**ANALYSIS OF TOOTHBRUSHES**

**GROUP 7**

**Predictive Analytics with SAS**

Barunjit Banerjee

Swati Bansal

Sai Harika Kunapareddy

Piyush Mittal

Dhananjay Sahu

Pranusha Yanamandala



**INTRODUCTION:**

The main purpose of this project is to analyze and provide valuable insights from the sales data, grocery stores and panel data obtained for various brands of toothbrushes in the market. Our main goal was to extract information useful in expanding the market share for our brand.

Three types of data that we handled:

Panel data: holds the dollar sales by units of different brands (by UPC) across different stores and week for each household.

Store level data holds the store level data and the information on price, display and features on the UPC level.

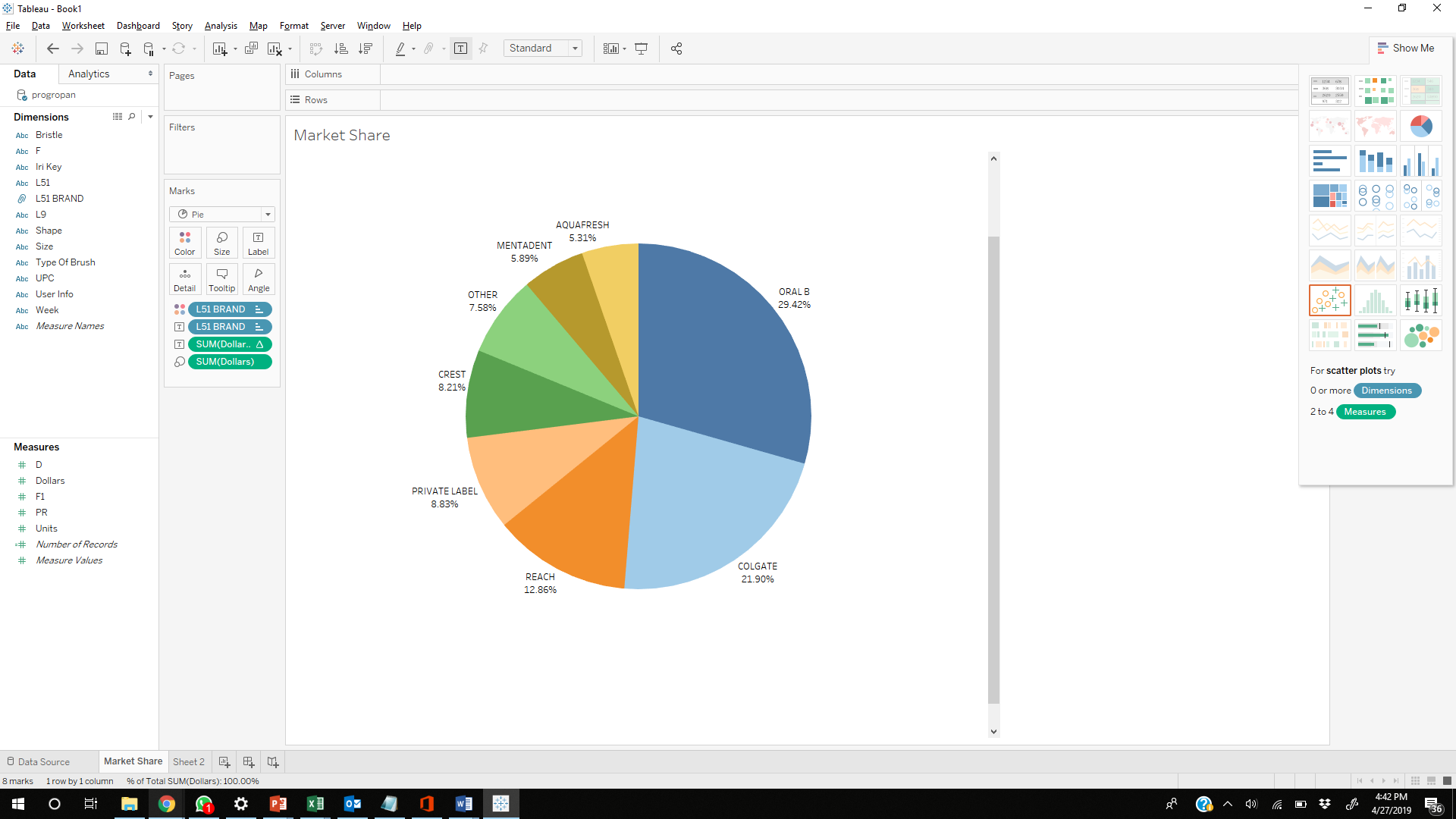
Product data: holds the brand names mapped to the UPC.

Our Brand Oral B Advantage is the most popular brand in terms of market share across the country. We have tried to provide certain insights that can be used in the future to grow the market share. We also performed clustering to understand the customer purchase behavior so that we can help identify segments of target markets.

The conclusion and recommendations are based on a combination of data and statistical analysis tools used for the completion of the project.

**Objective**

The purpose of this project is to help Oral B Advantage grow or sustain its market share. There are three major players in terms of Brands, capturing most of the market share. The predictions are all based on the scanner and panel data.



