SCM 651 Fall 2018 Group Project

Orange Juice Data Market Analysis

- **By Yash Pasar**

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.

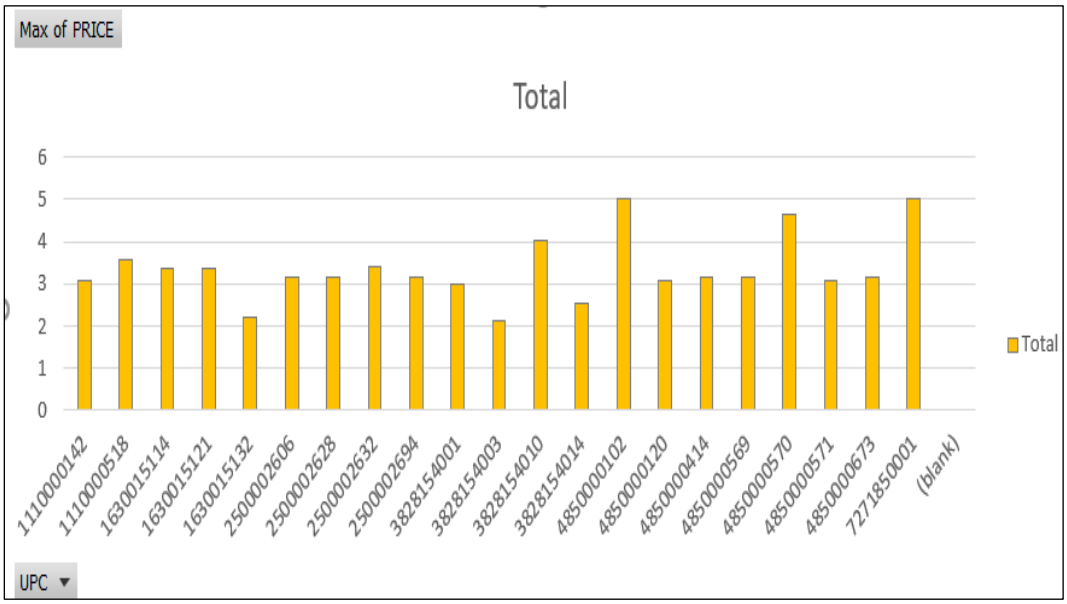

Figure 1: Rage of UPC's.

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.

Row Labels	Average of PRICE	
3828154003	0.053459411	HH ORANGE JUICE 64 OZ
1630015132	0.946237014	FLORIDA'S NAT HOME 64 OZ
4850000571	1.187909506	TROP PURE PREM PLUS 64 OZ
1110000142	1.488825075	FLORIDAGOLD VALENCI 64 OZ
3828154001	1.647997164	HH ORANGE JUICE 64 OZ
7271850001	1.851237815	TREE FRESH O J REG 64 OZ
2500002694	1.870947679	MM PULP FREE OJ 64 OZ
3828154014	1.945723039	DOM COUNTRY STYLE N. 64 OZ
2500002632	2.044831441	MIN MAID PRM CHOICE 64 OZ
4850000120	2.116716811	TROP SB OJ 64 OZ
2500002628	2.118489498	MIN MAID OJ CNTRY ST 64 OZ
2500002606	2.141074995	MIN MAID O J REGULAR 64 OZ
3828154010	2.222989374	DOM PREM OJ 64 OZ
1110000518	2.248916345	FLORIDAGOLD PREM SEL 64 OZ
1630015121	2.38416016	FLDA NAT PRM HOME SQ 64 OZ
1630015114	2.386271109	FLDA NAT PRM OJ 64 OZ
4850000414	2.394226218	TROP OJ PREM GROVEST 64 OZ
4850000570	2.483188743	TROP PURE PRM PLUS V 64 OZ
4850000569	2.51798203	TROP PURE PRM PLUS C 64 OZ
4850000102	2.622458992	TROP PURE PRM OJ 64 OZ
4850000673	2.667583001	TROP PREM TANGERINE 64 OZ
Grand Total	2.021303382	

1110000142	4
1110000518	14
1630015114	16
1630015121	15
2500002606	12
2500002628	11
3828154001	5
3828154010	13
4850000102	20
4850000120	10
4850000414	17
7271850001	6

