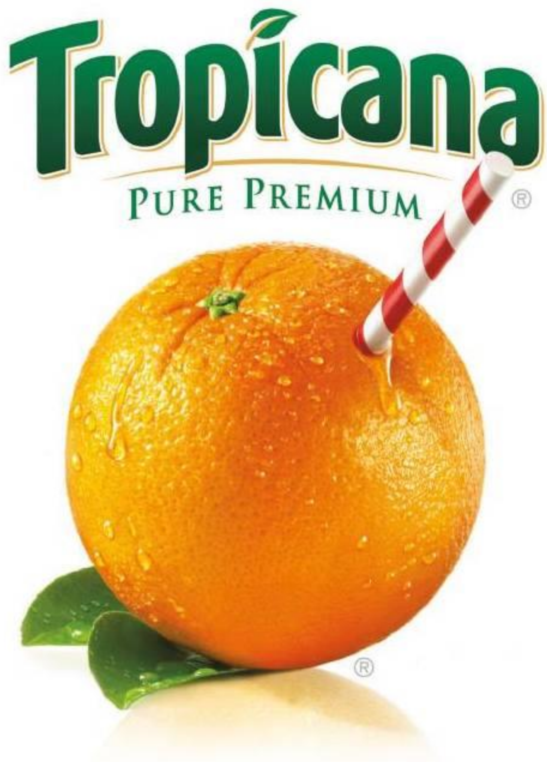
Pricing Decisions For Tropicana Orange Juice



By

Pranav Deo

Date: 01/09/2020

Executive Summary

As Nick's Grocer grows, the founder Nick Corpa would like to unify the retailer's pricing image. The objective of this analysis is to understand the factors affecting the quantity of orange juice sold across all 15 locations of Nick's Grocer and generate a recommended price across all locations. Our team at AMA Analytics accomplished this by running a semi-log model with the presented data. Based on our semi-log model and profit calculations, it is recommended for Nick's Grocer to price its 64 oz. Tropicana product at **$3.19 while running a promotional deal (in-store display and/or feature advertising)** for week one of January 2011. By doing so, the retailer's profits are projected to increase by **44.36%** relative to sales in the last week of December 2010.

Introduction & Background

Founded in 2016, Nick's Grocer is a fast-growing regional grocery retailer in the Chicago suburbs. Historically, the retailer made pricing recommendations on the chain level but defaulted to store managers for the final pricing decision. As the brand continues to make efforts to expand its business, the founder Nick Corpa has decided to pursue standardized pricing through all Nick's Grocer locations to unify the retailer's price image.

Our team at AMA Analytics was brought in to assist in the pricing of the retailer's products in selected categories. Our responsibility is to analyze the relationship between sales and variables such as price, psychological pricing, and seasonal trends and provide an optimal retail rate for a particular item within the refrigerated orange juice category: Tropicana (64 oz.). With our help, Nick's Grocer will be able to choose the most appropriate price point for its best-selling orange juice that will result in maximizing profit per store for the week.

Data & Methodology

Nick's Grocer provided AMA Analytics a dataset containing scanner data summarizing sales quantity per week per store, the presence of special promotions, and the price per week during two years between January 2009 to December 2010. In total, the data contained 1560 observations across all 15 Nick's Grocer stores. It should be noted that the sales quantity was noted in ounces to standardize units across all packaging sizes.

As stated before, our goal is to make an optimal price recommendation for the 64 oz. Tropicana product across all 15 Nick's Grocer stores. With this in mind, our team at AMA Analytics decided to pursue a regression model to analyze the effect of price on quantity. We explained our dataset using SAS software. First, the team ran a means analysis to check for any data-entry errors and confirmed there were no missing values in our data. Following that, our team took a look at the given variables in the dataset - store ID code, week, quantity, price, deal (dummy variable). Given the retail nature of the industry, our team decided to add three types of dummy variables to our model:

1. *Store* - This dummy variable type (store ID 2=store1, store ID 14=store2...etc.) is intended to capture the difference in quantity sold between each store during these two years. In our model, store 15 (store ID 137) was used as the baseline, with dummy variables store1, store2, store3, etc. measuring the difference in sales vs. baseline.
2. *Seasonal effects* - This dummy variable type (Q1, Q2, Q3) is intended to capture seasonal differences in orange juice bought over each quarter. In our model, Q4 was used as the baseline quarter, with dummy variables Q1, Q2, and Q3 measuring the difference of sales vs. Q4.
3. *Psychological pricing* - This dummy variable (end9) is intended to capture the effect of psychological pricing (prices that end in 9) on quantity bought per week. Psychological pricing is often used as a strategy to lead customers to believe they're getting a lower price.