**Exploratory Analysis**

## **Market Share:**

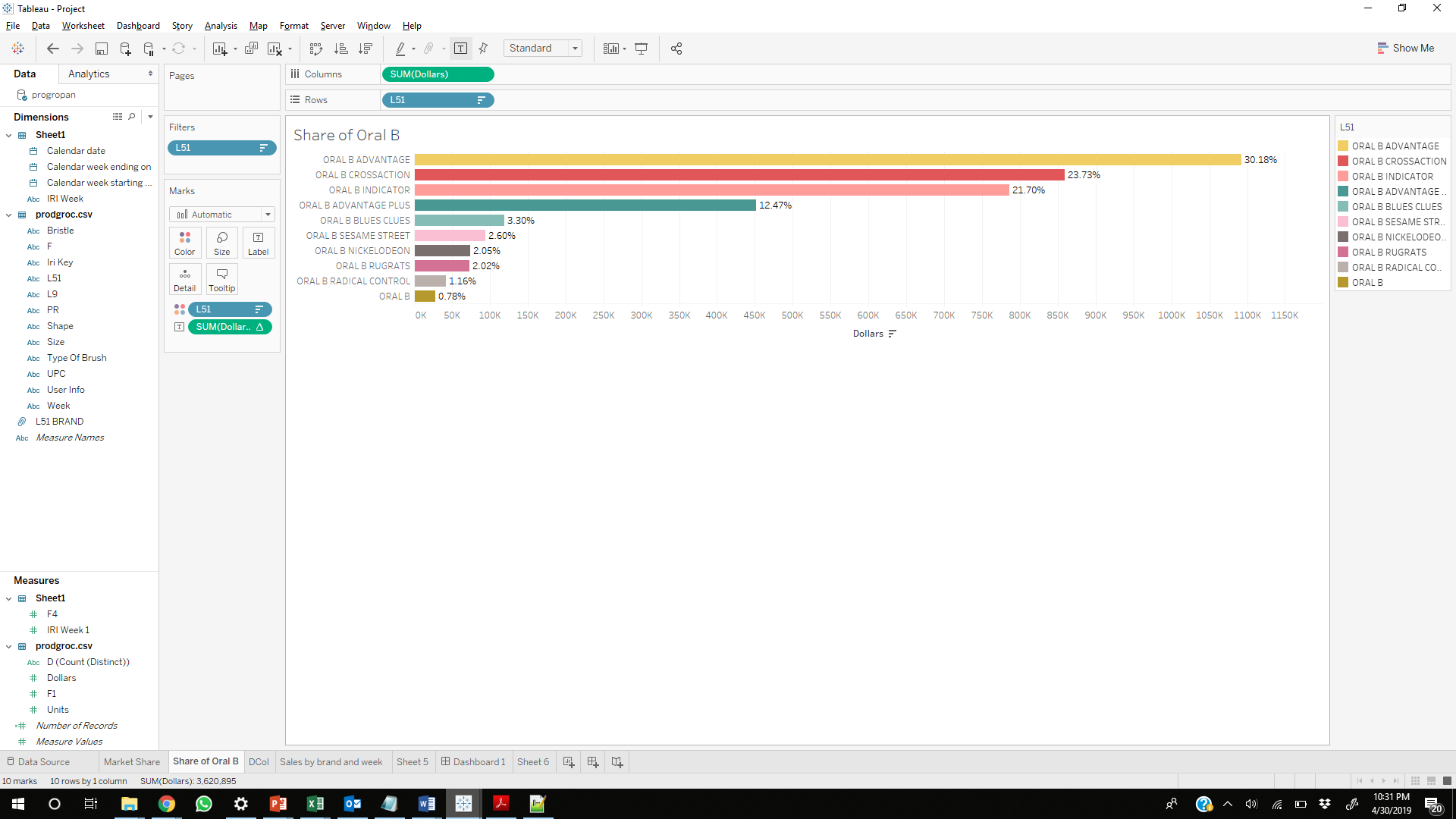
Oral B(29.42%) has the highest market share in terms of sales

