SCM 651 Fall 2018 Group Project

**Orange Juice Data Market Analysis**

**Team 8**

**- Mohammed Alnahdi**

**- Rajat Varshney**

**- Harsh Darji**

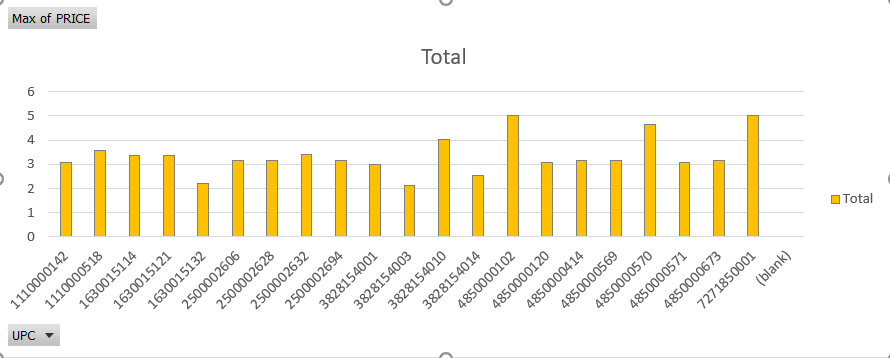
**- Lu Yin**

**Introduction**

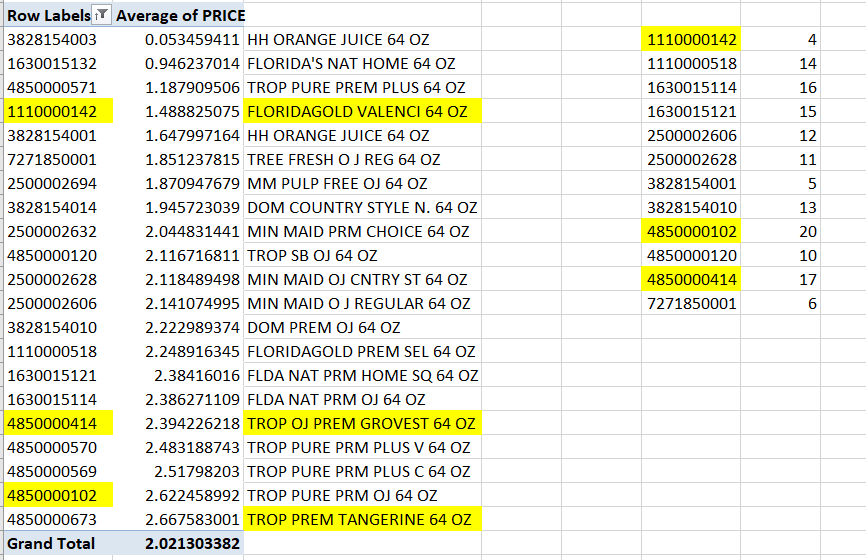
In this project we have selected orange juice data across 3 different brands, FG, Tropicana Pure and Tropicana Grove to study and analyze the affect of different variables on sales and profitability. Using different analytical tools to organize and sort the data in order to find an association between the variation of the variable and how the market place react accordingly. Two products we chose have higher prices, and one is lower priced. Then, using access to combine the data of all the products of our study in an excel file for further analysis. Implying visual tools in excel check for the patterns and changes in the data. Due to the limitations of excel package to run a detailed analysis another statistical package had to be used presented in R language. Different models have been created to study the impact of different changes by using regression analysis to generate a logical understanding of changes and how it can impact and enhance the decision-making process in the short and the long term to increase sales and profit margins for the intended products. This report includes potential questions and assumptions and how they impact outcomes. At the end a summary of all the findings will be presented for recommendations.

**Selection of Data**

Our goal is to include three UPC’s in our analysis with two higher priced national brands and one lower priced or store brand.

  
**Figure 1: Rage of UPC’s.**

We first calculated the average price of each UPC, showing in Figure 1. Then we looked up those UPCs in the high movement report. Our team have selected UPC 4850000102 (TROP Pure) and 1630015114 (Trop Grove) for products with higher average prices and 1110000142 (FG) for product with lower prices. All three items have high movements.