After considering these dummy variables, our team ran several regression models in SAS to determine the best-fitting model, considering the model's R_2 value, adjusted R_2 value, scatter plot of Residual Values against the Predicted Value, Q-Q plot normality, and Residuals histogram normality. Using the given variables and our additional variable types, three types of regression models were run through SAS: linear, log-log, and semi-log.

The linear regression model yielded the lowest R_2 of 0.206 and adjusted R_2 of 0.195 between the three model types. The log-log regression model showed significant improvements in terms of goodness-of-fit relative to the linear model, yielding an R_2 value of 0.468 and an adjusted R_2 value of 0.461. The model with the best fit for our data, however, was the semi-log regression model, yielding the highest R_2 of 0.481 and the highest adjusted R_2 of 0.474, edging out the log-log model. The fit diagnostic charts for the semi-log model also indicated the semi-log model to be the most favorable. The p-value of variables range from <.0001 to 0.03 indicate the effects of all variables to be statistically significant. The price elasticity of semi-log models is b*Pt, which is in accordance with the features of orange juice (a relatively price sensitive product). Considering all the above factors, our team decided to proceed with the semi-log model.

Key Findings

Pattern Description

Table 1[1] shows some fundamental statistics of the dataset we are provided. Sales quantity of Tropicana varies from 33 bottles (2112 oz) to 8539 bottles (1186496 oz), with a standard deviation of 769 bottles (49216.14 oz). Prices of orange juice sales per bottle range from 2.74 to 4.18, with a standard deviation of 0.35 and mean of 3.60. 69.55% of products in these 15 stores were promoted and the prices of 24.55% of products are ended with 9 cents.

From two years of sales data, it is easy to see that even with higher prices, less deal and End9 price, the sales quantity of the second year is much higher than that of the first year, which is probably a result of organic growth. Quarter1's average volume ranks highest in four seasons, which is likely to be a result of lower prices, more promotions, and psychological price. See seasonal fluctuation in **Chart 1**[1] and **Table 2**[2].

As it is shown in **Table 3**[3], sales quantities of six stores exceed the average line. Store15 shows an excellent performance in volume of sales, while the sales quantity of store11 varies widely with different products. Since consumers show different reactions to similar prices, we decided to explore the role of different stores.

Selected Model

Log(quantity) = A + B_0price + B_1deal +B_2week + B_3end9 + B_4qrt1 + B_5qrt2 + B_6qrt3 + B_7store1+ B_8store2 + B_9store3 + B_{10}store4 + B_{11}store5 + B_{12}store6 + B_{13}store7 + B_{14}store8 + B_{15}store9 + B_{16}store10 + B_{17}store11 + B_{18}store12 + B_{19}store13 + B_{20}store14 + ε

Factors Found to Influence Sales Quantity

The regression results tell us that in-store display or feature advertising, psychological pricing, organic growth, seasonal effects and operation of different stores have significant impacts on the number of sales. In this case, the company can take advantage of these factors to formulate the marketing and pricing strategy. The impact of these factors will be explained in detail as follows. **Table 4**[4] lists the coefficients and the p-values for the variables we used.