followed by Colgate(21.90%) and Reach(12.86%). The top 5

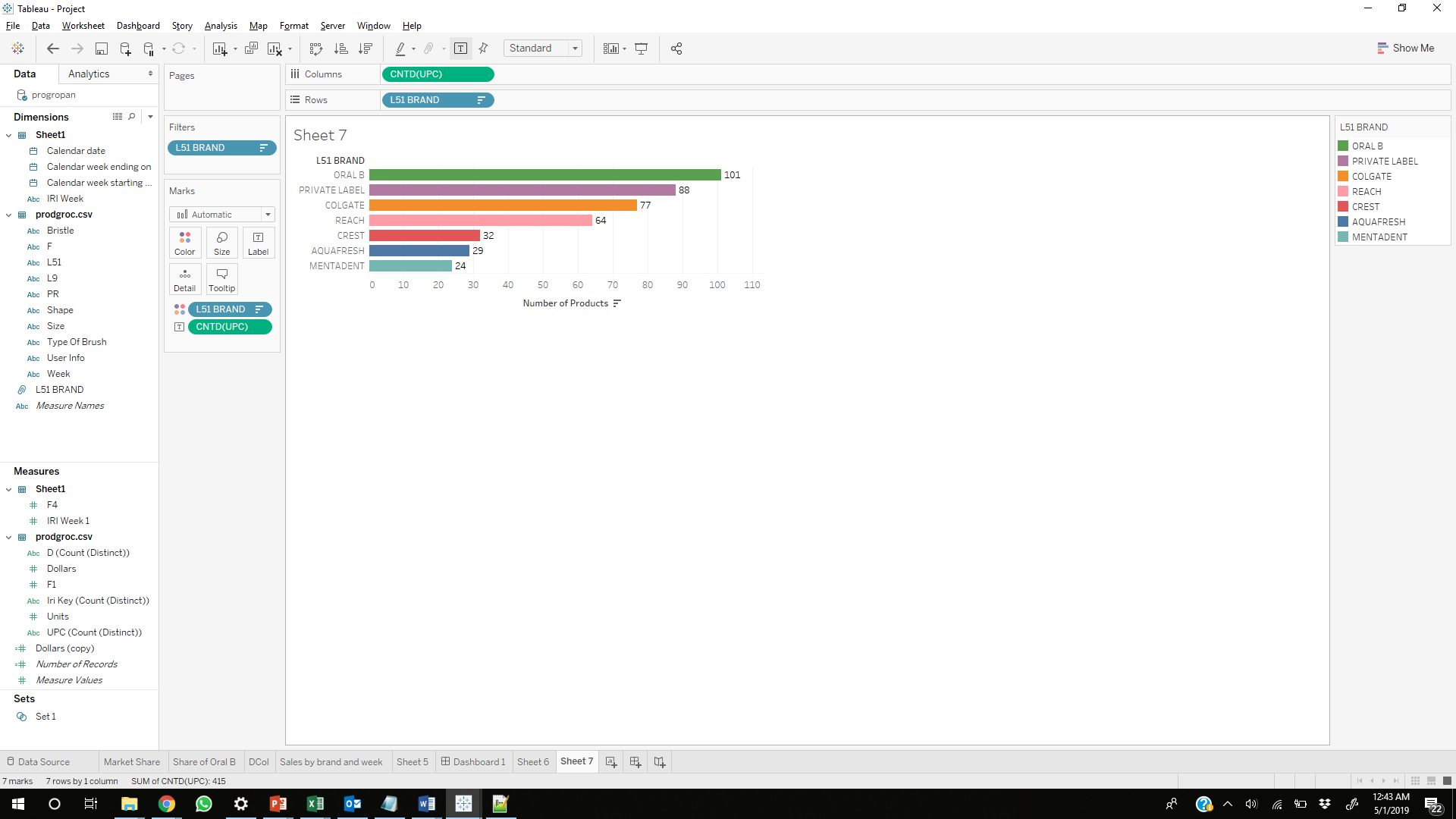
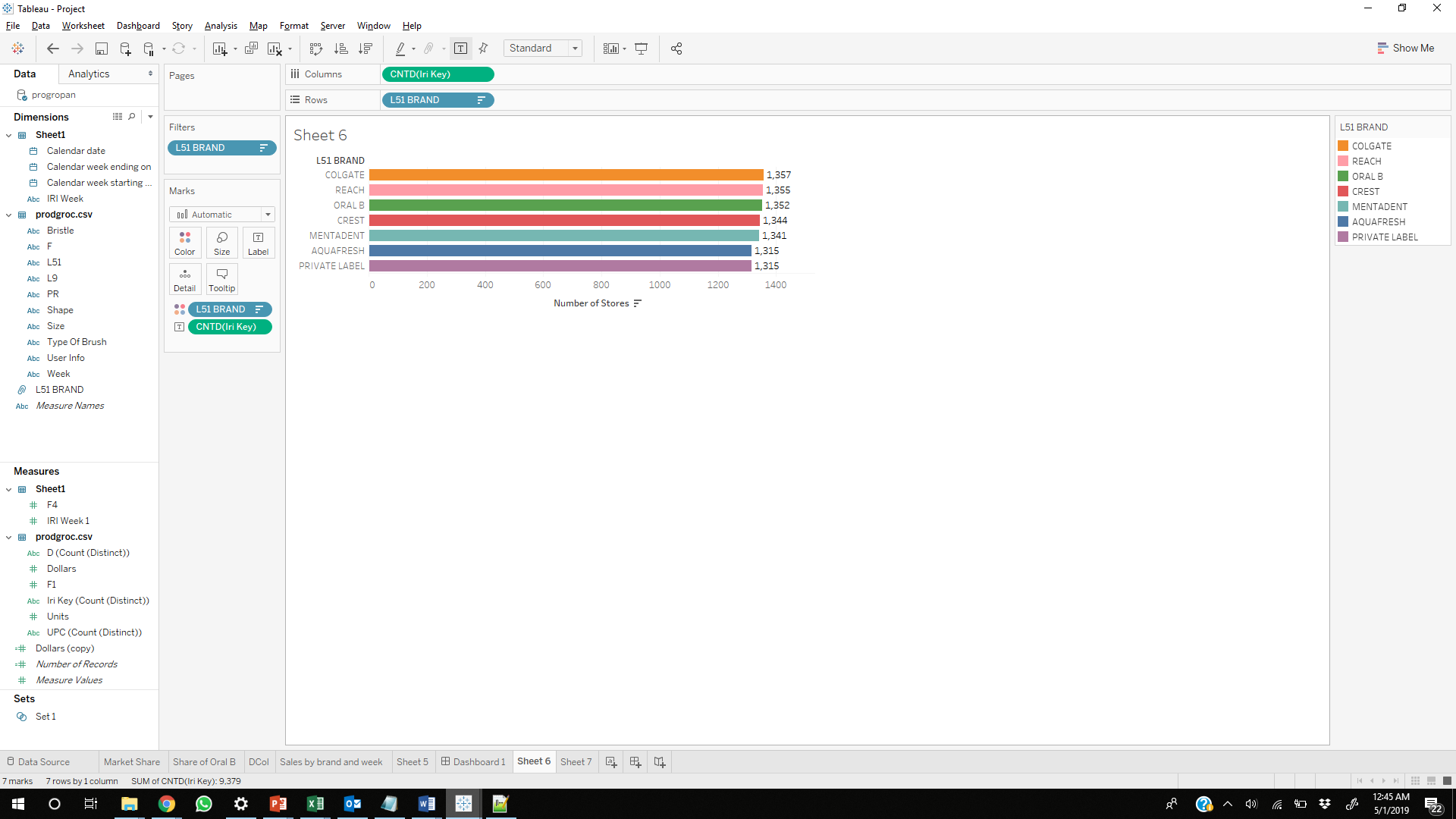
brands are Oral B, Colgate, Reach, Private Label and Crest

taking up to 80% of the market.

Within Oral B, Oral B Advantage has the highest sales and hence, has been chosen as the brand of interest for our further analysis.