  
**Figure 2: Different brands with Row labels and Average Price.**

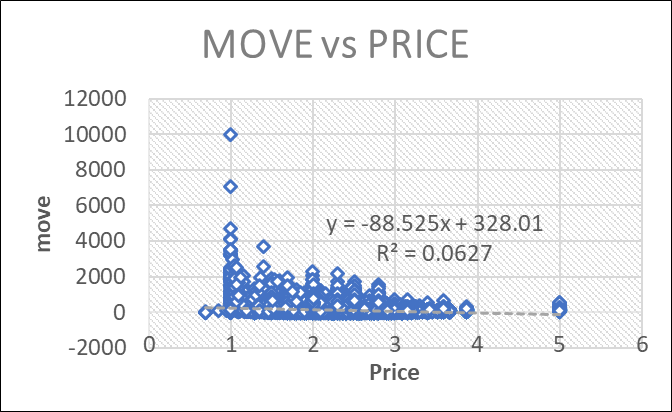
**Business Questions:**

The following are the list of questions that this project is going to cover in order to reach out for a conclusion after running the associated models:

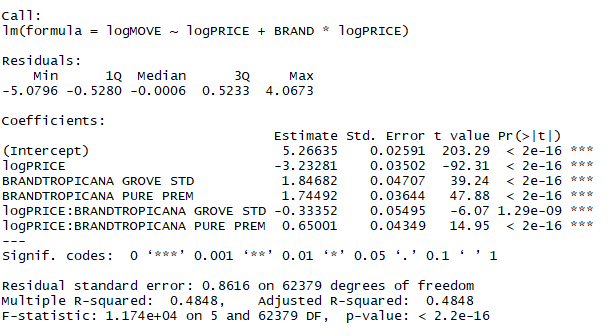
* How does the demand for a brand depend on price? What is the price elasticity of demand of a brand? Is price elasticity different for different brands?
* How does demand depend on whether the product is on sale (Feat =1)? Is this dependence same for all brands?
* How does the demand for a brand depend on the price of another brand?
* What Demographic factors affect demand?
* How does price vary across brands?
* How does the proportion of times a brand is on sale vary across brands?

**Descriptive Analysis & Modeling**

1. Price Elasticity

****  
**Figure 3: Move vs Price (Price elasticity)**

The scatter plot in figure 3 shows a high sensitivity to the price as the price increases the demand on the other hand decreases. Understanding how sensitive customers are to the changes makes the process of price setting quite challenging as any little change might impact the sales volume and as a result product profitability. However, R2 of this model is only 0.0627, meaning only 6% of the variations can be explained by this model. Our team explored more models in R.

**  
Figure 4: Price elasticity from the linear model**

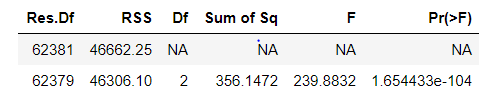
After running a linear model using R Studio, the price elasticity of demand of a brand was found to be:

TROPICANA GROVE STD: -3.56

TROPICANA PURE PREM: -2.58

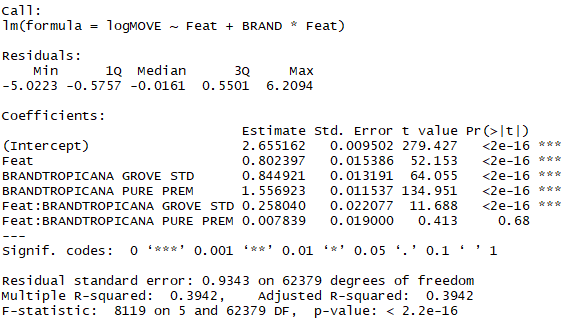
FG: -3.23

The coefficients for all 3 products are negative, meaning when price increases, the demand will decrease. By using logmove as dependent variable and logprice as one of the independent variables, the adjusted R2 for this new linear model is 0.4848, meaning 48% of the variations can be explained by logprice and brand.

