1. What are the top 6 brands in the category in terms of dollar sales? What are the market shares of the 6 brands (assuming there are only 6 brands in the market).

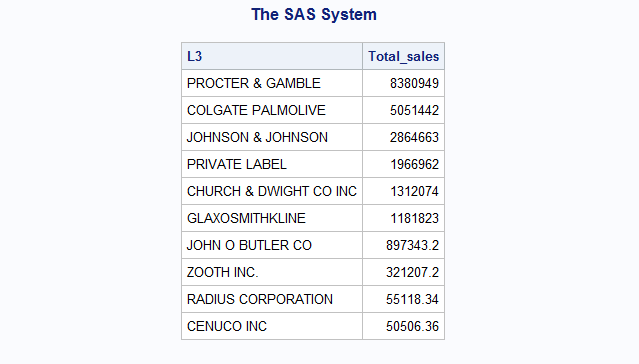




**Inference:**

The Top 6 brands for Tooth Brush and there market share are as follows:

1. ORAL => Oral B (By Procter & Gamble) has a market share of 30.65%
2. COLGA => Colgate (By Colgate Palmolive) has a market share of 24.17%
3. REACH => Reach (By Johnson & Johnson) has a market share of 15.59%
4. PRIVA => Private Label (In- House Brands of the Grocery Stores) has a market share of 11.55%
5. CREST => Crest (By Procter & Gamble) has a market share of 10.50%
6. MENTA => Mentadent (By Church & Dwight Co Inc) has a market share of 7.54%
7. Which companies are the major players in the category? Which company owns which brands?



**Inference:**

The major 10 companies in Tooth Brush Category are:

1. Procter & Gamble
2. Colgate Palmolive
3. Johnson & Johnson
4. Private Label (In- House Brands of the Grocery Stores)
5. Church & Dwight Co Inc
6. Galaxosmithkline
7. John O Butler Co.
8. Zooth Inc.
9. Radius Corporation
10. Cenuco Inc

All these brands are arranged in the increasing value of total sales over the given period.

The list of all the companies and their brands are listed below:





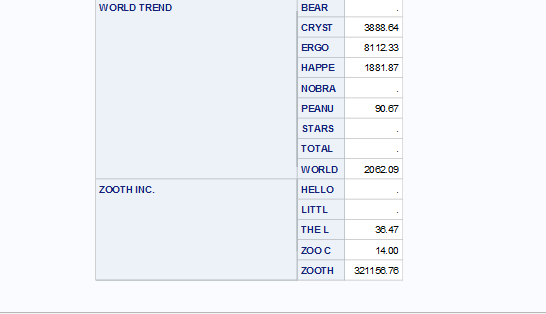




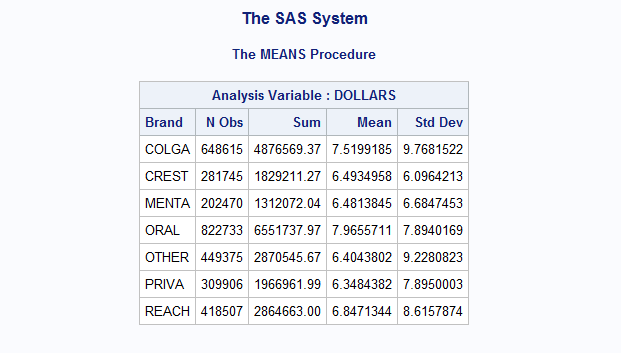








1. Create a 7th brand called “Other” that has all other brands that are not in the top 6.

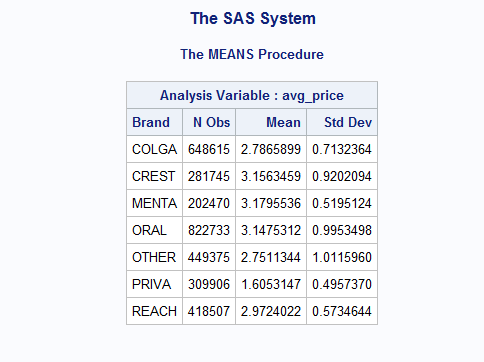


**Inference:**

Apart from the Top 6 brands that were mentioned in Question 1, have been merged under “Other”. And the total sales of these brands acquires 3rd position in terms of overall sales for the given period which amounts to $ 2870545.67.

1. Find average prices, display, features of each of the 7 brands.

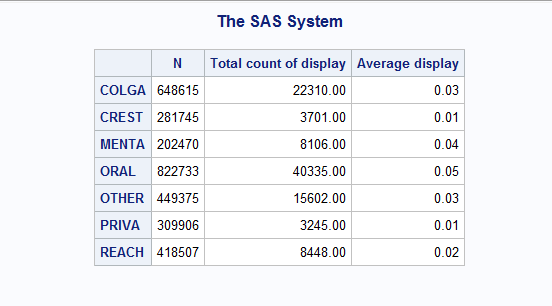
**Average Price:**



**Inference:**

The above table shows the average price per unit of all the 7 brands sold. Where Private Labels being the cheapest ($ 1.61 per tooth brush) and Mentadent ($ 3.18 per tooth brush) being the most expensive among all.

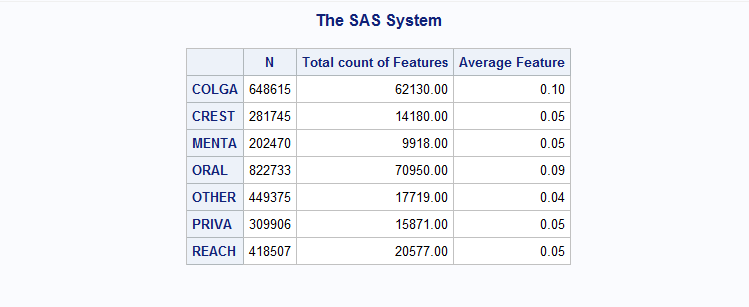
**Average Display:**



**Inference:**

The tooth brushes under the Brand name Oral B were on in- store display or in-store promotion for the maximum number of times. Followed by Colgate and the inhouse brands (Private Label) of stores were the on display for the least number of times.

