**Predictive Analytics Using SAS**

**Analysis on Laundry Scanner Data**

**Brand: ALL Detergent**

**Type: Liquid Detergent**

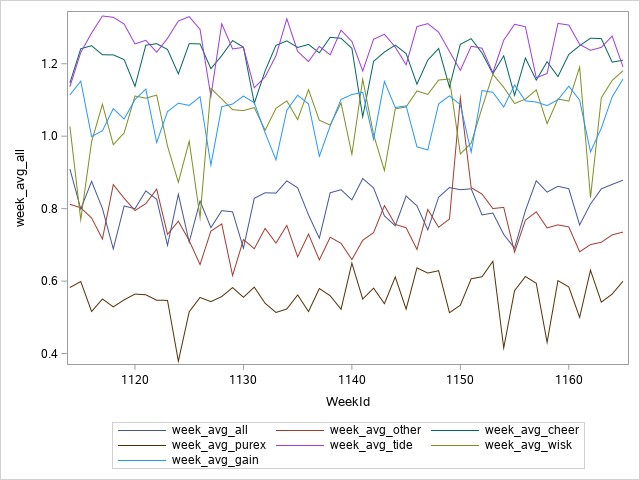
**ALL Liquid Detergent**

ALL is a brand owned by Unilever. It is among the top selling brands in the US with a market share of 10.2% in terms of dollar sales.

The top brands and their market share is shown in the table below:

\*\*\*\*\* diagrams here

\*\*\*demographic market share graphs



With data available from scanner data, we have designed some hypothesis to test and provide the answers to the following questions: -

1. What are the significant demographic features of the most valuable customers of ALL Liquid?
2. What is the effect of different characteristic features of the Store Panel data and what affects the sales?
3. What is the brand preference of different customers based on Store Panel Data and Panelist Demographic data?
4. What are the significant demographic features of the most valuable customers of ALL Liquid?

**Data Preparation:**

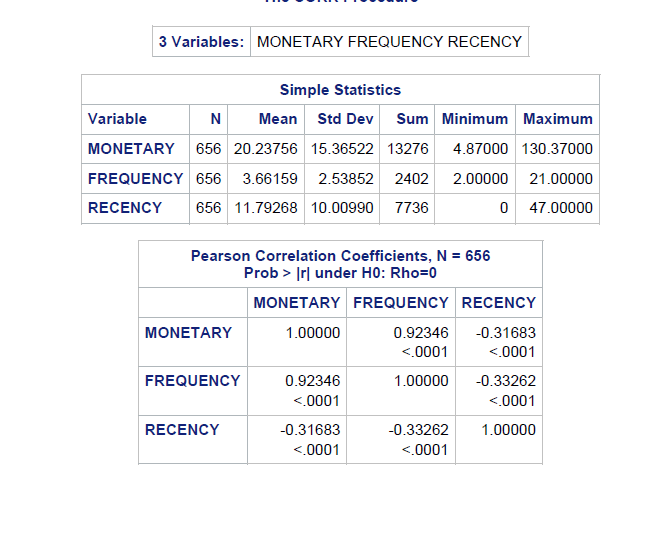
Panel Demographic data has all the data related to panelist, we have joined this data to the panel sales data to get the sales figure and also to the brand data to get the brand information.

We filter out the ALL Brand Liquid Detergent Category.

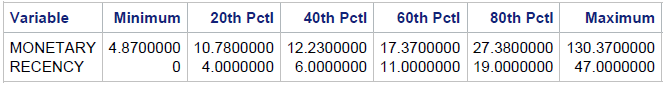
Analysis:

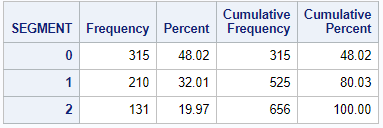
To segment the data we use RFM to build the bins where moentary is taken as the dollar sales, frequency as to the number of weeks customer made a purchase, and recency as to how recently he has made a purchase. We segment them into 3 segments of equal size and will do a analysis on the Top segment to get the demographic information.

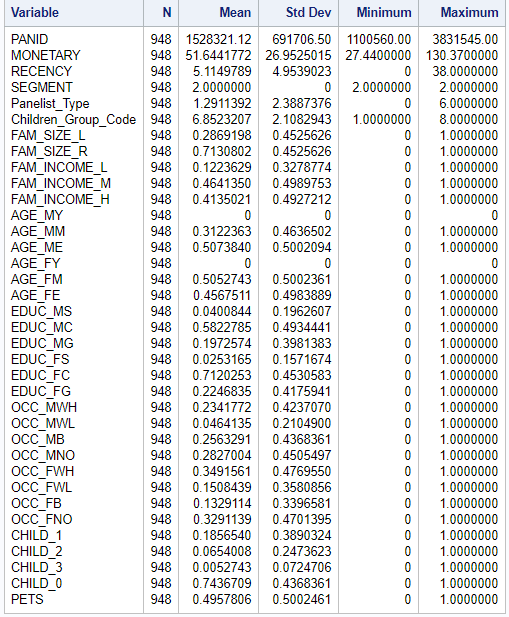
Correlation between RFM:



As the correlation between monetary and frequency is very high, we can remove any one of the 2 variables for the analysis further. And then run the segments over it.