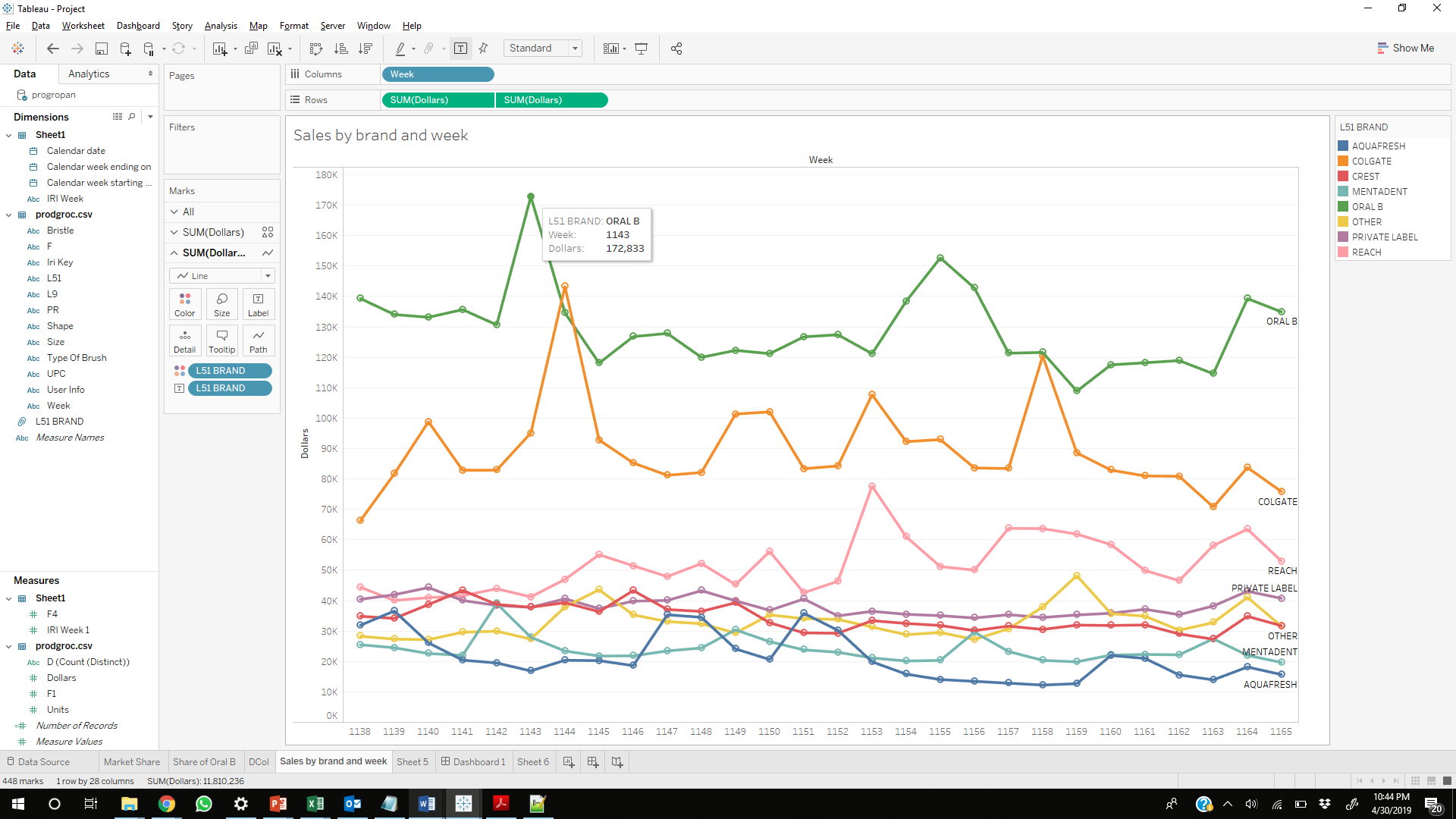


Our brand Oral B offers a wide variety of products and is the market leader. Whereas in terms of stores delivering those products, all the top brands maintain a close gap.

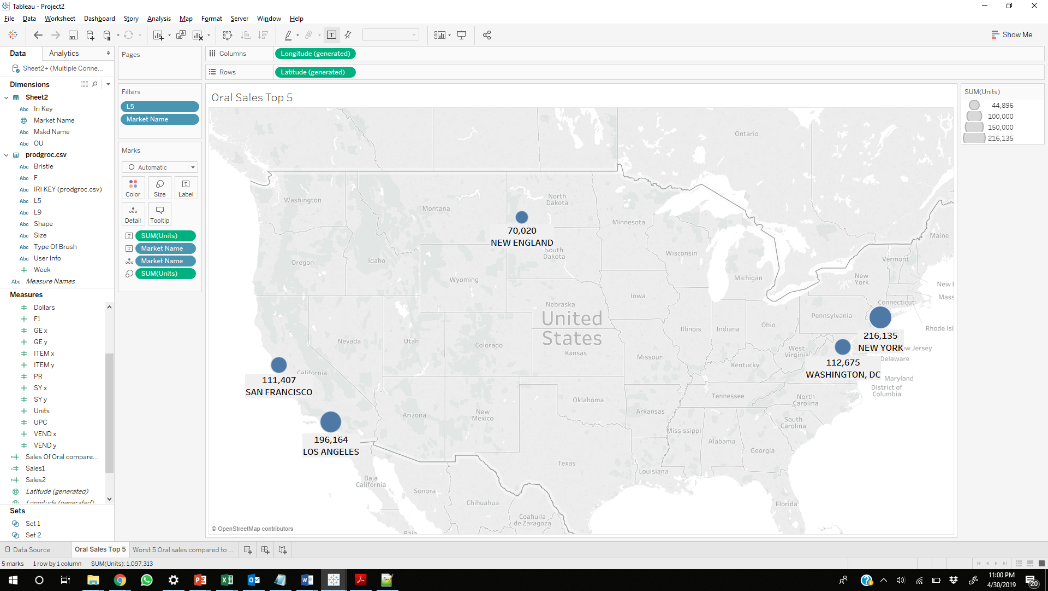
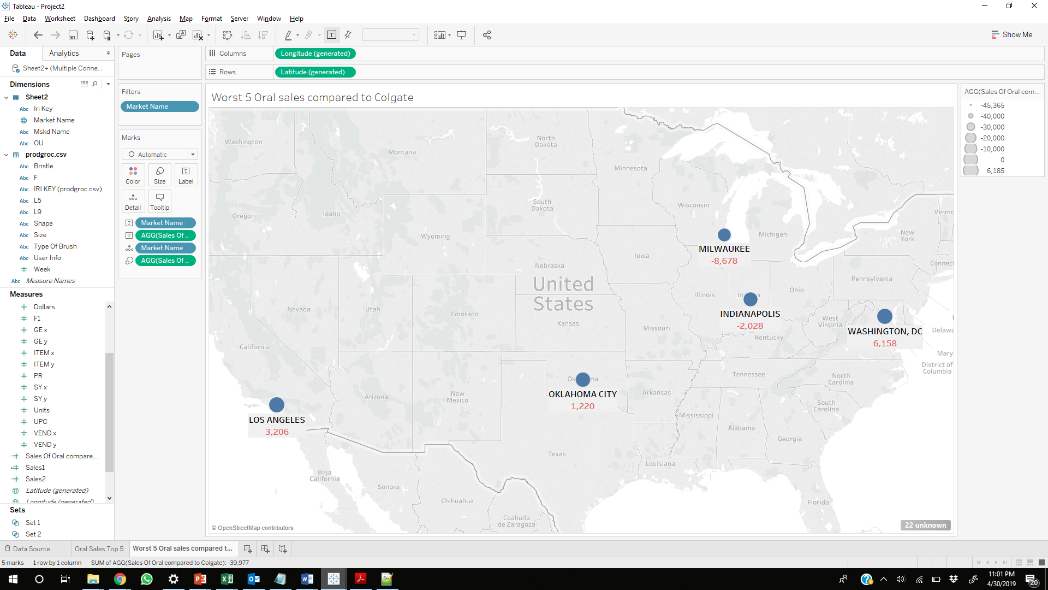


## **Sales by Week and Brand:**

Oral B has the highest sales across most of the weeks, performing well consistently. Maximum sales for Oral B in a week occurred during the week of Jul 23 - Jul 29 of 2001 in the given data.