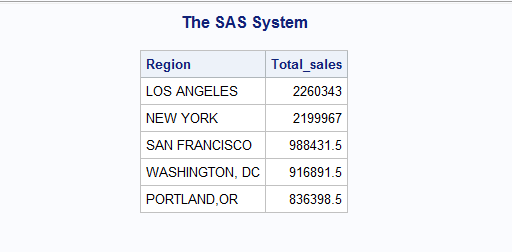
**Average Feature:**



**Inference:**

For the advertisement promotion on weekly basis here also Brand Oral B Tops. Followed by Colgate and the least number of promotions on the weekly basis is done by Brand Mentadent.

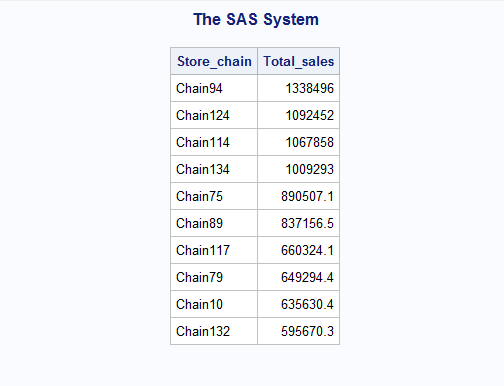
1. What are the top 5 regions in terms of dollar sales?



**Inference:**

Sales of Tooth Brushes of all brands among all the stores is highest in Los Angeles ($ 2,260,343). Followed by New York ($ 2,199,967). And lowest being in Portland, Oregon ($ 836,398.5).

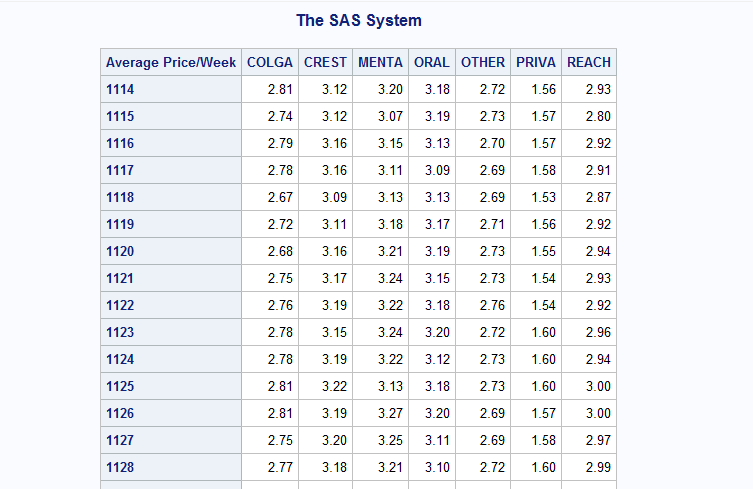
1. What are the top 10 store chains that sell a lot of your category in terms of dollar sales?

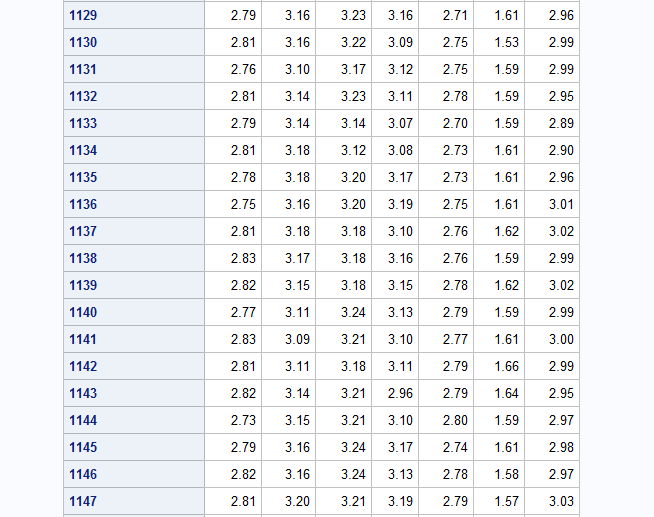


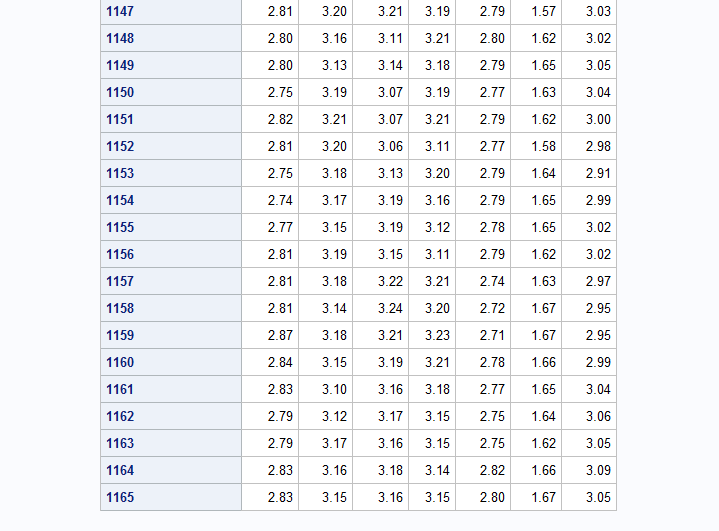
**Inference:**

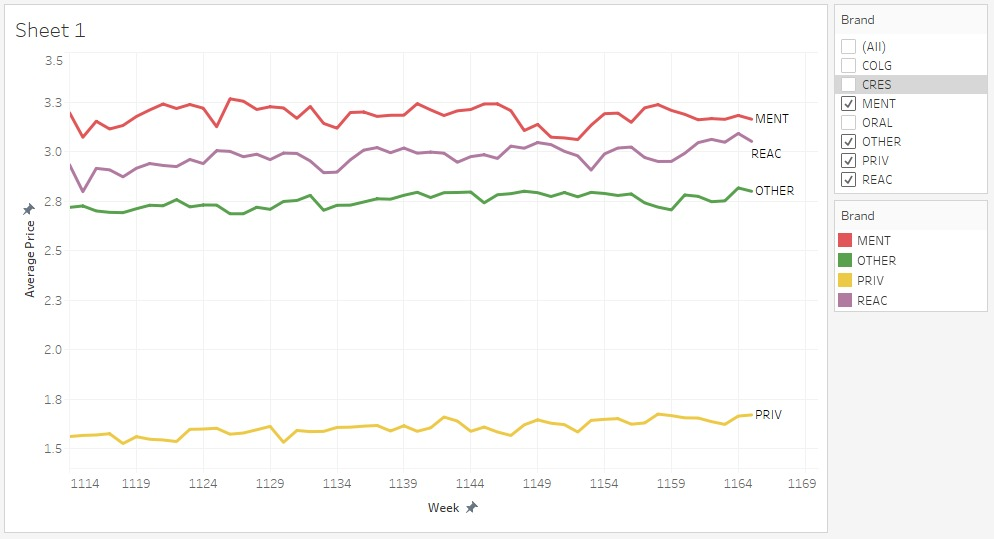
“Chain94” is the Store chain which has sold the highest number of Tooth brushes for the given period. The total number of sales being $ 1,338,496. Followed by ”Chain124” ($ 1,092,452 ) and “Chain132” at 10th position with sales of $ 595670.30.