  
**Figure 5: Sources of errors**

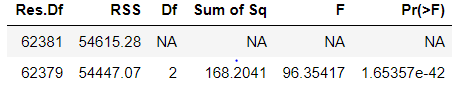
To see if price elasticity different for different brands, our team did a Hypothesis test showing in Figure 5. Since the P value is smaller than 0.05, we conclude that at 99% level of confidence, the price elasticity is different between those 3 bands we selected.

1. Feat vs. Demand

  
**Figure 6: Promotion impact on Demand**

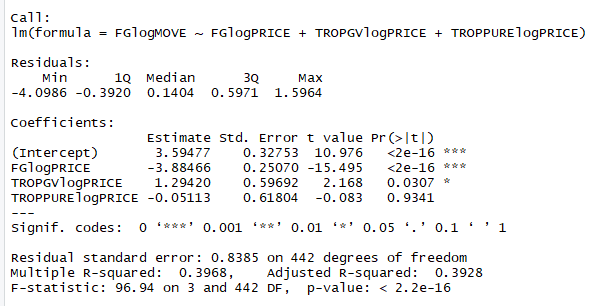
From the R result, all coefficients are positive, meaning having promotion has help to increase the sales for all three brands. Especially, Tropicana Grove will have the most significant impact from the promotion, since this band has highest coefficient of Feat, 1.06. When Tropicana Grove is on sale, the moves will increase 1.06%.

Figure 7 is the result from Hypothesis testing. Again, the P value is small enough for our team to conclude that the dependence between demand and product promotion are not the same for all 3 brands.

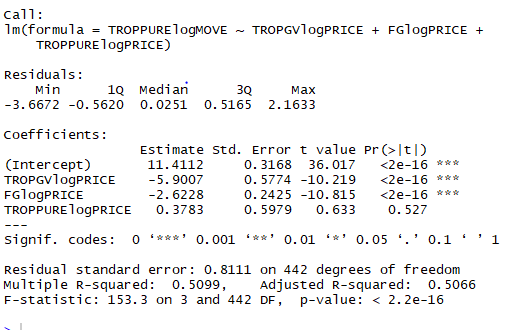
  
**Figure 7: Sources of error**

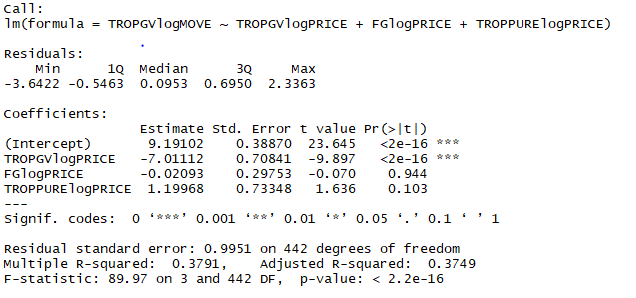
1. Demand vs. Price of Another Brand

From the model in figure 8, we can conclude that the demand for FG is significantly depended on price of both FG and TROPGV. Our model shows that as price of FG increases demand tends to decrease and as price of TROPGV increases demand for FG begins to increase.

  
**Figure 8: FG demand vs changes in other brands prices.**

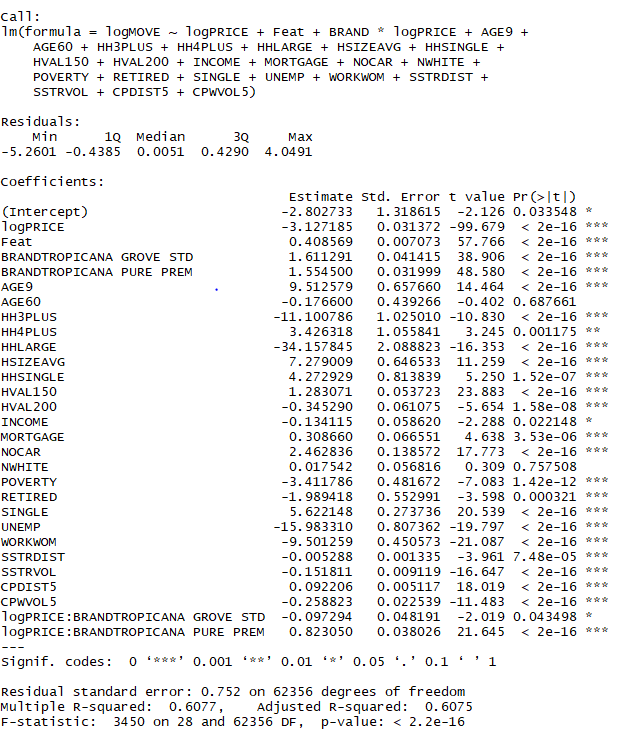
From the model in figure 9, we can conclude that the demand for TROPPURE is significantly depended on price of both FG and TROPGV. Our model shows that as price of FG increases demand tends to decrease and as price of TROPGV increases demand for FG begins to decrease.