- *Intercept* - The parameter estimate of the intercept is 15.92147, and the p-value of the intercept is less than 0.0001, which means intercept significantly affect sales. Intercept tells that when the price is zero, if not affected by holiday, repurchasing rate or economy prosperity, the weekly sales of store137 in season four would be 114,248 bottles (7,311,873 oz). Even though it doesn't seem to make sense in reality, it hints that the variables that we are considering now are not the only factors that would affect the quantity of sales.
- *Price* - The parameter estimate of price is -1.53699, indicating that sales quantity will be divided by $e_{1.53699}$ in response to one unit change of price. Since the p-value is less than 0.0001, we accepted that price is significant. This coefficient reflects that sales quantity of orange juice goes down as the price goes up.
- *In-store display or feature advertising (Deal)* - The parameter estimate for the dummy variable Deal is 0.081 and the p-value is 0.035, which is less than the significance level of 0.05, indicating that the variable of Deal is statistically significant and has an impact on the sales. In this case, theoretically, if we promote the product with the in-store display or feature advertising, the estimated sales volume will increase and be multiplied by $e_{0.081}$=1.084.
- *Psychological pricing (End9)* - Setting End9 as a dummy variable based on psychological pricing, we intended to observe how prices ending with 9 affect the quantity of sales. The parameter estimate and the p-value for the dummy variable of End9 are 0.187 and <0.0001, which met our expectation that psychological pricing has a significant impact on the quantity of sales. Hence, pricing the product with "odd prices" would likely increase the quantity of sales by a multiplier of $e_{0.187}$= 1.206.
- *Organic growth (Week)* - The p-value and the parameter estimate of the week variable, <0.0001 and 0.00347 respectively, show that the week variable has a significant effect on sales and as the operating time of the stores increases, sales quantity shows an upward trend. This may be because customers become more familiar with Nick's Grocer and trust the products it provides. Store prosperity and the word of mouth effect of consumers may also be reasons for positive effects on sales quantity.
- *Seasonal effects (Qrt1-Qrt3)* - By selecting Q4 as the baseline, we created three dummy variables to explore whether varying quarters is influences sales quantity. The changes among quarters reflect seasonal fluctuation, and the parameter estimate of Q1 (0.21152) is the largest, which means it has the largest sales quantity among these four quarters. Because originally consumers may have different demands according to different seasons. What's more, according to seasonal analysis (refer to **Table 2**[2]), it may be the result of lower prices, more promotion and psychological price. Therefore, demand in season one is much higher, and consumers are more probably price sensitive in season one.
- *Operation of different stores (Store)* - By selecting store 137 as the baseline, we created fourteen dummy variables to explore if the quantity of sales changes with each store. Since all parameter estimates are negative (range from -0.91075 to -0.32729), and all the p-values are less than 0.0001, store sales differences are statistically significant, meaning each store has varying abilities to attract consumers and generate sales quantities. We speculate that some stores have better (more popular) locations, and others may enjoy a good reputation or have higher repurchase rates.

The semi-log regression model provided the highest goodness-of-fit statistics for the dataset. The optimal price formula for semi-log is defined as P = C-(1/b), where C=unit variable cost and b is the parameter estimate for the price variable. According to our model and the $2.57 reference wholesale price, the initial optimal price is calculated to be $3.22. However, our team determined the significant effect of the end9 variable, so profit calculations were performed (Table 5, Table 6) for test optimal prices of $3.19 and $3.29. Ultimately, the price of $3.19 accompanied by a deal for week 105 was projected to be the most profitable.

Conclusion & Recommendations

Based on our team's semi-log analysis, we were able to determine the relationship of price, week, deal, store ID, seasonality, and psychological pricing variables with quantity of orange juice sold. It was demonstrated that as the price variable increased, quantity decreased. The data showed that in-store displays and/or feature advertising and psychological pricing were effective in increasing sales quantity. The week variable was shown to also have a positive impact on sales, which is in line with our expectations that organic growth is a factor in the increasing sales (as Nick Grocer's retail locations are open longer, more people will become aware of the brand through word of mouth, potentially increasing store visits and orange juice sales). It is evident that orange juice quantity is also influenced by seasonality (with sales varying between quarters), and stores.

The recommended price point for Nick's Grocer stores for week one of January 2011 (week 105) is $3.19 while running a promotional deal, generating a 44.36% increase to sales compared to sales in week 104 across all stores. Additional recommendations can be found below:

- Using unique product IDs to identify orange juice products of different sizes and breaking down the data by said product IDs can help analyze variable effects on quantity per product size. By analyzing data per product size, specific recommendations may be made for particular sized Tropicana bottles, which will influence pricing.
- Currently, seasons three and four have lower rates of deals and psychological pricing (**Table 2**[2]). Sales for these two quarters may be increased through enacting deals and psychological pricing rates.
- As store137 (store15) has the best performance in sales among others, Nick's Grocer should consider analyzing the operation strategy that store137 conducted and encourage other stores to emulate and apply similar effective tactics.

Limitations & Opportunities for Future Research

This analysis is first and foremost limited by the lack of competitor data in our explanatory variables (i.e., the effect of competitor promotions may influence quantity sold by Nick's Grocer locations). Data collection on store membership systems and coupons could also be helpful for future research in identifying the behavior of individual customers and the effect of those variables on sales. Additionally, it should be noted that the prices given in the data set reflect temporary price reductions - there is an opportunity for data on temporary discounts to be collected separately to aid a more nuanced analysis in the future. While our model accounted for sales differences between individual Nick's Grocer stores, the lack of data regarding each store's regions (downtown vs. rural, north/south/east/west areas) prevented further interpretations of why certain stores appeared more successful than others.