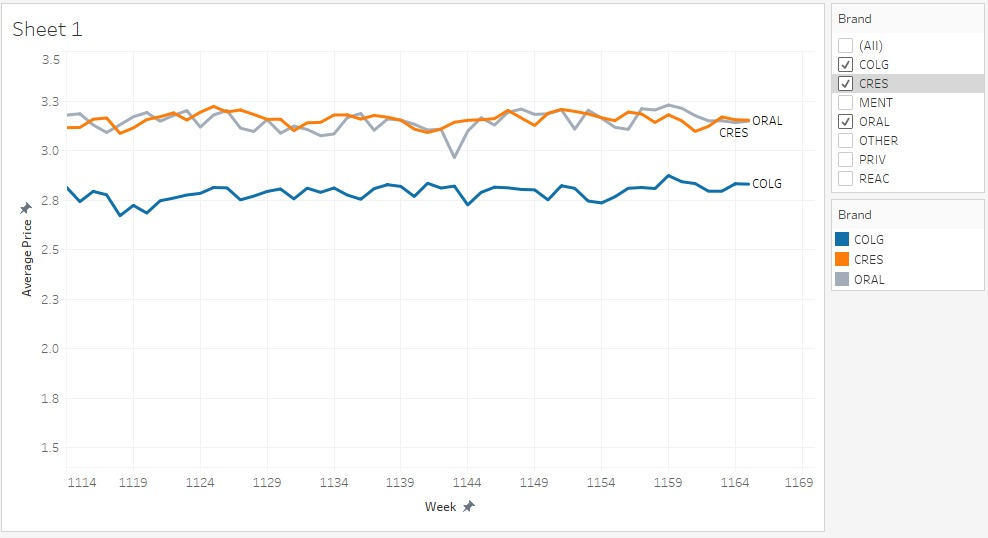
1. What is the average price per unit of 7 brands by week? Plot the average price by week (I wish to see a line plot of price by week). Comment on your findings.











**Inference:**

The average weekly price of all the tooth brushes have been nearly constant throughout the year. There have been minor variations in the brands average weekly prices which might have been because of some product under that brand being on Promotional Display or on Featured Promotion. This might have led the average weekly prices of the Tooth brush to change.

1. Assume you are manager of a brand (out of the top 6). Write a short paragraph stating what you learned from this descriptive analysis (steps 1-7).

Being the Brand Manager of “Oral B” which is owned by PROCTER & GAMBLE. Oral B Tooth Brush is the most selling tooth brush for the given time frame approximately having a market share of 31%. There is a stiff competition from Brand “Colgate”. Private Labels (In-House Brands of Stores) also play a significant role in the market and have a comparatively higher market share in comparison to some of the renounced brands Like “Crest”, also owned by PROCTER & GAMBLE. The average weekly Price per unit of tooth brush for my brand is ranging between $ 2.96 to $ 3.21, which is higher than “Colgate”. Under the Brand “Oral B” we provide our customer a wide range of tooth brushes, trying to cater every demographic segment by providing them tooth brushes with different feature which is convenient for them to use and indirectly take care of their teeth.

**Statistical Analysis**

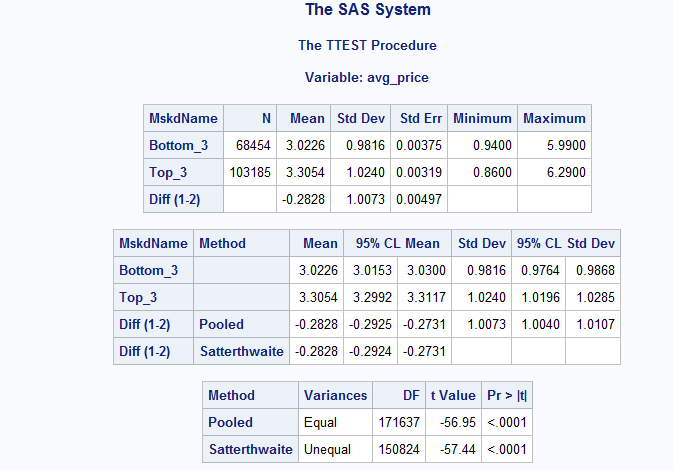
1. Do large stores (top 3 stores) have higher average price per unit than small stores (stores ranked 8-10) for brand 1 (the top brand in Q1). Test and report your results and comments.

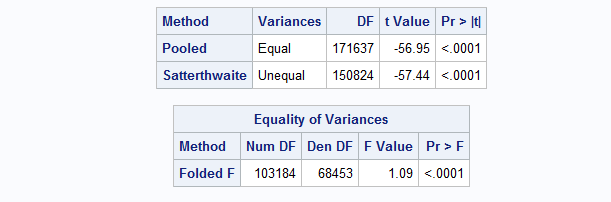
**Answer**:

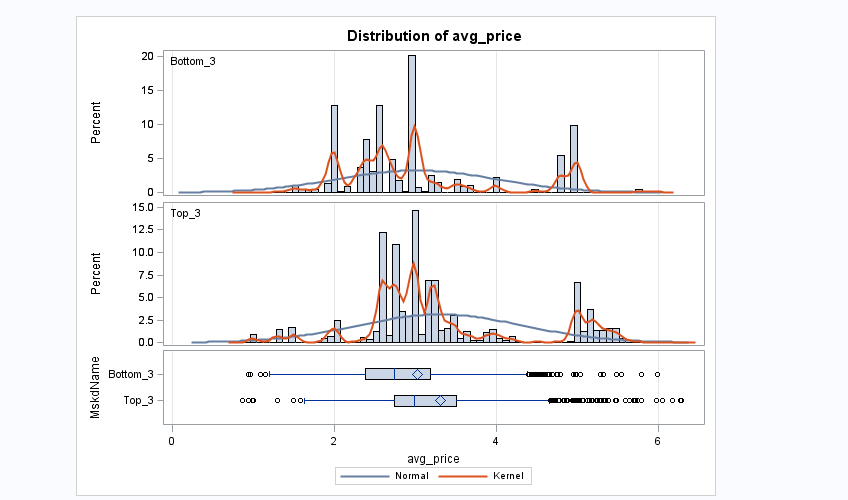
We applied a **t test on Brand “Oral B”** and designed the Hypothesis as follows:

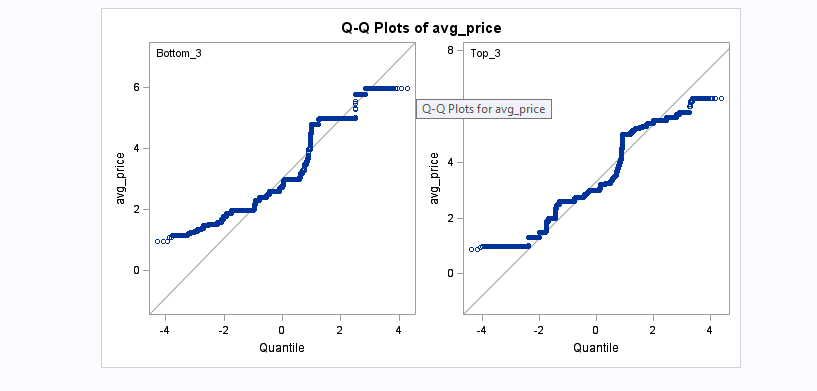
H0: Mean average price per unit of larger stores (Top 3) is equal to Mean average price per unit of Small stores (Store 8-10)

H1: Mean average price per unit of larger stores (Top 3) is greater than Mean average price per unit of Small stores (Store 8-10)









**Inference:**

As the **significance probability for the F-test is less than 5% (<0.01%) we reject the null.** Hence the variances for the two groups Man and Women are not equal. The **correct t Critical value is -57.44**, based on the probability which is less than 5% (<0.01%) **we reject the null.** So, there is an evidence that the Mean average price per unit of larger stores (Top 3) is greater than Mean average price per unit of Small stores (Store 8-10).

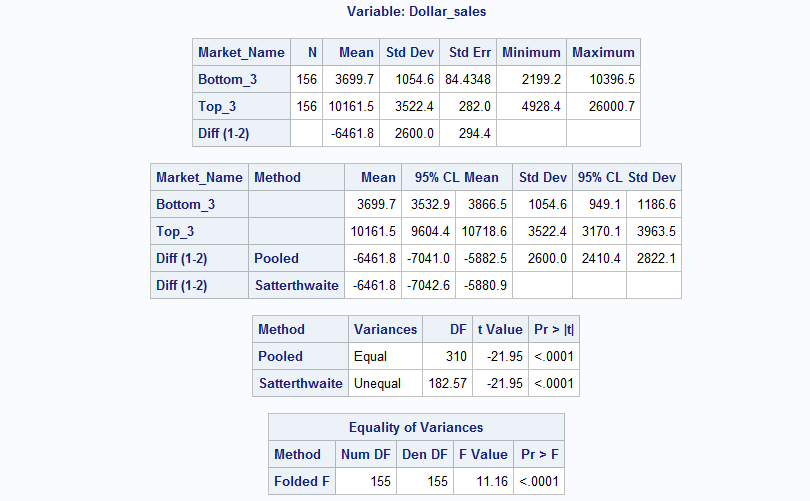
1. Develop three additional hypotheses linking useful variables to dollar sales, test them and report your findings.
2. Weekly dollar sales in the top 3 regions for ORAL B brand is higher than that of the bottom 3 regions

**Answer**:

We applied **t-test on Brand “Oral B”** and designed the Hypothesis as follows:

H0: Mean Weekly dollar sales in the top 3 regions for ORAL B brand is less than that of the bottom 3 regions

H1: Mean Weekly dollar sales in the top 3 regions for ORAL B brand is higher than that of the bottom 3 regions



**Inference:**

As the **significance probability for the F-test is less than 5% (<0.01%) we reject the null.** Hence the variances for the two groups are not equal. The **correct t Critical value is 182.57**, based on the probability which is less than 5% (<0.01%) **we reject the null.** So, there is an evidence that the Weekly dollar sales in the top 3 regions for ORAL B brand is higher than that of the bottom 3 regions.

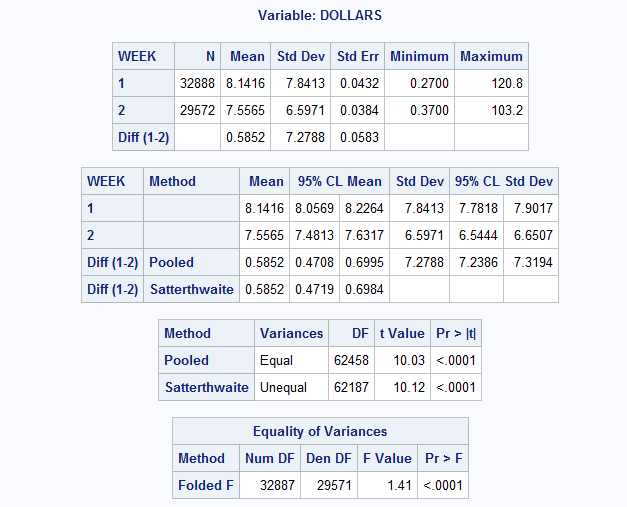
b) For ORAL B, there is a significant decrease in dollar sales during the OFF season (WEEK 1159 to 1163) compared to the HOLIDAY season (WEEK 1140 to 1144)

**Answer**:

We applied **t test on Brand “Oral B”** and designed the Hypothesis as follows:

H0: Mean dollar sales during the OFF season (WEEK 1159 to 1163) is greater than compared to the HOLIDAY season (WEEK 1140 to 1144)

