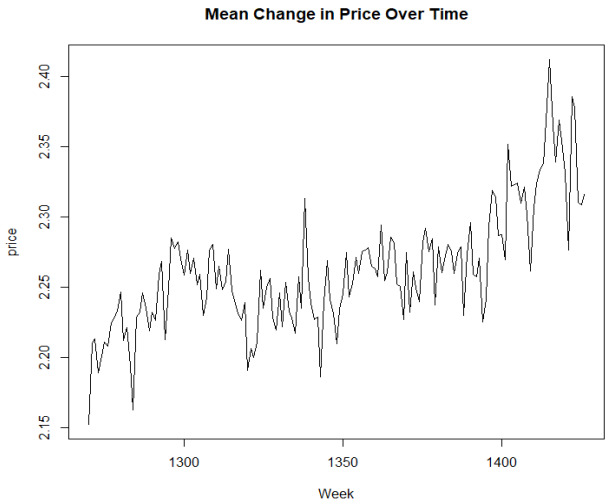


Figure 15. Liter vs. Revenue Distribution

In figure 14, one can see the price distribution of the beverages with regards to revenue they bring. Likewise, in figure 15, one can observe the the distribution of beverages in liters according to revenue they bring. Deciding on best price that brought the highest revenue in total is a hard task both because they change over time with discounts and promotions and also because they always change in different amounts, there are a lot of price categories that would need to be examined. However, there is one price-liter combination that brings the highest amount of revenues and possibly the highest amount of revenues in total also. In figure 14, we see a peak at the price of 2.98 dollars that resulted in high revenues. Examining the data set, it is found that the products with 4.67 liters (possibly a cartoon or box of beverages) that bring the highest revenues in figure 15 are sold at the price of 2.98 dollars also. Likewise this price-liter combination, the relationship with 2.33 liter beverages, 9.34 liter beverages and how they interact with prices might be further investigated to potentially come up with a optimal price per liter that would yield highest revenues.

Changes Over Time



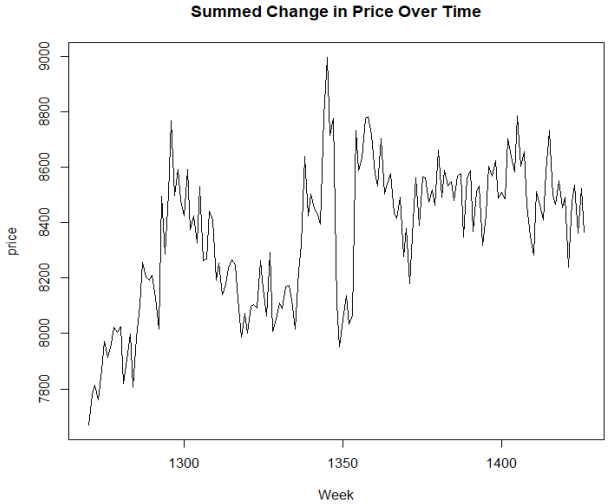


Figure 16. Average Price Change Over Time Figure 17. Overall Price Differences Between the Weeks

In the figures above the changes in price over time are analysed. In figure 16, the mean of the prices is calculated. There it is possible to observe some constant variation through out the weeks. One other observation that can be made based on figure 16 is that over the span of two years prices of carbonated beverages show an increasing trend, namely that while in the first weeks the lowest average price used to be around 2.15 dollars, towards the end of year 3 it becomes 2.30, which is something that is worth paying attention to. In figure 17, the total sum of prices within the weeks over 3 years is calculated. As we saw in figure 16, the prices show an upward trend here too. Also, there are some sudden changes in prices. These might be during some special events throughout the country or holidays. The reason why some of the changes in the prices are so steep could be that some big companies in the industry such as Coca-Cola or Pepsi increasing or decreasing its products' prices and the smaller companies being forced to go along with this decision of bigger firms to keep up the competition.

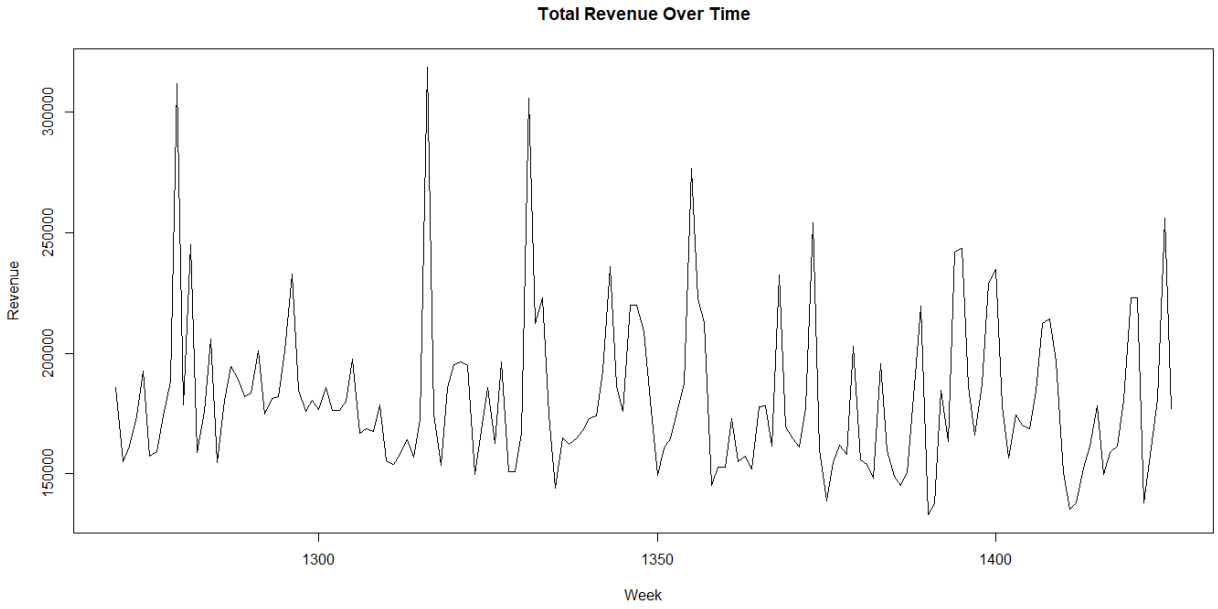


Figure 18. Total Revenue Over Time

The plot in figure 18 shows how the total revenue changes over time and the patterns it exhibits. Examining it, one may come up with the conclusion that there is slight decrease of revenues in the industry in total as peaks and downs correspond to less revenue than they used to in 2004. Another clear insight that might be taken from figure 18 is that there are very steep increase in revenues in certain weeks. As mentioned before, these might stem from the fact that there are some holidays, national days, religious ceremonies, big events like Super Bowl or similar ones that result in the sales to be high temporarily.

Figure 19. Total Revenue Over Time Based on Packaging

In Figure 19, the change in revenue based on beverage packaging is displayed. Blue line indicates cans, green line indicates plastics and the pink one indicates glasses. As one might conclude, the variation that we observed in previous plots result from cans and they seem to be the focus of the competition. Plastic packaged products don't show much variation in revenues while glass products show almost no variation in total over 3 years.

Conclusion

In general, the data set of IRI contains much more information across cities, different product categories and over the years. In this first special work performance, a general look through carbonated beverages market based on stores in Eau Claire and Pittsfield in US over the years of 2004, 2005 and 2006 was provided. With the data set available a more detailed analysis and examination is also possible, when desired.