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

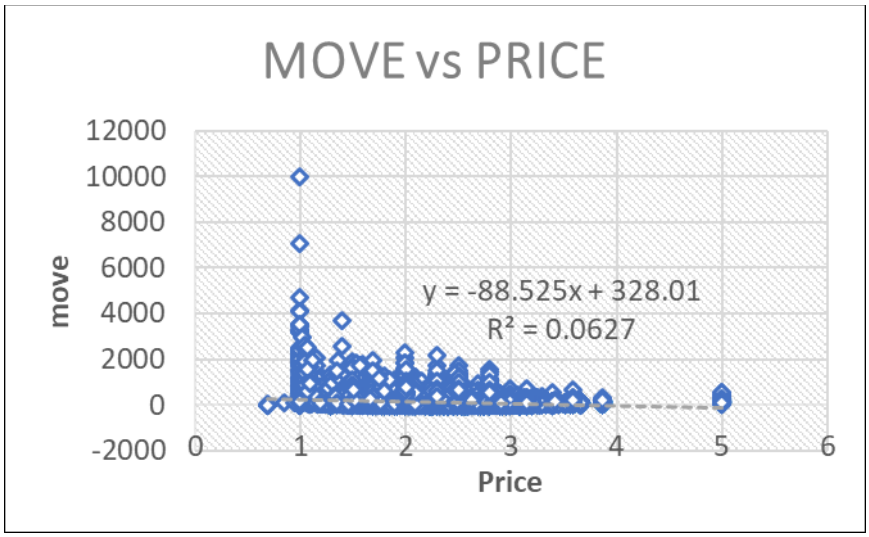

Figure 3: Move vs Price (Price elasticity)

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, R^2 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.

```
Call:
lm(formula = logMOVE ~ logPRICE + BRAND * logPRICE)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0796 -0.5280 -0.0006  0.5233  4.0673

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        5.26635    0.02591  203.29  < 2e-16 ***
logPRICE                          -3.23281    0.03502  -92.31  < 2e-16 ***
BRANDTROPICANA GROVE STD           1.84682    0.04707   39.24  < 2e-16 ***
BRANDTROPICANA PURE PREM           1.74492    0.03644   47.88  < 2e-16 ***
logPRICE:BRANDTROPICANA GROVE STD -0.33352    0.05495   -6.07 1.29e-09 ***
logPRICE:BRANDTROPICANA PURE PREM  0.65001    0.04349   14.95  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8616 on 62379 degrees of freedom
Multiple R-squared:  0.4848,    Adjusted R-squared:  0.4848
F-statistic: 1.174e+04 on 5 and 62379 DF,  p-value: < 2.2e-16
```

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 R^2 for this new linear model is 0.4848, meaning 48% of the variations can be explained by logprice and brand.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
62381	46662.25	NA	NA	NA	NA
62379	46306.10	2	356.1472	239.8832	1.654433e-104

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.

2. Feat vs. Demand

```
Call:
lm(formula = logMOVE ~ Feat + BRAND * Feat)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0223 -0.5757 -0.0161  0.5501  6.2094

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   2.655162   0.009502 279.427   <2e-16 ***
Feat                          0.802397   0.015386  52.153   <2e-16 ***
BRANDTROPICANA GROVE STD      0.844921   0.013191  64.055   <2e-16 ***
BRANDTROPICANA PURE PREM      1.556923   0.011537 134.951   <2e-16 ***
Feat:BRANDTROPICANA GROVE STD 0.258040   0.022077  11.688   <2e-16 ***
Feat:BRANDTROPICANA PURE PREM 0.007839   0.019000   0.413     0.68
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9343 on 62379 degrees of freedom
Multiple R-squared:  0.3942,    Adjusted R-squared:  0.3942
F-statistic:  8119 on 5 and 62379 DF,  p-value: < 2.2e-16
```

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.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
62381	54615.28	NA	NA	NA	NA
62379	54447.07	2	168.2041	96.35417	1.65357e-42

Figure 7: Sources of error

3. 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.

```
Call:
lm(formula = FGlogMOVE ~ FGlogPRICE + TROPGVlogPRICE + TROPPURElogPRICE)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0986 -0.3920  0.1404  0.5971  1.5964

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       3.59477    0.32753  10.976   <2e-16 ***
FGlogPRICE       -3.88466    0.25070 -15.495   <2e-16 ***
TROPGVlogPRICE    1.29420    0.59692   2.168   0.0307 *
TROPPURElogPRICE -0.05113    0.61804  -0.083   0.9341
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8385 on 442 degrees of freedom
Multiple R-squared:  0.3968,    Adjusted R-squared:  0.3928
F-statistic: 96.94 on 3 and 442 DF,  p-value: < 2.2e-16
```

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.

```
Call:
lm(formula = TROPPURElogMOVE ~ TROPGVlogPRICE + FGlogPRICE +
    TROPPURElogPRICE)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6672 -0.5620  0.0251  0.5165  2.1633

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       11.4112     0.3168  36.017   <2e-16 ***
TROPGVlogPRICE    -5.9007     0.5774 -10.219   <2e-16 ***
FGlogPRICE        -2.6228     0.2425 -10.815   <2e-16 ***
TROPPURElogPRICE   0.3783     0.5979   0.633    0.527
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8111 on 442 degrees of freedom
Multiple R-squared:  0.5099,    Adjusted R-squared:  0.5066
F-statistic: 153.3 on 3 and 442 DF,  p-value: < 2.2e-16
```

