**DIAPERS CASE STUDY**

**GROUP 6**

**BUAN6337.005- Predictive Analytics with SAS**

Leo Liu

Sandeep Chandra Nutakki

Sai Rohit Jammu

Toma Hasei

Yechen Hua



**Overview:**

The data of various brands with respect to diaper sales across the US was given and we performed various analysis on the data a to increase the sales of the brand (HUGGIES) with two different approaches. The data had different types of Brands and their sales in the drug stores, grocery stores, and bulk purchases. The demographics linking to the diaper sales and IRI week data was also present. The data analysis is done on grocery stores which is has the highest market share among all the channels.

Different approaches are studied and solutions to the problem statement were found by understanding the datasets, analyzing the panel datasets and demographic information of the customers, studied cross-price elasticity and effect of competitors, created RFM model to target certain customers and predictive model based on effect of advertisement, found interesting insights from the model and finally, provided recommendations based on the analysis.

**Step 1**

**ndvsm**

**Step 2**

**Step 3**

Customer Analysis

COMPETITOR ANALYSIS

MARKET ANALYSIS

CUSTOMER ANALYSIS

**Problem:**

Understanding of market analysis on the brand Huggies and effect of competitors on the brand as a Manager. To increase the sales by sales in 2002 by creating strategies of approach.

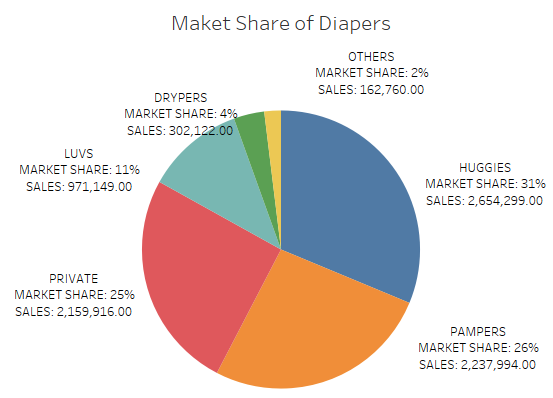
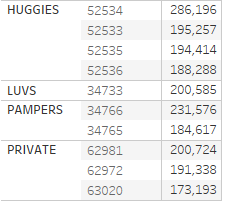
**Understanding the datasets:**

PANEL Data of the year “2001”, product data, delivery store data of three different channels were given. Demographic data of the panelists were also given. We used store data to understand the demographics of the store and transactional product data to understand the effect of features on the sales of the Brand “HUGGIES”. We used transactional data of all the brands and delivery store data to find the effect of advertisements on the brand. We calculated a new table to calculate the cross-price elasticity of the all the brands to find the effect on each other. We used the panel data to create an RFM model to target the most loyal customers.

**Analyzing the HUGGIES Market Performance:**

# Market share:

# Top 10 products which has the most sales(dollars)

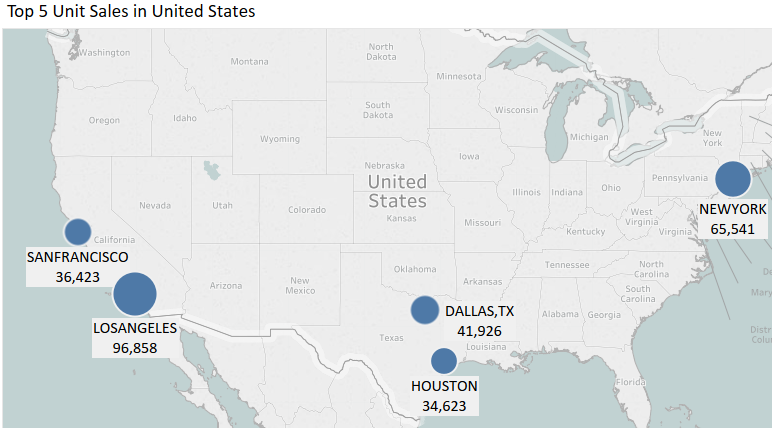
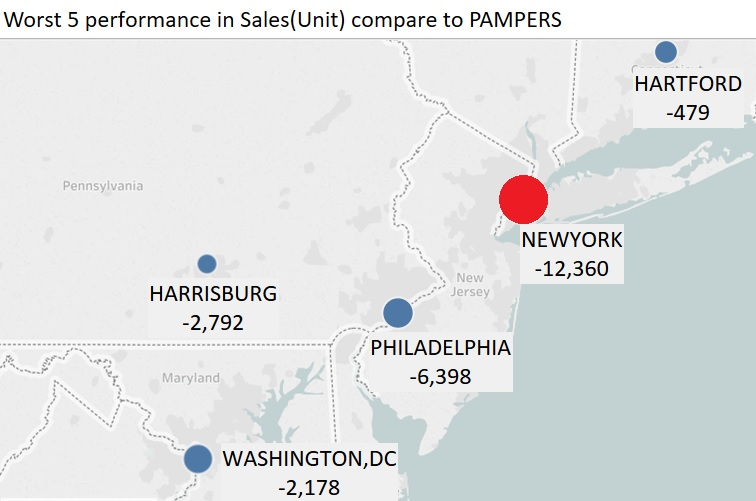
Top 10 products take up 24 % of the market