## **Location Analysis:**



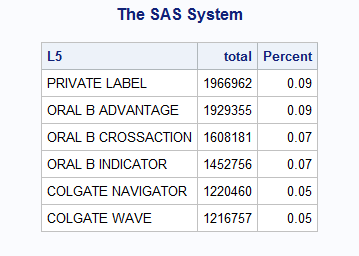
East and west regions show a high number of transactions of sales of Oral B. In the central US, we are losing the most customers, especially in Milwaukee. Los Angeles has second largest units of sales in the United States and is also a very competitive market with Colgate sales almost on par with that of our brand.

**Regression Analysis:**

We carried out panel regression on the merged product data and store level data.

**The market scenario:**

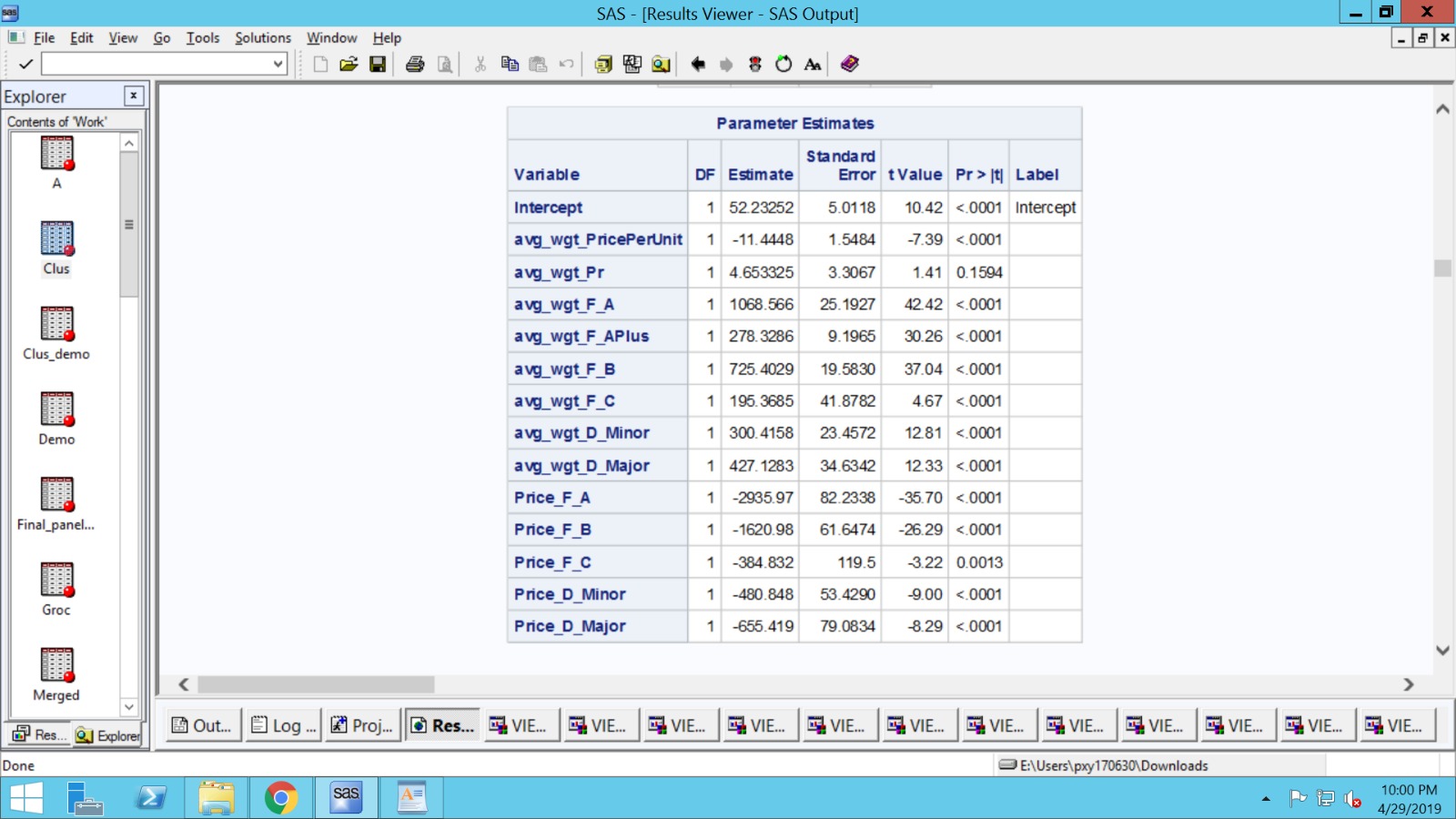
ORAL B Advantage holds the highest market share as a brand:



**Approach:**

* Introduced Dummy variables for “display” and “feature”.
* Considered possible interactions of the variables.
* Calculated the weighted average of the price, each display, each feature for the brand “ORAL B ADVANTAGE’ at a store level.
* Ran panel regression (Random effects). The Hausman test rejected the null hypothesis hence we ran fixed effects model. (Refer to Appendix1)

**Results:**



**Interpretations:**

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $11.

Price reduction is not significant since its p value is 0.1594.

For a given store, holding all other variables constant, having a large size ad (F\_A) when compared to not having large sized ad, increases weekly sales by $1068.

For a given store, holding all other variables constant, having a retail coupon or rebate (F\_APlus) when compared to not having coupon or rebate, increases weekly sales by $278.

For a given store, holding all other variables constant, having a medium size ad (F\_B) when compared to not having medium sized ad, increases weekly sales by $725.40.

For a given store, holding all other variables constant, having a small size ad (F\_C) when compared to not having small sized ad, increases weekly sales by $195.37.

For a given store, holding all other variables constant, having a minor display (D\_Minor) when compared to not having minor display, increases weekly sales by $300.42.

For a given store, holding all other variables constant, having a major display (D\_Major) when compared to not having major display, increases weekly sales by $427.13.

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $2935.97 in presence of F\_A.

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $1620.98 in presence of F\_B.

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $384.83 in presence of F\_C.

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $480.85 in presence of D\_Minor.

For a given store, holding all other variables constant, increasing the price per unit by $1 is associated with a decrease in sales by $655.42 in presence of D\_Major.