H1: Mean dollar sales during the OFF season (WEEK 1159 to 1163) is less than compared to the HOLIDAY season (WEEK 1140 to 1144)



**Inference:**

As the **significance probability for the F-test is less than 5% (<0.01%) we reject the null.** Hence the variances for the two groups are not equal. The **correct t Critical value is 10.12**, based on the probability which is less than 5% (<0.01%) **we reject the null.** So, there is an evidence that there is a significant decrease in dollar sales during the OFF season (WEEK 1159 to 1163) compared to the HOLIDAY season (WEEK 1140 to 1144)

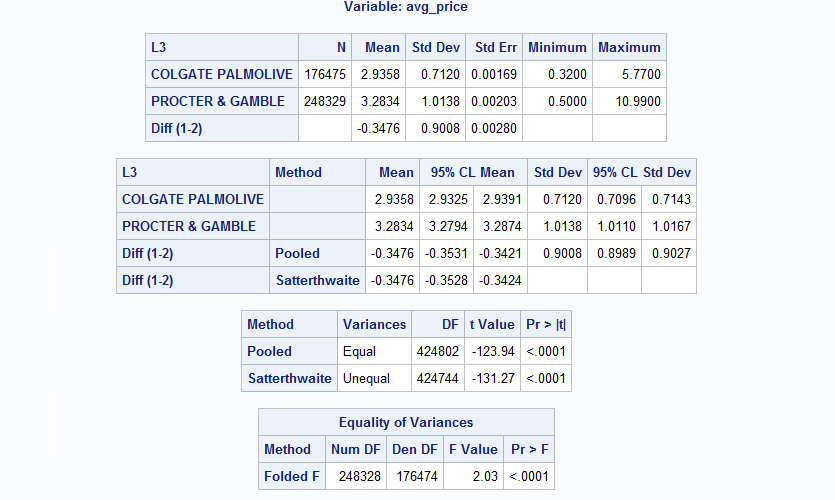
c) In the top-4 regions according to the dollar sales, there is a significant difference in the average price per unit for the top-2 companies.

**Answer**:

We applied the **t-test on Brand “Oral B”** and designed the Hypothesis as follows:

H0: Mean price per unit is not equal for the top-2 companies

H1: Mean price per unit is equal for the top-2 companies



**Inference:**

As the **significance probability for the F-test is less than 5% (<0.01%) we reject the null.** Hence the variances for the two groups are not equal. The **correct t Critical value is -131.27**, based on the probability which is less than 5% (<0.01%) **we reject the null.** So, there is an evidence that according to the dollar sales, there is a significant difference in the average price per unit for the top-2 companies

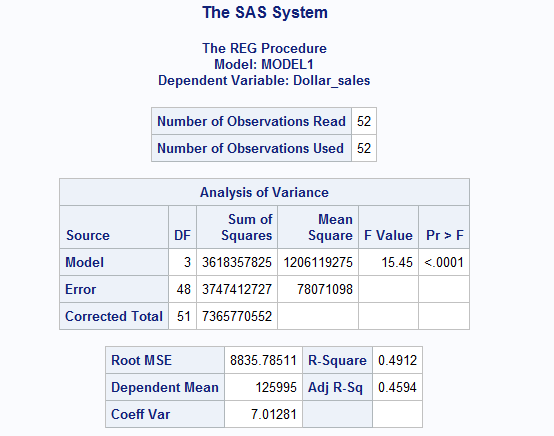
1. For the top brand: run a regression model with weekly dollar sales as dependent variable. Use average weekly price per unit, average display, average feature, and other useful variables in your regression and answer the following questions:

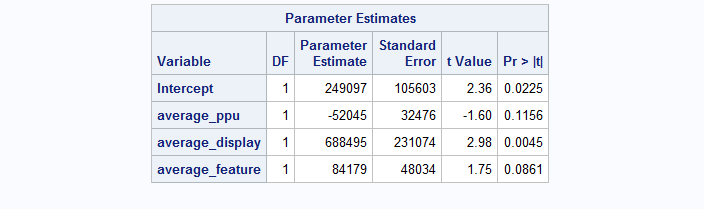
**Answer:**

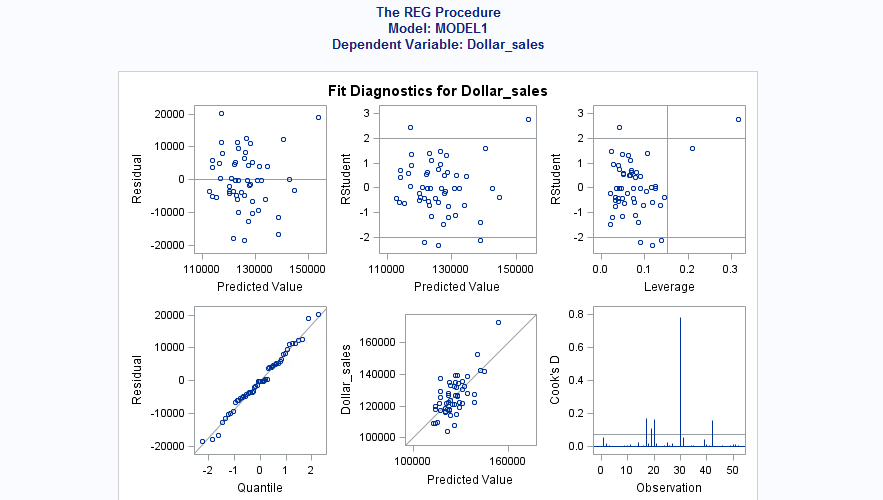
**Model is:**

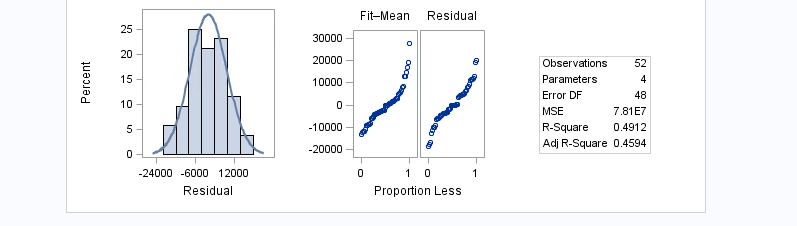
**Dollar Sales = Constant + x1\*Avg. Price per unit + x2 \* Avg Feature + x3 \* Avg Display + error**