Figure 9: TROPPURE demand model

```
Call:
lm(formula = TROPGVlogMOVE ~ TROPGVlogPRICE + FGlogPRICE + TROPPURElogPRICE)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6422 -0.5463  0.0953  0.6950  2.3363

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.19102    0.38870  23.645   <2e-16 ***
TROPGVlogPRICE  -7.01112    0.70841  -9.897   <2e-16 ***
FGlogPRICE      -0.02093    0.29753  -0.070    0.944
TROPPURElogPRICE 1.19968    0.73348   1.636    0.103
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9951 on 442 degrees of freedom
Multiple R-squared:  0.3791,    Adjusted R-squared:  0.3749
F-statistic: 89.97 on 3 and 442 DF,  p-value: < 2.2e-16
```

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.

4. Demand vs. Demographic Variables

```
Call:
lm(formula = logMOVE ~ logPRICE + Feat + BRAND * logPRICE + AGE9 +
    AGE60 + HH3PLUS + HH4PLUS + HHLARGE + HSIZEAVG + HHSINGLE +
    HVAL150 + HVAL200 + INCOME + MORTGAGE + NOCAR + NWHITE +
    POVERTY + RETIRED + SINGLE + UNEMP + WORKWOM + SSTRDIST +
    SSTRVOL + CPDIST5 + CPWVOL5)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2601 -0.4385  0.0051  0.4290  4.0491

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -2.802733   1.318615  -2.126 0.033548 *
logPRICE                          -3.127185   0.031372 -99.679  < 2e-16 ***
Feat                               0.408569   0.007073  57.766  < 2e-16 ***
BRANDTROPICANA GROVE STD           1.611291   0.041415  38.906  < 2e-16 ***
BRANDTROPICANA PURE PREM           1.554500   0.031999  48.580  < 2e-16 ***
AGE9                               9.512579   0.657660  14.464  < 2e-16 ***
AGE60                             -0.176600   0.439266  -0.402 0.687661
HH3PLUS                          -11.100786   1.025010 -10.830  < 2e-16 ***
HH4PLUS                            3.426318   1.055841   3.245 0.001175 **
HHLARGE                          -34.157845   2.088823 -16.353  < 2e-16 ***
HSIZEAVG                           7.279009   0.646533  11.259  < 2e-16 ***
HHSINGLE                           4.272929   0.813839   5.250 1.52e-07 ***
HVAL150                            1.283071   0.053723  23.883  < 2e-16 ***
HVAL200                           -0.345290   0.061075  -5.654 1.58e-08 ***
INCOME                            -0.134115   0.058620  -2.288 0.022148 *
MORTGAGE                           0.308660   0.066551   4.638 3.53e-06 ***
NOCAR                              2.462836   0.138572  17.773  < 2e-16 ***
NWHITE                             0.017542   0.056816   0.309 0.757508
POVERTY                           -3.411786   0.481672  -7.083 1.42e-12 ***
RETIRED                           -1.989418   0.552991  -3.598 0.000321 ***
SINGLE                             5.622148   0.273736  20.539  < 2e-16 ***
UNEMP                            -15.983310   0.807362 -19.797  < 2e-16 ***
WORKWOM                           -9.501259   0.450573 -21.087  < 2e-16 ***
SSTRDIST                          -0.005288   0.001335  -3.961 7.48e-05 ***
SSTRVOL                           -0.151811   0.009119 -16.647  < 2e-16 ***
CPDIST5                            0.092206   0.005117  18.019  < 2e-16 ***
CPWVOL5                           -0.258823   0.022539 -11.483  < 2e-16 ***
logPRICE:BRANDTROPICANA GROVE STD -0.097294   0.048191  -2.019 0.043498 *
logPRICE:BRANDTROPICANA PURE PREM  0.823050   0.038026  21.645  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.752 on 62356 degrees of freedom
Multiple R-squared:  0.6077,    Adjusted R-squared:  0.6075
F-statistic:  3450 on 28 and 62356 DF,  p-value: < 2.2e-16
```

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

5. 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.

6. 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).