**Findings**

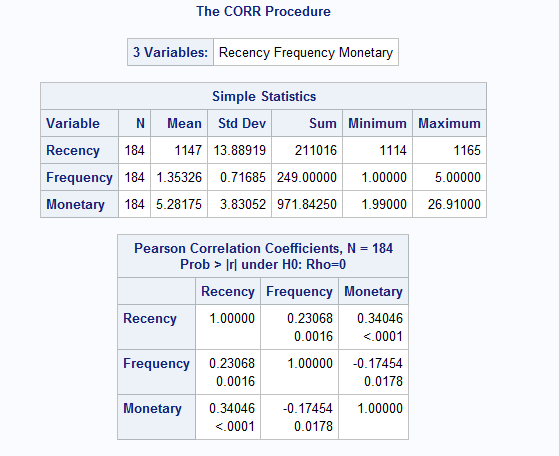
Features, display and promotions have a significant impact on the weekly sales for ‘ORAL B ADVANTAGE’.

The interaction terms of price per unit with feature and display have a synergy effect. A unit reduction of price increases sales to a much higher level in presence of feature and display.

**RFM / Clustering Analysis:**

We did RFM Analysis on panel grocery data and demographics data. The RFM analysis was done on our brand of focus (Oral B Advantage) so that we can target the market based on the different results and to identify those customers who are likely to respond to our campaigns and promotions.

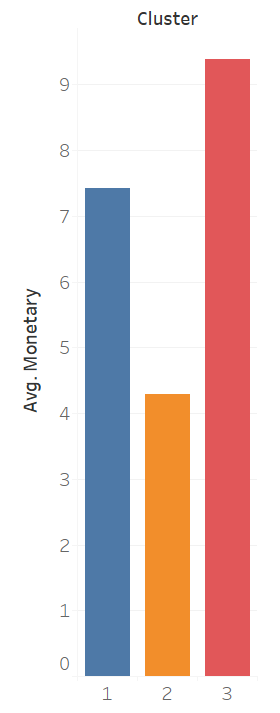
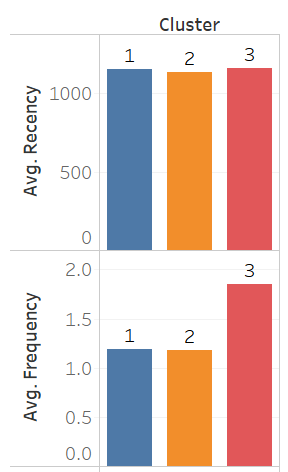
We then selected only those panelists who purchased “ORAL B ADVANTAGE” toothbrushes during their visits. We then created the recency, frequency and monetary variables for these customers by their most recent purchase, number of visits and the average dollars they spent on during each visit. We next checked the correlations for the 3 variables as below:



There is no strong correlation between recency, frequency and monetary. Therefore, we have included all the 3 variables. PROC RANK was used to rank these customers on these three variables with 5 being the highest and 1 being the lowest. RFM score for each customer is then obtained by combining these three variables.

After doing RFM, we performed Clustering such that the RFM scores are divided into 3 clusters. And then merged the demographic information to find out what characteristics each household in the different clusters have.

We can see households in the 3rd cluster spend more on Oral B advantage frequently since the average monetary value and average frequency of buying is high and is the most valuable cluster.



From the above bar charts and graph we can observe that customers in cluster 3 are our loyal customers, customers in cluster 1 are our potential profitable customers and those in cluster 2 are less profitable customers.

The most common characteristics of households using ORAL B advantage are:

* Overall majority of Households are ‘Married’ and earn more than 25K/ year
* Most common family size of Households is ‘two’ followed by ‘four’
* Households that prefer ‘ORAL B Advantage’ Age between ‘45-54’, the distribution is similar with respect to both females and males
* Households mostly belong to the occupation category ‘Professional or Technical’
* Most of the household males are working ‘Full-time’
* About 84% of the householders are ‘owners’

**CLUSTER 1**: Potentially Profitable customers

Customers Differentiating Characteristics:

* High M, R and Avg F scores.
* Cluster size: 63 households

Demographics:

* Households predominantly belong to ‘Age’ between ’45-54’ and ‘65+’.
* Households mostly holds job titles such as ‘clerical’, ‘operative’ and ‘laborer’
* Income lies between '$45,000 to $54,999 per yr.'
* Males of the household are between the age ’55-64’

**CLUSTER 2**- Potentially Profitable Customers

Customers Differentiating Characteristics:

* Low M score and High R score and Avg F score
* Cluster size: 74 households

Demographics:

* Households predominantly belong to age between ’45-54’
* Income lies between '$55,000 to $64,999 per year'

**CLUSTER 3** – Loyal Profitable customers

Customers Differentiating Characteristics:

* High R, F and M scores.
* Cluster size: 47 households

Demographics:

* Households predominantly belong to ‘age’ above ‘65+ and between ’45-54’
* Belong to occupation category ‘Professional or Technical’ and 'clerical’ and ‘Retired’
* Males belong to the age category ’45-54’

**Recommendations:**

1. Since most of them are professionals, send discounts/coupons on special timings such as weekends, holidays etc.
2. As a store manager of any retail chain, one can send promotions/discounts or advertisements to customers who belong to the age group ’45-54’ with a family size of 2 and who particularly own their houses and work as professional. This way we can see maximum profit and response rate to the promotions the company invested in and market share can be increased.