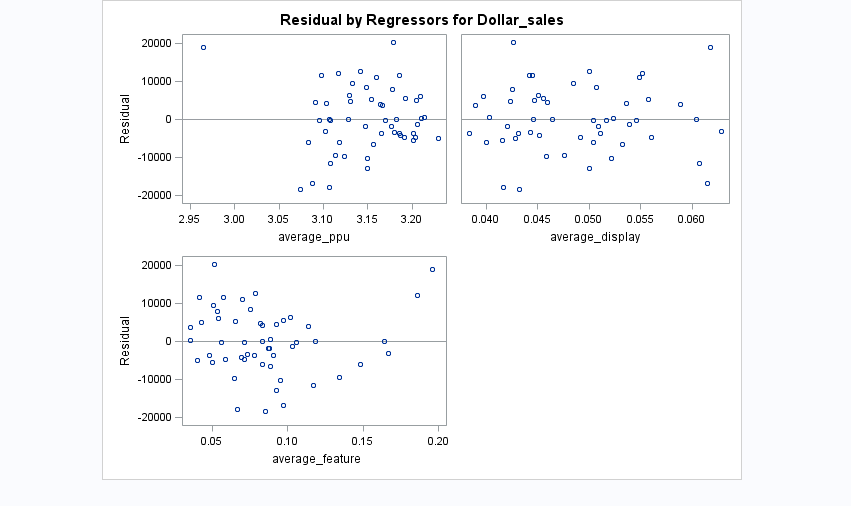
**Regression out puts is as follows:**











1. What is the R-sq and adj R-sq of the model?

Answer:

The R-square value for the model is 0.4912 and Adjusted R-square value for the model is 0.4594.

1. Which coefficients are significant?

**Answer:**

Average display(average\_display) is statistically significant at 95% confidence interval, Average feature(average\_feature) is statistically significant at 90% confidence interval but average price per unit(average\_ppu) is not statistically significant at either 90% or 95% confidence interval.

1. Which variables are most important in explaining sales?

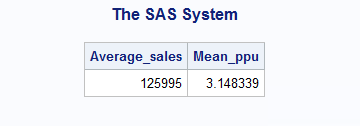
**Answer:**

Average Display is the most important variable that is significant at alpha = 5% followed by Average feature which is significant at alpha=10%.

1. Interpret the meaning of the price coefficient? What is the price elasticity?

**Answer:**

If there is a $1 increase in average price of tooth brush the sales would decrease by $52045 Weekly across all store. This is inferred from the negative sign on the coefficient which suggests that as the price increases, sales decreases.



Price elasticity= (-52049)\*(3.148/125995) = -1.29

**Inference**: means 1% increase in Avg price of the product will decrease the Weekly sales of Brand=Oral B across all stores by 1.29%

1. Interpret the meaning of the display coefficient?

**Answer:**

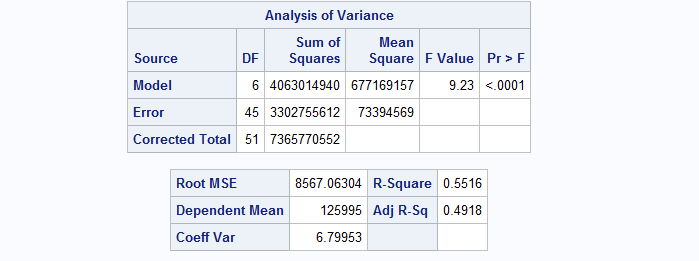
The interpretation of the coefficient of display coefficient is that when there is a price promotion (in store), the sales of the “Oral B” tooth brush would increase by 688495 units.

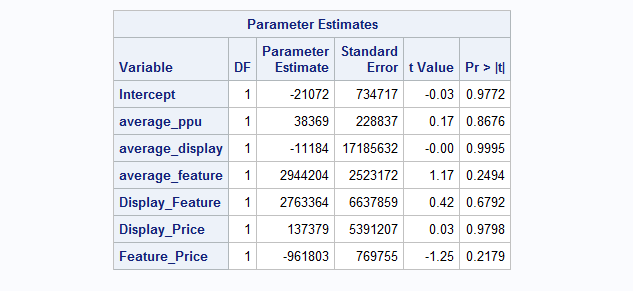
1. Test whether there is an interaction between display, feature and price. Comment on your findings.