7. 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:

```
Call:
lm(formula = logMOVE[1:41590] ~ logPRICE[1:41590] + Feat[1:41590] +
    BRAND[1:41590] * logPRICE[1:41590] + AGE60[1:41590] + HH3PLUS[1:41590] +
    HH4PLUS[1:41590] + HHLARGE[1:41590] + HSIZEAVG[1:41590] +
    HHSINGLE[1:41590] + HVAL150[1:41590] + HVAL200[1:41590] +
    MORTGAGE[1:41590] + NOCAR[1:41590] + NWHITE[1:41590] + POVERTY[1:41590] +
    SINGLE[1:41590] + WORKWOM[1:41590] + SSTRDIST[1:41590] +
    SSTRVOL[1:41590] + CPDIST5[1:41590] + CPWVOL5[1:41590])

Residuals:
    Min      1Q  Median      3Q     Max
-4.2963 -0.4201  0.0019  0.4130  3.8161

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       -22.861056   1.692639 -13.506  < 2e-16 ***
logPRICE[1:41590]                                  -3.166441   0.037705 -83.979  < 2e-16 ***
Feat[1:41590]                                       0.395444   0.008472  46.678  < 2e-16 ***
BRAND[1:41590]TROPICANA GROVE STD                   1.489064   0.049761  29.924  < 2e-16 ***
BRAND[1:41590]TROPICANA PURE PREM                   1.417372   0.038246  37.059  < 2e-16 ***
AGE60[1:41590]                                      4.191487   0.191929  21.839  < 2e-16 ***
HH3PLUS[1:41590]                                  -12.619473   1.226216 -10.291  < 2e-16 ***
HH4PLUS[1:41590]                                   -5.742694   1.317004  -4.360 1.30e-05 ***
HHLARGE[1:41590]                                  -37.130175   2.491386 -14.903  < 2e-16 ***
HSIZEAVG[1:41590]                                  13.008061   0.799348  16.273  < 2e-16 ***
HHSINGLE[1:41590]                                  12.104011   0.976205  12.399  < 2e-16 ***
HVAL150[1:41590]                                    1.345520   0.060518  22.233  < 2e-16 ***
HVAL200[1:41590]                                   -0.341713   0.069033  -4.950 7.45e-07 ***
MORTGAGE[1:41590]                                   0.768487   0.076191  10.086  < 2e-16 ***
NOCAR[1:41590]                                      1.638428   0.159828  10.251  < 2e-16 ***
NWHITE[1:41590]                                    -0.304749   0.077970  -3.909 9.30e-05 ***
POVERTY[1:41590]                                   -1.768823   0.432605  -4.089 4.34e-05 ***
SINGLE[1:41590]                                     2.910434   0.227616  12.787  < 2e-16 ***
WORKWOM[1:41590]                                   -1.206249   0.202764  -5.949 2.72e-09 ***
SSTRDIST[1:41590]                                  -0.013464   0.001504  -8.950  < 2e-16 ***
SSTRVOL[1:41590]                                   -0.126855   0.009690 -13.092  < 2e-16 ***
CPDIST5[1:41590]                                    0.065243   0.007039   9.269  < 2e-16 ***
CPWVOL5[1:41590]                                   -0.392344   0.023722 -16.539  < 2e-16 ***
logPRICE[1:41590]:BRAND[1:41590]TROPICANA GROVE STD  0.025704   0.057931   0.444    0.657
logPRICE[1:41590]:BRAND[1:41590]TROPICANA PURE PREM  0.952879   0.045453  20.964  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7355 on 41565 degrees of freedom
Multiple R-squared:  0.6104,    Adjusted R-squared:  0.6102
F-statistic:  2713 on 24 and 41565 DF,  p-value: < 2.2e-16
```

Figure 15: Verification model

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

```
          fit          lwr      upr
1    2.768965 0.8738714797 4.664058
2    4.538691 2.6436724611 6.433710
3    2.768965 0.8738714797 4.664058
4    4.147044 2.2519737671 6.042115
5    3.164409 1.2692656709 5.059552
6    4.147044 2.2519737671 6.042115
7    2.768965 0.8738714797 4.664058
8    4.147044 2.2519737671 6.042115
9    2.428441 0.5332700416 4.323611
10   4.790504 2.8954701190 6.685537
11   2.440882 0.5457157131 4.336049
12   4.147044 2.2519737671 6.042115
13   2.910347 1.0152691998 4.805425
14   4.147044 2.2519737671 6.042115
15   2.910347 1.0152691998 4.805425
16   4.200872 2.3058141001 6.095930
17   2.440882 0.5457157131 4.336049
18   5.185948 3.2908996567 7.080996
19   2.440882 0.5457157131 4.336049
20   4.200872 2.3058141001 6.095930
21   2.555115 0.6599801230 4.450250
22   4.200872 2.3058141001 6.095930
23   2.555115 0.6599801230 4.450250
24   4.790504 2.8954701190 6.685537
25   2.727752 0.8326527681 4.622852
26   4.200872 2.3058141001 6.095930
27   2.727752 0.8326527681 4.622852
28   5.098789 3.2037380584 6.993839
29   3.123197 1.2280439640 5.018349
30   4.200872 2.3058141001 6.095930
31   3.123197 1.2280439640 5.018349
32   4.200872 2.3058141001 6.095930
33   2.727752 0.8326527681 4.622852
34   4.200872 2.3058141001 6.095930
35   2.741430 0.8463329097 4.636527
```

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.