  
**Figure 9: TROPPURE demand model**

**  
Figure 10: TROPGV demand model.**

From the model above, we can conclude that the demand for TROPGV is significantly depended on price of itself, other brands do not affect its demand, because the P value for both FG and TROPOURE are greater than 0.05.

1. Demand vs. Demographic Variables

**  
Figure 11: The effect of demographics on the demand.**

Our team included all the demographic variables in the new model and the result from R is showing in Figure 11. At 99% level of confidence we conclude that the following variables have significant impact on logmove, since their P value is smaller than 0.05: AGE9,HH3PLUS,HHLARGE,HSIZEAVG,HHSINGLE,HVAL150,HVAL200,MORTGAGE,NOCAR,POVERTY,RETIRED, SINGLE,UNEMP,WORKWOM,SSTRDIST,SSTRVOL,CPDIST5,CPWVOL

1. Price vs. Brand

|  |  |  |  |
| --- | --- | --- | --- |
| **Row Labels** | **Average of logPRICE** | **Average of PRICE** | **StdDev of PRICE** |
| **FG VALENCIA** | **0.711201866** | **2.075668098** | **0.404986774** |
| **TROPICANA GROVE STD** | **0.913576135** | **2.523648221** | **0.365848245** |
| **TROPICANA PURE PREM** | **0.974220452** | **2.695135144** | **0.491665959** |
| **Grand Total** | **0.893993319** | **2.498983117** | **0.509072125** |

**Figure 12: Brands Prices**

The above pivot table in figure 12 shows how the price varies across brands. We used Access to sort the data by STOREWEEK, and then we calculated the average prices of FG, Tropicana Pure and Tropicana Grove in Excel.

The average price for FG Valencia is 2.0756 while for Tropicana Grove and Tropicana Pure average price is slightly higher.

1. Proportion Of Times A Brand Is On Sale

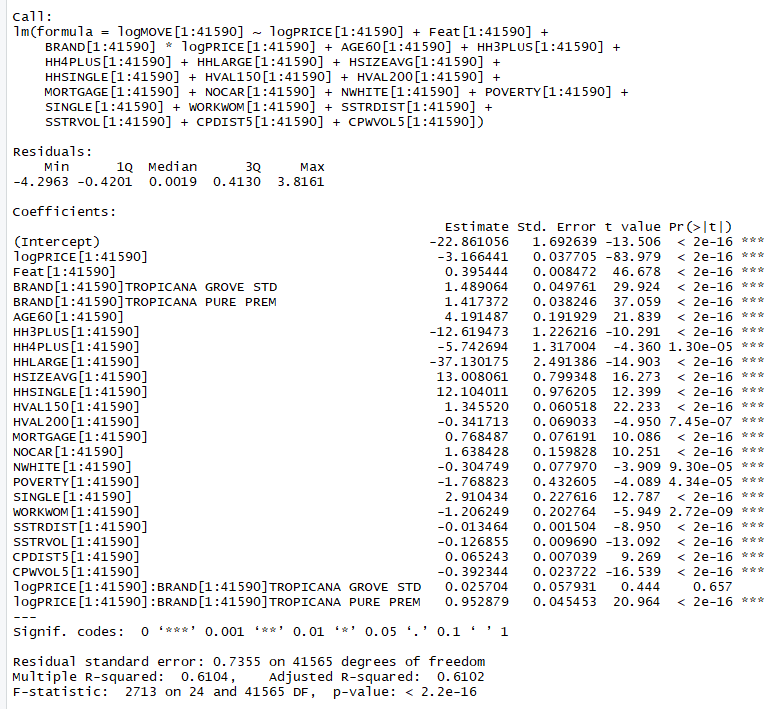
|  |  |  |  |
| --- | --- | --- | --- |
| **Sum of logMOVE** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand Total** |
| **FG VALENCIA** | **55.46%** | **44.54%** | **100.00%** |
| **TROPICANA GROVE STD** | **60.46%** | **39.54%** | **100.00%** |
| **TROPICANA PURE PREM** | **61.37%** | **38.63%** | **100.00%** |
| **Grand Total** | **60.06%** | **39.94%** | **100.00%** |