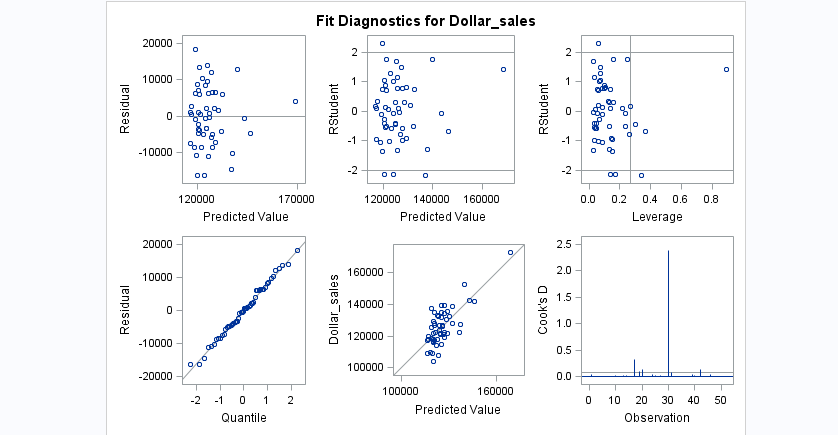
**Answer:**

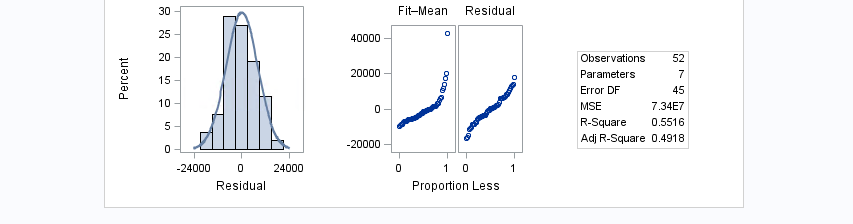
**Model is:**

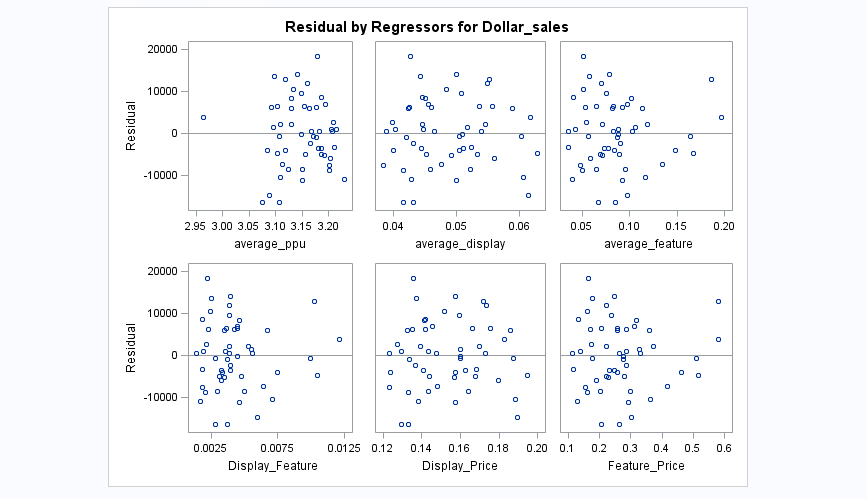
**Dollar sales(Dollar\_sales)= constant + x1 \* Avg. Price per unit(average\_ppu)+ x2 \*Avg. display( average\_display) + x3 \* Avg. Feature(average\_feature) + x4\* Avg. Display\* Avg. Feature (Display\_Feature)+ x5\* Avg. Display\* Avg. Price (Display\_Price)+ x6\* Avg. Price\* Avg. Feature (Feature\_Price) + error**











**Inference:**

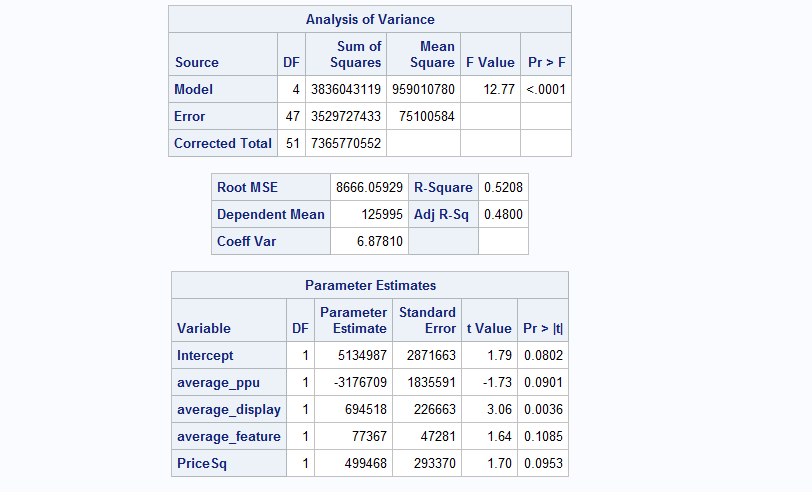
With the inclusion of interaction variables to determine it effect on the model defined. We can see that **the P value for the model independent variable and the interaction variable has become insignificant even at 90% confidence level**. Making the overall model insignificant and is insignificant at 90% confidence interval. The residual plots of independent variable also suggest the existence of Heteroscedasticity

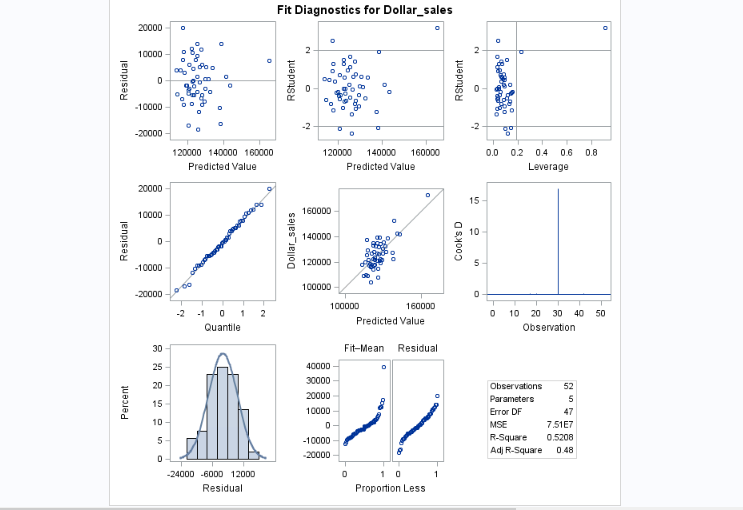
1. Test whether the effect of price is non-linear. Comment on your findings.

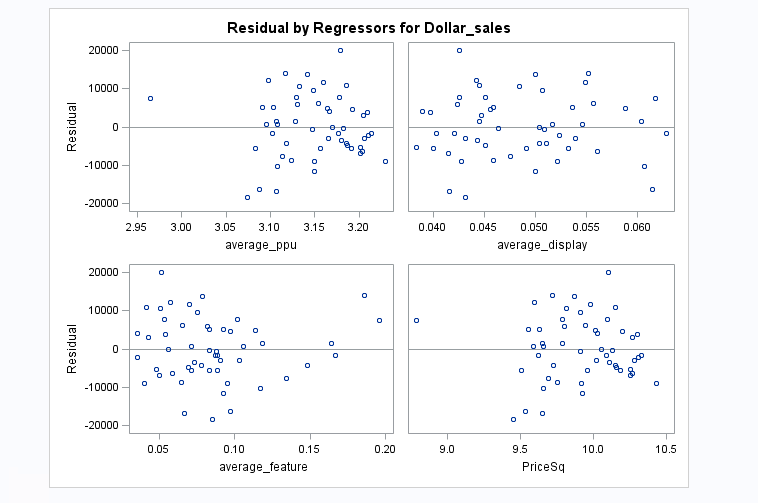
**Answer:**

**Model Is:**

**Dollar sales(Dollar\_sales)= )= constant + x1 \* Avg. Price per unit(average\_ppu)+ x2 \*Avg. display( average\_display) + x3 \* Avg. Feature(average\_feature) + x4\* PriceSq + error**







**Inference:**

In order to determine the Linear/Non Linear ship of Variable Price. We are including variable PriceSq (Price \* Price) in the Regression model to check the relation-ship. On observing the P value of Variable **PriceSq** , we can say that the variable is Significant at 90% confidence level and other independent variables included in the model are also significant at 90% confidence level except for variable average\_feature, which is insignificant even at 90% confidence level.