**MDC – Multinomial Discrete Choice Model**

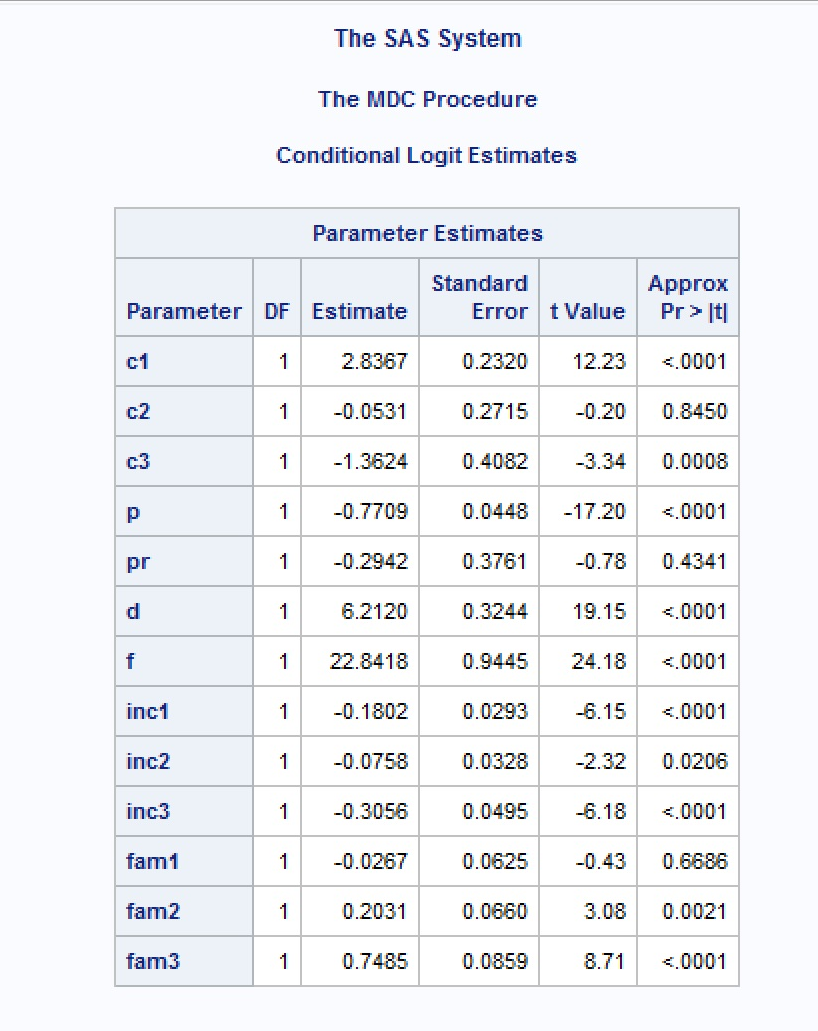
A conditional logit model was run after combining the product, grocery, panel and demographics data and choosing the top 4 brands in the panel data - MENTADENT PROCARE (brand choice 1), ORAL B ADVANTAGE (brand choice 2), BUTLER PROTECT (brand choice 3) and ORAL B INDICATOR (brand choice 4) for the final analysis.

The Log Likelihood of Null model = -3487

The Log Likelihood of model with covariates = -1930

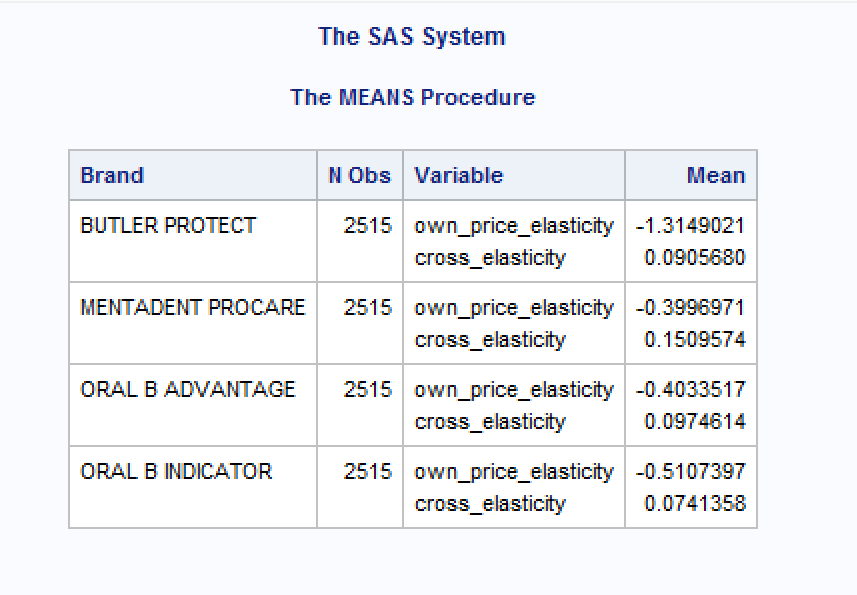
McFadden’s R2 = 44.65

This shows that the model with covariates performs better than the NULL model.



**MDC Insights:**

* Customers who purchase Tooth brush from grocery stores are price sensitive across brands. Since coefficient of Price is negative, with every unit increase in Price the probability of choice of each brand decreases.
* With every unit increase in Income of the household, the probability of choice of all 3 brands decreases when compared to brand 4.
* With every unit increase in family size of the household, the probability of choice for brand 2 and brand 3 increases when compared to brand 1 and brand 4, as brand 1 is insignificant.
* BUTLER PROTECT is most popular brand among bigger families. From this it can be inferred that bigger families tend to prefer buying the cheapest products.
* Feature and display both have a positive effect on the probability of a person picking up a brand.
* BUTLER PROTECT has the highest own price elasticity whereas MENTADENT PROCARE has the highest cross price elasticity.
* ORAL B Advantage and MENTADENT PROCARE has almost similar own price elasticity which is the lowest among the 4 brands.