**Figure 13: Promotion impact on sales.**

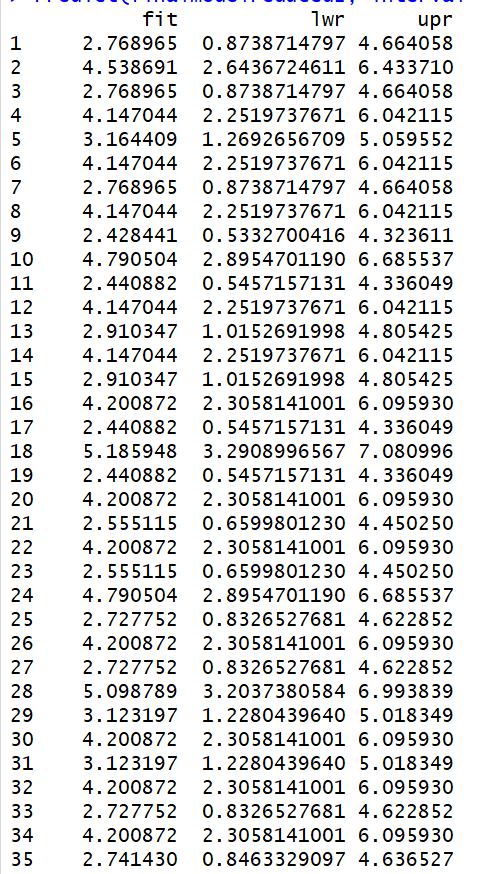
The pivot chart in figure 13 shows that: FG Valencia has almost equal proportion whether it’s on sale or not.(1:1ratio) While, for Tropicana grove std and Tropicana pure the ratio is (3:2).

1. Additional Analysis

Our team used 2/3 of dataset to create a linear regression model and removed all the nonsignificant independent variables. The result shows below:

  
**Figure 15: Verification model**

We then used the rest 1/3 of data points to validate the model. The result shows in Figure 16.



Our final model in figure 15 shows that R-squared value is 0.6102 which is higher than the full model. Also, P-value indicates that all the variables are significant in predicting the demand.

**Executive Summary:**

In this project we have selected orange juice dataset to study possible factors that might impact the sales of different brands of orange juice. The main purpose was to find out ways to analyze the behavior of different factors and they contribute in the volume of sales or the amount of demand in the marketplace, while considering the influence of price fluctuation at the same time. Three brands were included in the study two are highly priced, and one is low priced. These brands are TROP PREM, TROP OJ PREM, and FLORIDA GOLD.

In the modeling and descriptive analysis part three software packages were used Access to collect, sort, and filter data of interest in one file, then using Excel to run simple analysis and for further modeling and hypothesis testing R-Studio had been utilized. Based on the analysis all brands tend to be price sensitive as demand kept changing due to price variation. In addition to that several demographical factors were impacting the demand presented in the following factors:

AGE9,HH3PLUS,HHLARGE,HSIZEAVG,HHSINGLE,HVAL150,HVAL200,MORTGAGE,NOCAR,POVERTY,RETIRED,SINGLE,UNEMP,WORKWOM,SSTRDIST,SSTRVOL,CPDIST5,CPWVOL.

Also analysis showed that the average price for FLORIDAGOLD is $2.0756, while the prices for TROPICANA PURE and TROPICANA GROOVE is slightly higher. However, when it come to the proportion whether product is on sale or not it was (1:1) for FLORIDAGOLD and (3:2) for both TROPICANA GROOVE and PURE. It was also observed over a long period of time that demand tend to be stable across the three brands.