1. Test using VIF and COLLIN whether there is multicollinearity in the model? Comment on your findings.

**Answer:**

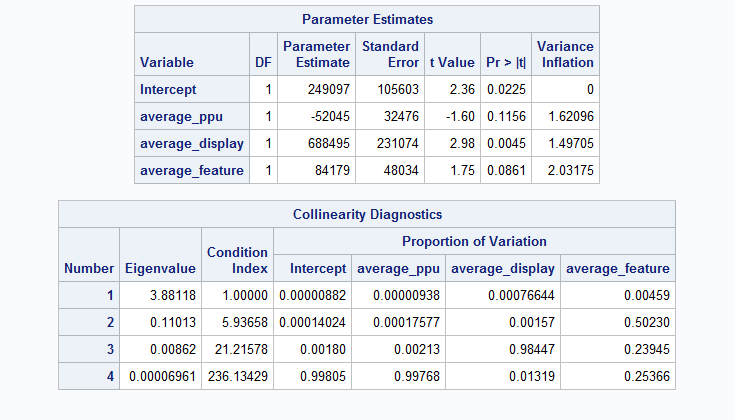
**Model Is:**

**Dollar sales(Dollar\_sales)= constant + x1 \* Avg. Price per unit(average\_ppu)+ x2 \*Avg. display( average\_display) + x3 \* Avg. Feature(average\_feature) + error**

(**SAS code**:

proc reg data=W.Weekly\_sales;

model Dollar\_sales=average\_ppu average\_display average\_feature /vif collin;run;)



**Inference:**

VIF:

Checking for multicollinearity via VIF and COLLIN function

For VIF value if greater than 10 there could be multicollinearity , but as we see the VIF values for the independent variables in the model we can say that VIF value is less than 10 but we cannot conclude that there’s no multicollinearity.

COLLIN:

Checking : Condition Index is more than 100 & two or more variables have high proportion of variance. We can see from the delivered output that Condition index in 236.13 for variable average\_feature which is greater than 100 and also the proportion of variance is high between variable average\_ppu & average\_feature.

Hence, we can say that Multicollinearity exists in out model.

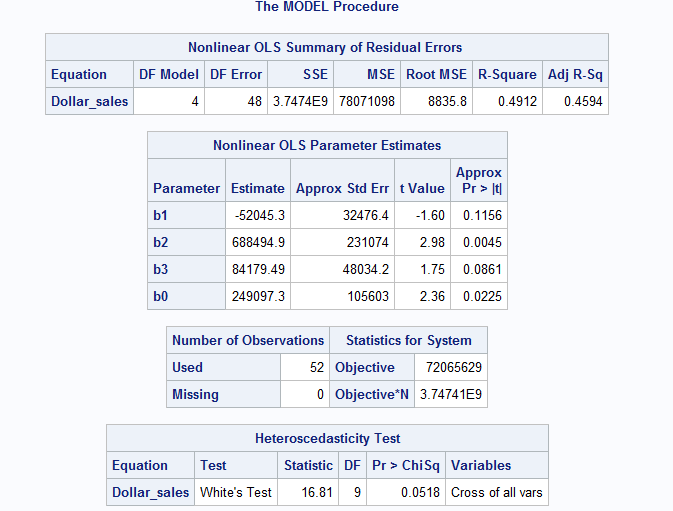
1. Test for presence of heteroscedasticity using White test. Do A WLS if needed. Comment on your findings.

**Answer for White Test**:

The Hypothesis as follows:

H0: Variance of Errors are constant

H1: Variance of errors are not constant



**Inference:**

Since White's Test statistic is 16.81 and p value 0.0518 which is insignificant at 90% confidence interval. Therefore, we reject the null hypothesis and conclude that variance of

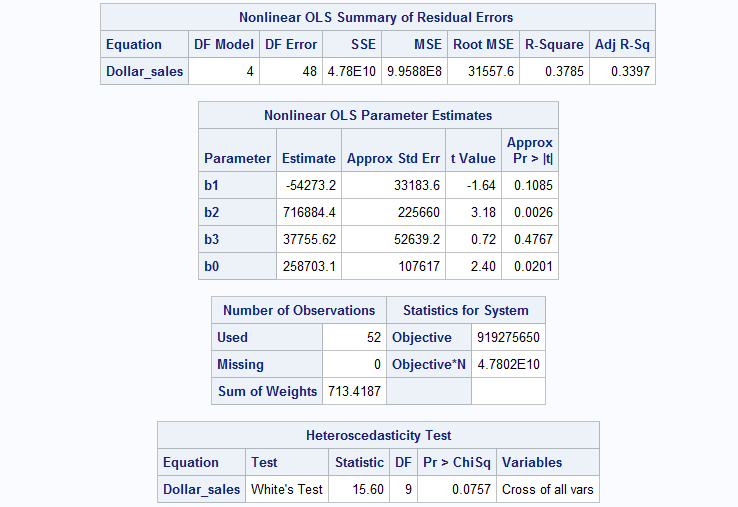
error is not constant.

We can see in the Residuals plots obtained from the regression for model

**Dollar Sales = Constant + x1\*Avg. Price per unit + x2 \* Avg Feature + x3 \* Avg Display + error**

(Dollar\_sales= average\_ppu average\_display average\_feature) from the **residual plots** observed for each independent variable obtained from the PROC REG model output, the variation in residual error is evident in variable average\_feature. so we will consider the Weighted Variable as avergae\_feature as compared to other independent variables included in the model.

**For WLS:**



**Inference:**

We can see from the result that the after WLS implementation there is a very slight change in the value of White T Statistics and P value which of 0.0757 which has only increased by 0.0239, which is insignificant at 90% confidence interval. We can say that Heteroscedasticity is present in the model and model needs to be corrected to negate the effects of Heteroscedasticity