Furthermore, our team considered testing the effects of holiday weeks on the quantity of orange juice purchased, as intuition told us that orange juice purchase quantity might increase due to celebratory events. However, upon running our semi-log model with the holiday weeks as a variable, we found that the holiday variable was not statistically significant. This may be explained by consumers pre-buying orange juice before the holidays as opposed to during the holiday weeks. Additionally, because we were only given two years' worth of data, there may not have been enough data points of holiday weeks. Considering all these factors, our team decided to exclude the holiday week variable from our model.

Our team also attempted to test potential interaction effects between the psychological pricing, holiday weeks, and deals on quantity sold through interaction dummy variables end9*holiday, end9*deal, and deal*holiday. Although it was found that the parameter estimate for deal*holiday was statistically significant ($p<0.0001$), we did not include the interaction effect in our final analysis due to the lack of observation counts occurring during a holiday week when a deal was present. A more extensive data set may be necessary to test the interaction effect between the above dummy variables.

Appendices

Table 1: Descriptive Statistics for Quantity, Price, Deal, and End9

Variable	Mean	Standard Deviation	Minimum	Maximum
Quant	40914.54	49216.14	2112.00	1186496.00
Price	3.60	0.35	2.74	4.18
Deal	0.70	0.46	0.00	1.00
End9	0.25	0.43	0.00	1.00

Table 2: Performance in Each Season

Quarter	avgQuant	avgPrice	pctDeal	pctEnd9	SDQuant	SDPrice
1	55112.80	3.50	0.74	0.34	81337.01	0.40
2	34775.21	3.70	0.71	0.29	29775.64	0.38
3	37289.96	3.63	0.72	0.22	28549.99	0.3
4	36480.19	3.58	0.61	0.13	33427.74	0.24

Chart 1: Seasonal Fluctuation

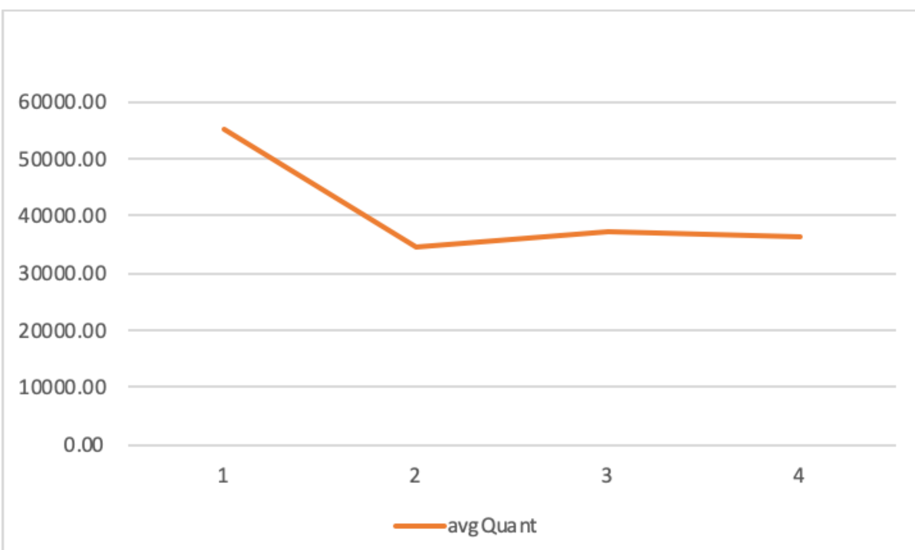


Table 3: Performance of Different Stores

store	avgQuant	avgPrice	pctDeal	pctEnd9	SDPrice	SDQuant
1	27818.79	3.58	0.67	0.26	0.34	21659.02
2	31783.38	3.65	0.72	0.28	0.37	22349.88
3	46072.62	3.61	0.72	0.27	0.35	37474.19
4	38584.09	3.63	0.71	0.28	0.35	28021.35
5	33152.66	3.61	0.71	0.27	0.35	19671.72
6	34583.60	3.62	0.70	0.27	0.35	29465.33
7	54601.24	3.56	0.65	0.26	0.36	125736.49
8	37563.69	3.61	0.72	0.27	0.35	32008.22
9	35760.27	3.66	0.71	0.26	0.36	37676.50
10	29345.85	3.62	0.72	0.28	0.35	31458.89
11	51078.77	3.57	0.68	0.13	0.35	63980.04
12	46139.08	3.59	0.68	0.20	0.34	45463.95
13	35575.38	3.61	0.68	0.22	0.35	33062.36
14	46978.69	3.54	0.66	0.20	0.35	49490.24
15	64680.00	3.56	0.67	0.24	0.35	42209.46