The Market Share(MS) of the brand “Huggies” is the first in the market. Pampers and Private have the similar share in the market. Among the top 10 products, 4 of them belongs to our company.

# 

Based on brands, PAMPERS and HUGGIES have the highest number of products and have more stores delivering those products.

**Location Analysis:**



East and west regions show a high number of transactions of sales on Huggies. On the east coast, we are losing the most customers, especially in NYC. New York City has second largest units of sales in the United States. But PAMPERS sold 12,360 more units of diapers than us in NYC. We still have big market growth potential in NYC.

**Effect of Advertisement on Sales:**

Different variables are created to find the effect of features on the sales of Huggies. The variables below are created and different models are modeled- Pooled OLS, fixed effects, and Random effects.

F\_AD = F\_A\*D

F\_Adisc = F\_A\*disc

F\_AplusD = F\_Aplus\*D

F\_Aplusdisc = F\_Aplus\*disc

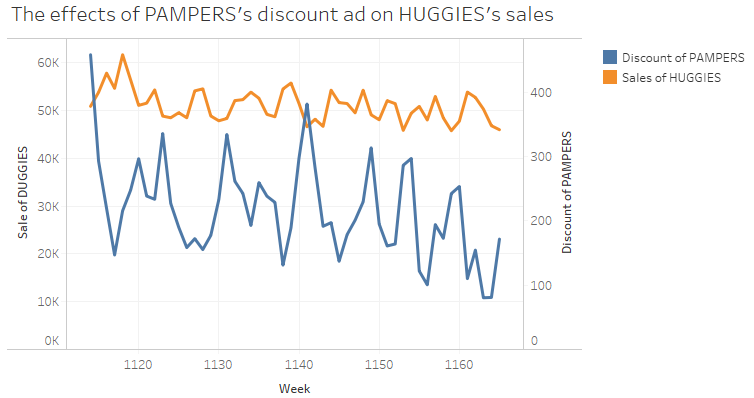
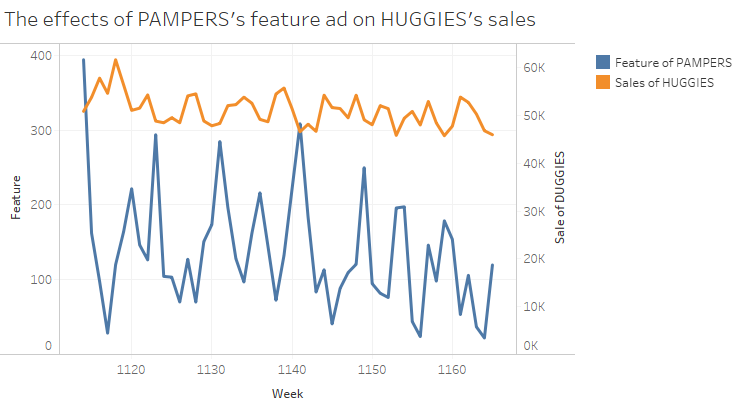
F\_BD = F\_B\*D;