We can drive the following statastics from the above data:

1. Family Size: Small family prefer our brand more than large families
2. Income Group: Medium and High income groups prefer our product over low income groups.
3. Male and Female Education: People who have graduated high school, or in some college prefer our brand more ie the college going students and the people who graduated high school and never attended college prefer our brand more.
4. Male Household Occupation: Clerical workers and craftsmen are less likely to buy our product.
5. Female Household Occupation: Females at professional or managerial level jobs and the females who are retired prefer our brand more than other type of occupation.
6. Child: Household with no children are more likely to buy our product. Almost 75% of the top demographic are the household with no children.

Recommendations:

This demographics will help us target similar customers which fit to the top 20% of our buyers. Special promotion and ad targetting could be devised for the customers with similar demographics.

So the households to be targetted will be:

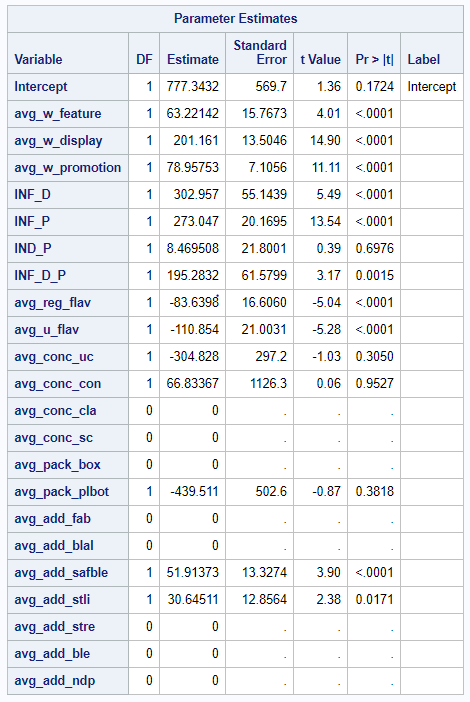
1. People with small family size should be targetted (Family Size < 4)
2. Currenlty ongoing college students and the people who have graduated high school but not continued should be targetted.
3. Household with no children should be targetted.
4. What is the effect of different characteristic features of the Store Panel data and what affects the sales?

Cross sectional data is grouped by stores and time series data is grouped by week

We are going to test if any heterogeneity exists in store purchase data over time and then check for effects of advertisements/promotions and product characteristics on total volume sold weekly based on the appropriate model

On performing Hausman test to check for random effects across the panel data, the p value came to be significant implying that the unobserved factors are time invariant and since there are no random effects the ideal regression model for this data would be fixed effects

**Interpretation of coefficients:**



We can clearly see that advertisement, features and promotion are significant along with observed heterogenic variables which are basically the interaction effects of feature and display, feature and price reduction, display and price reduction and have a huge impact on total volume sales.

As we can see from the output, most of the product characteristics are insignificant when it comes to predicting the total volume sold. Only two characteristics which are flavor/scent and additives have a significant impact on total sales.

**Weighted Feature:** when products are featured then the total number of units sold per week increases by 63

**Weighted Display:** when products are on display then the total number of units sold per week increases 201

**Weighted Price reduction:** when there is a price reduction on products then the total number of units sold per week increases by 79

**Feature and display interaction:** The variable is significant and is showing a synergistic effect due to combination of feature and display on total sales

**Feature and price reduction interaction**: The variable is significant and is showing a synergistic effect due to combination of feature and price reduction together on total sales

**Feature, display and promotion interactions:** The variable is significant and is showing a synergistic effect due to combination of feature, display and promotion on total sales

**Weighted regular flavor:** The total number of units sold in a week with regular flavor are 83 less than the citrus breeze flavor

**Weighted unscented flavor:** The total number of units sold in a week with unscented flavor are 110 less units than the citrus breeze flavor

**Weighted color safe bleach additive**: The total number of units sold in a week with additive as color safe bleach are 51 more than the units sold with additive as orange essence

**Weighted stain lifter bleach additive**: The total number of units sold in a week with additive as stain lifter bleach additive are 30 more units than the units sold with additive as orange essence.

**Recommendations:**

1. As we can see the largest contribution to sales volume is given by the synergistic effect of feature and display followed by the synergistic effect of feature and price reduction, so we would suggest the brand to feature a product and simultaneously put it on display instead of only featuring or just putting it on display. By doing both simultaneously will most likely boost sales by at least 101 units
2. Another recommendation that we could suggest would be to reduce the production of of products which are unscented because this is the worst performing flavor and reducing this would help save production cost.