Table 4: Coefficient & P-value for variables in model

Variable	Parameter estimate	Pr > |t|
intercept	15.921	<.0001
price	-1.537	<.0001
deal	0.081	0.0315
week	0.003	<.0001
end9	0.187	<.0001
qrt1	0.212	<.0001
qrt2	0.176	0.0002
qrt3	0.146	0.0013
store1	-0.842	<.0001
store2	-0.608	<.0001
store3	-0.326	0.0002
store4	-0.439	<.0001
store5	-0.545	<.0001
store6	-0.636	<.0001
store7	-0.682	<.0001
store8	-0.556	<.0001
store9	-0.545	<.0001
store10	-0.909	<.0001
store11	-0.612	<.0001
store12	-0.48	<.0001
store13	-0.667	<.0001
store14	-0.628	<.0001

Table 5: Expected Sales for Week 105 without Deal

	Expected Sales (oz) in Week 105			Expected Gross Profit in Week 105		
Store	P = 3.22 w/ Deal	P = 3.19 w/ Deal	P = 3.29 w/ Deal	P = 3.22 w/ Deal2	P = 3.19 w/ Deal4	P = 3.29 w/ Deal6
Store 1 (code=2)	48402.280	61136.962	52426.785	491.586	592.264	589.801
Store 2 (code=14)	61125.976	77208.274	66208.418	620.811	747.955	744.845
Store 3 (code=32)	81093.833	102429.691	87836.543	823.609	992.288	988.161
Store 4 (code=52)	72373.262	91414.730	78390.882	735.041	885.580	881.897
Store 5 (code=62)	65092.994	82219.017	70505.281	661.101	796.497	793.184
Store 6 (code=68)	59457.806	75101.206	64401.545	603.868	727.543	724.517
Store 7 (code=71)	56809.691	71756.369	61533.246	576.973	695.140	692.249
Store 8 (code=72)	64432.420	81384.645	69789.783	654.392	788.414	785.135
Store 9 (code=93)	65120.339	82253.556	70534.900	661.378	796.831	793.518
Store 10 (code=95)	45255.170	57161.844	49018.002	459.623	553.755	551.453
Store 11 (code=111)	60880.743	76898.519	65942.795	618.320	744.954	741.856
Store 12 (code=123)	69489.587	87772.358	75267.439	705.754	850.295	846.759
Store 13 (code=124)	57641.738	72807.329	62434.476	585.424	705.321	702.388
Store 14 (code=130)	59929.982	75697.613	64912.981	608.664	733.321	730.271
Store 15 (code=137)	112304.836	141852.335	121642.647	1140.596	1374.194	1368.480
Total	**979410.7**	**1237094.4**	**1060845.7**	**9947.1**	**11984.4**	**11934.5**
Increase over current practice				19.82%	44.36%	43.76%

Table 6: Expected Sales for Week 105 without Deal

	Expected Sales (oz) in Week 105			Expected Gross Profit in Week 105		
Store	P = 3.22 No Deal	P = 3.19 No Deal	P = 3.29 No Deal	P = 3.22 No Deal	P = 3.19 No Deal	P = 3.29 No Deal
Store 1 (code=2)	44647.884	56394.781	48360.222	453.455	546.324	544.053
Store 2 (code=14)	56384.648	71219.497	61072.863	572.657	689.939	687.070
Store 3 (code=32)	74803.668	94484.578	81023.369	759.725	915.319	911.513
Store 4 (code=52)	66759.521	84324.008	72310.375	678.026	816.889	813.492
Store 5 (code=62)	60043.958	75841.575	65036.433	609.821	734.715	731.660
Store 6 (code=68)	54845.872	69275.868	59406.142	557.028	671.110	668.319
Store 7 (code=71)	52403.161	66190.478	56760.327	532.220	641.220	638.554
Store 8 (code=72)	59434.622	75071.923	64376.433	603.633	727.259	724.235
Store 9 (code=93)	60069.181	75873.435	65063.754	610.078	735.024	731.967
Store 10 (code=95)	41744.885	52728.000	45215.847	423.971	510.802	508.678
Store 11 (code=111)	56158.437	70933.770	60827.843	570.359	687.171	684.313
Store 12 (code=123)	64099.523	80964.163	69429.206	651.011	784.340	781.079
Store 13 (code=124)	53170.670	67159.919	57591.652	540.015	650.612	647.906
Store 14 (code=130)	55281.423	69826.013	59877.908	561.452	676.440	673.626
Store 15 (code=137)	103593.742	130849.345	112207.251	1052.124	1267.603	1262.332
Total	**903441.2**	**1141137.4**	**978559.6.2**	**9175.6**	**11054.8.8**	**11008.8.5**
Increase over current practice				10.52%	33.16%	32.61%