F\_Bdisc = F\_B\*disc;

F\_Ddisc = D\*disc

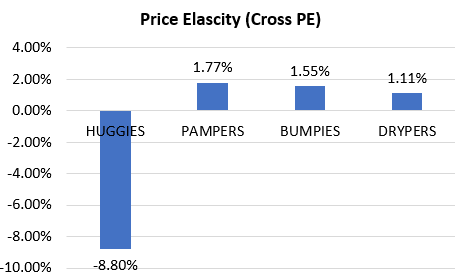
Pooled OLS model is not reliable which can have unobserved heterogeneity as it panel data, therefore Fixed effects are used with store\_n and week as fixed effects. Conducted Hausman test to find if there is endogeneity present and concluded that there is endogeneity present, so we decided to use the fixed effects with time and entity effects to understand the effect of advertising on the sales.

# **Competitors analysis of Star product (52534):**

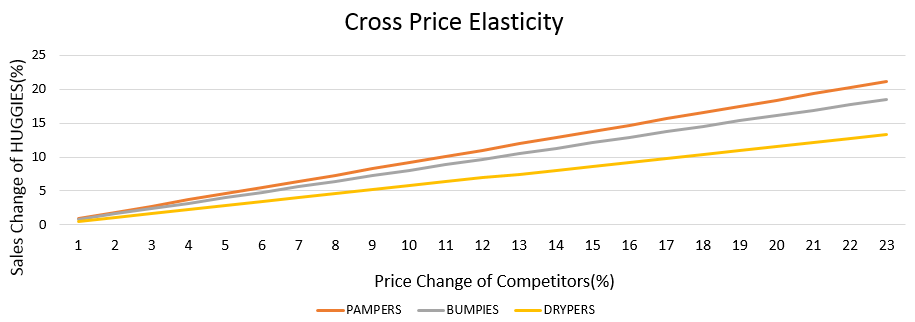


# With a change in feature and discount ads of Pampers, there is an effect on sales on Huggies. This shows that there is an effect of cross-price elasticity of the brands. We had 243,035 units of sales in New York City, but our competitor pampers had 376,565 units of sales greatly outperforming us in NYC market. Based on the increase of PAMPER’s feature ads and discount promotion, we found a decreasing trend in HUGGIES’ sales.

# **Company summary based on Cross Price Elasticity:**



A regression model is modeled to understand the cross-price elasticity on Huggies buys other brands. All the brands and their elasticity is calculated accordingly as the data is in time series and it needs to be DE trended. The data is grouped on brands and weeks and residual data is calculated. Linear regression is used to the model the data and following results were generated.



With a unit increase in Pampers price, there is 1.77 units increase in sales of Huggies. With the increase of Huggies price itself, there is a self-association of the decrease in sales by 8.80%. The following table explains the competitor feature of Huggies and Pampers. The line chart shows how the changes in prices of our competitors affect the sales of HUGGIES. It shows that the sales of HUGGIES are most sensitive to the price of PAMPERS. Decreasing the product price is the most effective way to increase sales. Increasing Feature\_A+ ad can also boost sales. Every 1 percent decrease in price will lead to 4.7 % increase in sales.

Comparison between Pampers and Huggies to understand the competitions and the difference in demographics. Most stores are common between both brands and average product price is also same between the brands.

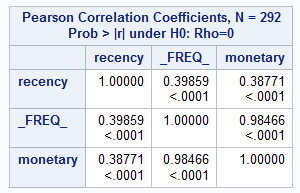
|  |  |  |
| --- | --- | --- |
| **Company name** | **HUGGIES** | **PAMPERS** |
| Product number | 98 | 101 |
| Product features | ULTRA TRIM | THIN, ASSORTED |
| user stage | Premature - NEW BORN – Stage 6 | NEW BORN - STAGE 6 |
| User gender | BOY OR GIRL | BOY OR GIRL |
| Number of Store covered | 1431 | 1430 |
| Product average price | 0.31 | 0.304 |

**RFM Analysis for customer targeting:**

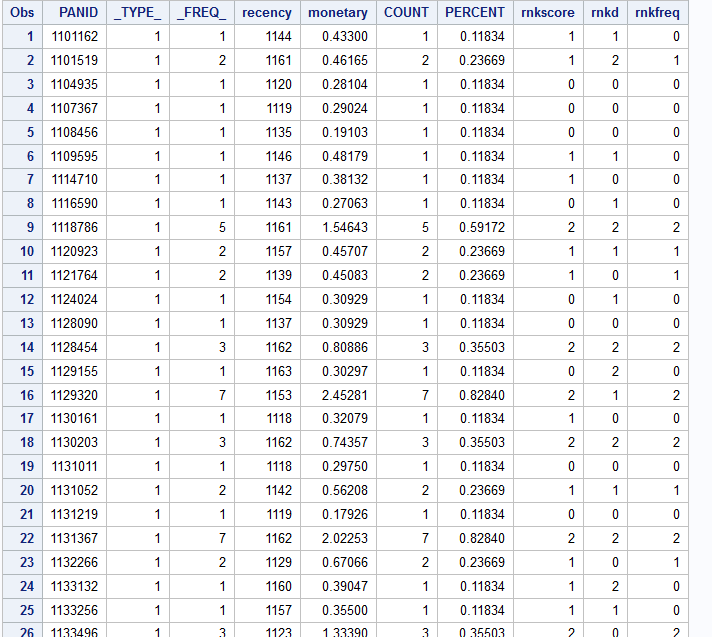
We performed RFM analysis on the dataset to segment our customers based on RFM scores and to identify those customers who are likely to respond to our campaigns and promotions.

The panel data for grocery stores was combined with the product data to identify the brands chosen by different panelists.

We then selected only those panelists who purchased “HUGGIES” diapers during their visits. We then sorted these customers by their most recent purchase date, number of visits and the total dollars they spent on HUGGIES brand during their visits. PROC RANK was used to rank these customers on three variables: Recency, Frequency and Monetary with 2 being the highest and 0 being the lowest. RFM score for each customer is then obtained by combining these three variables.

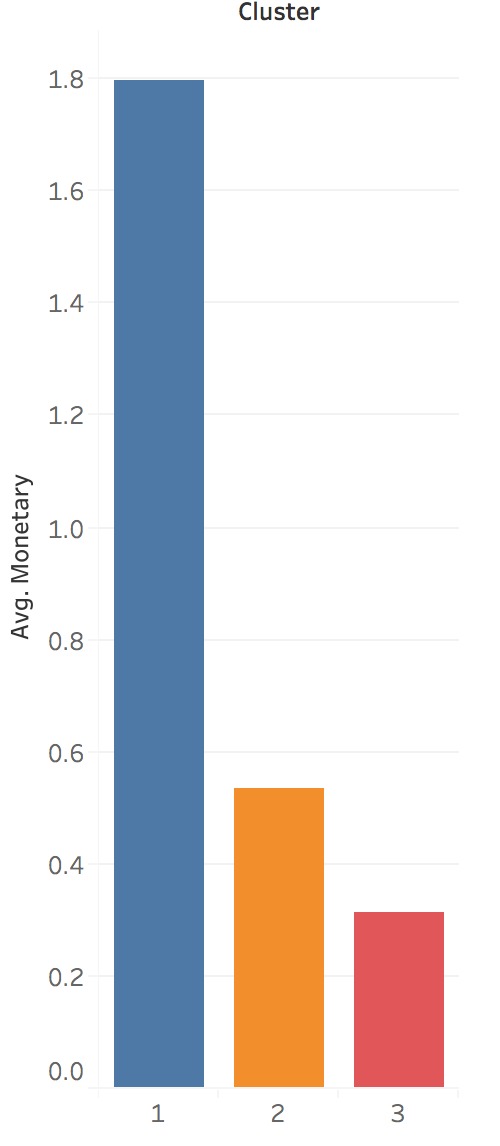


After sorting and ranking customers based on recency, frequency and monetary attributes, we ran a correlation matrix between recency frequency and monetary values which shows us that there is high correlation between frequency and monetary values for the households. So, because of the high correlation we can reduce the differentiating dimensions from three to two (Recency, Monetary).

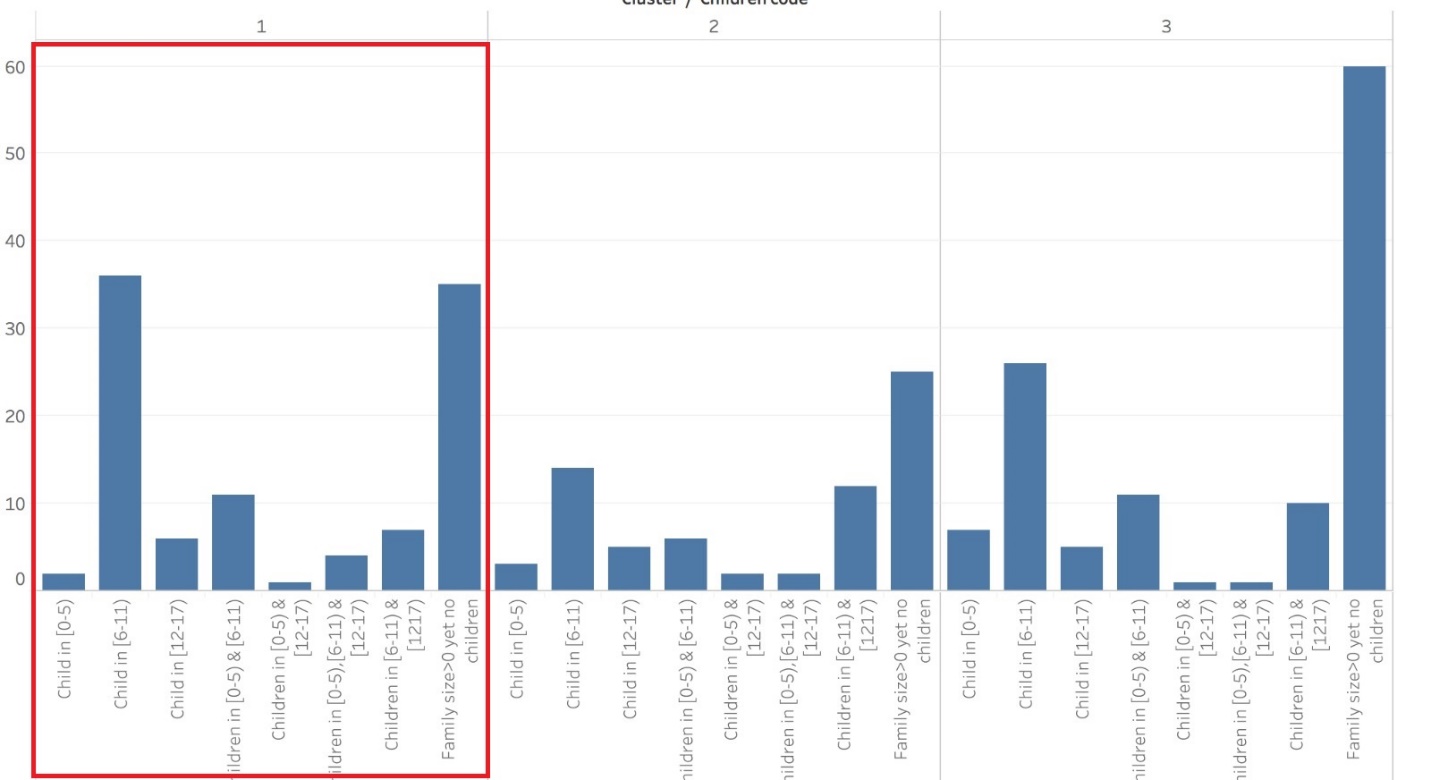


RFM Scores Dataset

From our analysis, we observed that cluster 3 is high valued cluster as it consists of panelists who spend more on HUGGIES frequently. Cluster 1 is of least importance to us as it has customers who spend less on our brand.

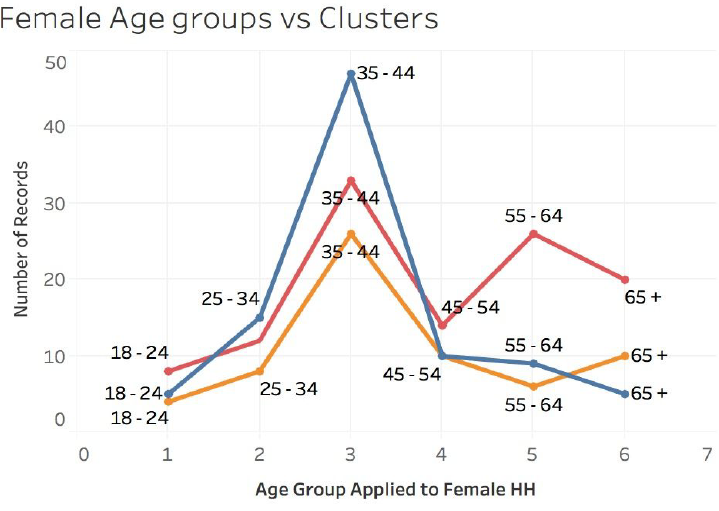


Average price per unit diaper spent per customers in different clusters



Child age in families grouped in clusters

As we can look at the above bar chart, we can see that in cluster1(loyal customers) the most families either are expecting their first child or having their child between 6 -11.



Female age groups vs Clusters

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

Moreover, the most females in the cluster1(orange line) are between ages 25 -44 yrs. Combining these clusters with demographics, we can now profile our customers.

CLUSTER 1- Loyal Customers

Differentiating Characteristics:

* High R, F and M scores.
* Cluster size: 102 households

Demographics-

* High Income
* Male households in the age group of 25-54.
* Female households in the age group of 25-44.
* Female in households-professional jobs, clerical jobs, part-time or full-time workers.
* Males in Households- full time workers, homemakers.
* Females in “Professional” field are higher than in other clusters
* Males in “Full Time” Jobs are higher than in other clusters

CLUSTER 2-Potentially Profitable Customers

Differentiating Characteristics:

* High M score and Low R score or Low M score and High R score or average R and M scores
* Cluster size: 69 households

Demographics-

* High Income
* Male households in the age group of 25-54 years
* Female households in the age group of 35-54 years.
* Male in Households- full time workers, retired
* Female in Households- full-time, part time workers; managerial jobs and retired

CLUSTER 3 – Less Profitable customers

Differentiating Characteristics:

* Low R and M scores.
* Cluster size: 121 households

Demographics-

* Low Income
* Male households in the age group of 35-65+ years
* Female households in the age group of 35-65+ years.
* Males in Retirement are higher than in other clusters.
* Females in “Full-time “and” Part-time” Jobs are higher than in other clusters.

Based on the characteristics of these clusters we can now develop strategies accordingly to better target these customers.

**Recommendations:**

* To compete with our competitor, we need to focus mainly on NYC market.
* We can close the sales gap between HUGGIES and PAMPERS in NYC market by:
  + - * + Decreasing the product price in NYC stores as Pampers sales are highly co-related with HUGGIES price, which can result in customers switching brand to ours.
        + Increasing Feature A+ ads in NYC stores because this might be the reason PAMPERS has high sales in NYC (pampers uses A+ ads in NYC).
        + Launch Thin and Assorted diapers as this model in PAMPERS diapers has most sales in NYC.
* Discount timing special for them -Making Coupon accessible to professionals’ timings (e.g. weekends, holidays etc).
* Sending out Customized flyers tocustomers who have full time and part time jobs in professional and managerial fields.
* Reach out to the customers in their preferred line of communication i.e., social media platforms because the huge portion of our loyal customers are young working professionals.
* Develop strategies to encourage customers to do repeat purchase as many customers in cluster 1 have low frequency.
* Create “value packs” or “economy packs” to attract the sales of the customers in cluster3(less profitable customers).