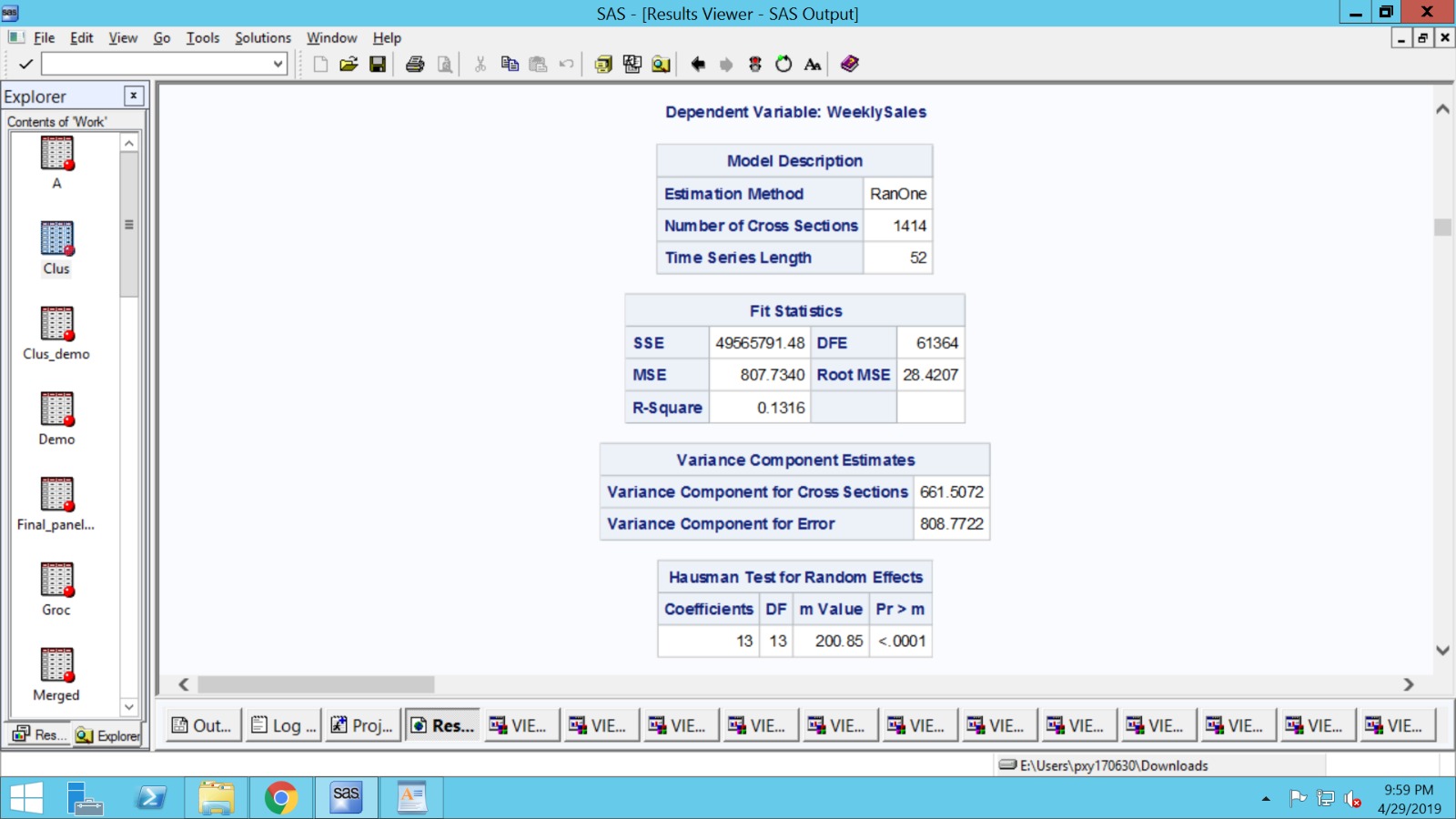
****

Recommendations:

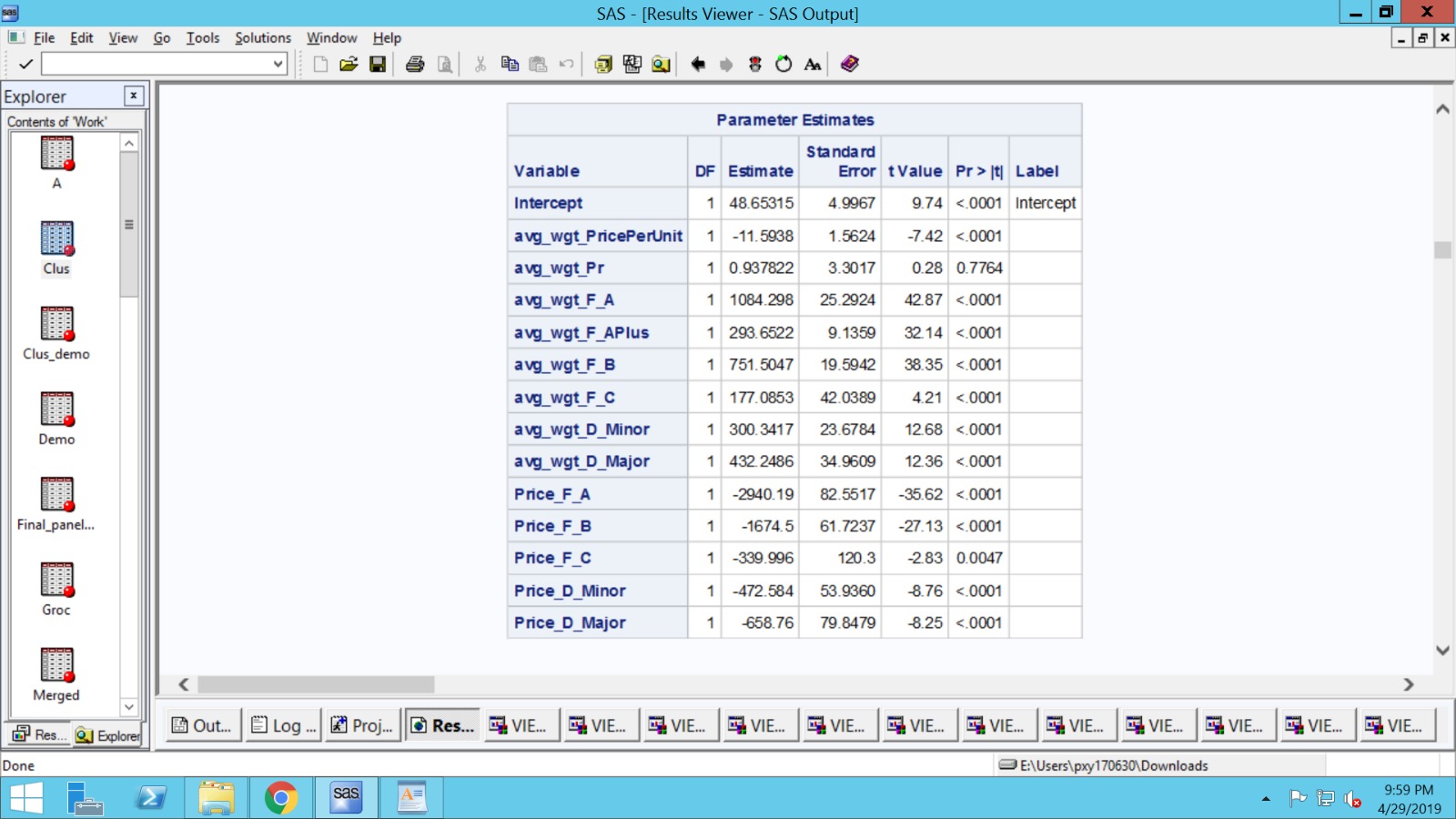
* **Focus on bigger families : BUTLER PROTECT** should focus more to target bigger families more to increase its market share and to also retain the existing customers, as it is most popular among bigger families.
* **Reduce price to further expand the customer base: The own price elasticity of ORAL B ADVANTAGE** brand is among the lowest hence it would be good idea to experiment by reducing its price to target the bigger families as bigger families appear to buy the cheaper products. Putting this kind of promotion on display might help the sales further.

**APPENDIX 1:**

Hausman test



Fixed one-way estimates:



Fixed two-way estimates:

