

BUAN 6337.502- Marketing Predictive Analytics Using SAS- S17



HUGGIES MARKET ANALYSIS PROJECT REPORT



GROUP 10

TEAM MEMBERS

Harish Rao Pallepati

Nimesh Mohandas

Santosh Shankar

Nakul Pyasi

Viraj Shah

Table of Contents

1. INTRODUCTION.....	3
2. OBJECTIVE	3
3. STRATEGY.....	3
4. EXPLORATORY DATA ANALYSIS.....	3
5. EXECUTIVE SALES REPORT	4
6. ANALYSIS ON PRICE ELASTICITY	5
6.1 Data Preparation.....	5
6.2 SAS Results	6
7. BRAND CHOICE PREFERENCE.....	13
7.1 Store Feature Study	13
7.1.1 Data Preparation.....	13
7.1.2 SAS Results	13
7.1.3 Insights	17
7.2 Demographic Study.....	17
7.2.1 Data Preparation.....	17
7.2.2 SAS Results	18
7.2.3 Insights	22
8. MONETARY SEGMENTATION using RFM	22
8.1 RFM Study	22
8.2 RFM Results	23
8.3 Insights	25
9. RECOMMENDATION	28
10. APPENDIX.....	29

1. INTRODUCTION

Huggies is one of the largest and most recognized brand in the diaper industry. The analysis is made on the weekly sales incurred for Huggies against its competitors for the year 2001 across different markets of United States. The data comprises information on weekly store level sales, customer panel data, demographics of every customer and descriptive information about markets.

2. OBJECTIVE

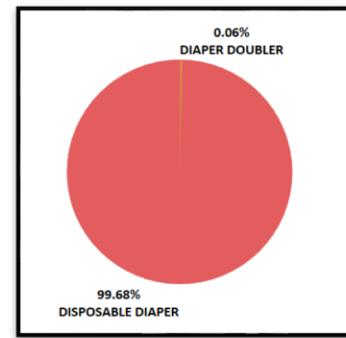
To find insights from data that would assist the brand manager of Huggies in making decisive actions to increase their market share for the upcoming year. We have store-level and customer level sales information for the year 2001 across each week. Using exploratory analysis and statistical concepts and procedures, we study the customer demographics and the market scenario of diaper industry.

1. To find out the market share of other competitors and study how demographics and store-promotions affect brand choice
2. To figure out the most profitable and highly frequent buyers using RFM method and understand their demographic characteristics to gain insights
3. To find the self-price elasticity and cross-price elasticity of Huggies with respect to the sales and pricing of other leading brands to gain insights on market pricing
4. To provide recommendations from the insights on what type of groups need to be targeted or which markets needs changes to increase sales and capture brand recognition in less profitable markets

3. STRATEGY

To study the collected data and gain insights out of it, we have followed the following techniques to come out with recommendations for improvement in market share. People generally prefer '**Disposable diapers**' and the exploratory market analysis confirms the same (around 90% of market sales), therefore the whole analysis is made on this product type alone.

1. **Price Elasticity Study** – Propose strategy for store level pricing and promotions
2. **Monetary Segmentation through RFM** - Understand demographics of customer who have bought our products to target promotions on specific group of people
3. **Logistic Regression** – Study Brand choice preference of customers over competitors through store level and customer demographic study



4. EXPLORATORY DATA ANALYSIS

To gain a better understanding of the data and to delineate our area of focus, we have done an initial study on the number of observations, sales and revenue of Huggies along with Luvs, Pampers and Other brands (all fringe brands grouped together)

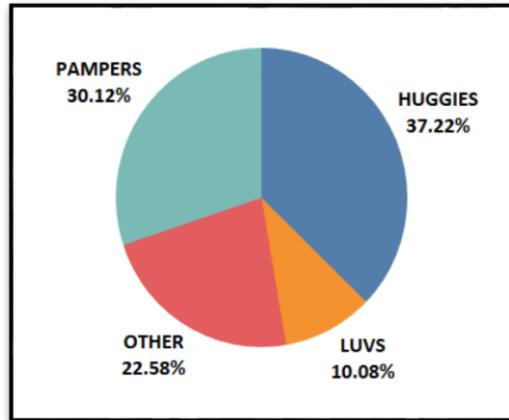
Brand		Brand				
	Metrics	HUGGIES	LUVS	PAMPERS	OTHER	TOTAL
Store	Dollars	\$40,559,415.51	\$10,984,935.73	\$34,767,813.79	\$23,349,023.99	\$109,661,189.02
	Units	2,978,403	1,001,459	2,483,350	2,832,700	9,295,912
	Number of Records	1,099,744	359,375	891,195	915,628	3,265,942
Panel	Dollars	\$14,659.90	\$10,545.24	\$12,998.47	\$14,561.84	\$52,765.45
	Units	1,082	976	950	1,840	4,848
	Number of Records	1,042	939	902	1,547	4,430
	Distinct Panel ID	346	299	332	407	946

5. EXECUTIVE SALES REPORT

A. Market Share

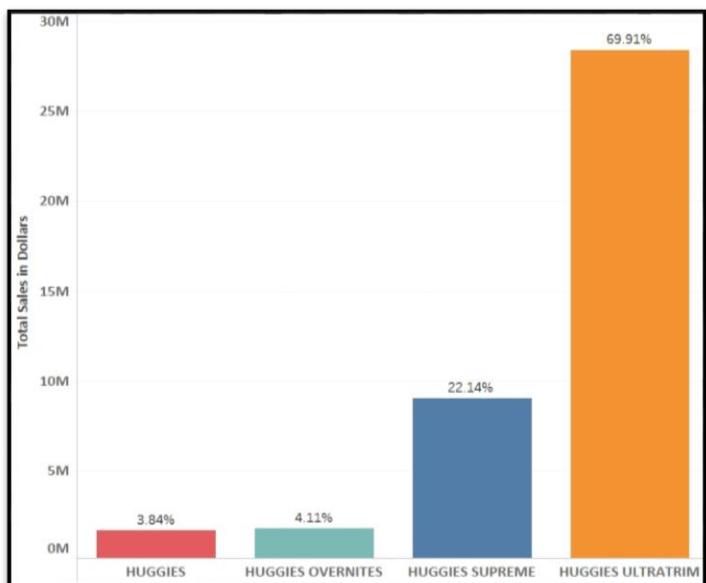
Clearly, Huggies is the overall market leader for current year in comparison with their competitors Pampers, Luvs and Other (all other brands in dataset excluding Huggies, Pampers and Luvs).

Huggies should be more watchful about Pampers on deciding about prices, promotions and advertisements of diapers for the next year.



B. Affinity towards Huggies Variants

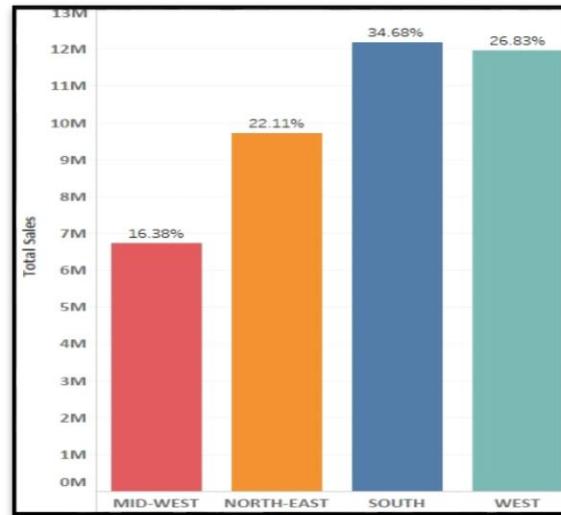
Huggies UltraTrim is the most preferred products followed by Huggies Supreme. From the graph, both together contribute to 88.25% of total sales across all variants of Huggies.



C. Region-wise Revenue

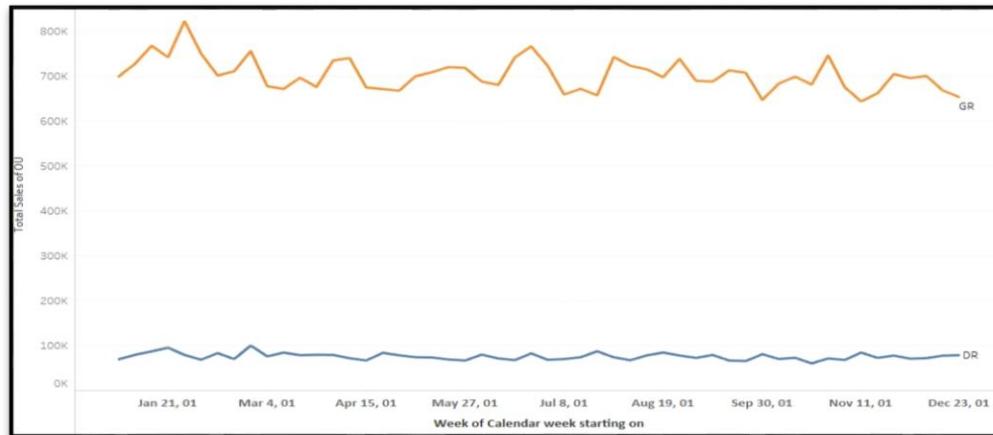
Strong market presence in the southern regions (35%) of United States. Followed by 27% in the Western regions.

Though the revenue is on the lower side in the Mid-West regions but still Huggies is the market leader in Detroit, Cleveland, Grand Rapids compared to all other brands.



D. Weekly Sales Trend

There is a downward trend in sale of products for the year 2001. From the below chart, we can infer that the sales have gradually declined over the months this can be due to declining birth rate in the United States.



6. ANALYSIS ON PRICE ELASTICITY

To devise an effective pricing structure, it is important to study the effects of price on sales in relation with the competitors.

6.1 Data Preparation

Since, the diapers come in different pack-sizes, we will need to first get Per Diaper Price. And since there are several products under one diaper-brand, to compare the prices of one brand with other, it is important to convert the prices of each product to a weighted price and get a brand-level total of weighted price for a week and store.

Steps to get the total weighted price of a brand:

1. The store-level data from Grocery and Drug Store are clubbed into a single data extract.
2. All store level data are grouped into four brands – Huggies, Pampers, Luvs and Others (Rest of the Brands).
3. Weighted Price for a product is calculated for a store for a particular week (i):

$$WeightedPrice_i = \sum Price\ per\ Diaper_i * \frac{Sales\ of\ Diaper_i}{Total\ Sales\ of\ the\ Brand_i}$$

Similarly, weighted display, feature and price reduction score are calculated.

4. Now we will have one observation for each brand for each store for every week that the store was open.
5. The data is converted from long-form to wide-form, where we will have single observation for every week when the store is open, and the price and store-promotion details are stored in columns.

6.2 SAS Results

The data is now arranged as a Panel Data with store-wise details for all the weeks. We apply a simple regression (PROC PANEL) over the panel-data with IRI_KEY as Cross-section ID and Week as Time-series ID, with weighted-price and weighted-features (Feature, Display, and price-reduction) as independent variables and Total-sales of the brand as the dependent variable.

Hausman Test:

H_0 : correlation between error term U_i and X variables

H_1 : No correlation between error term U_i and X variables

Hausman Test for Random Effects			
Coefficients	DF	m Value	Pr > m
28	28	858.67	<.0001

Since, the p-value is less than 0.05(confidence-level at 95%), so we reject null hypothesis and confirm that Fixed-Effects model is best model for this dataset.

The SAS System			
The PANEL Procedure Fixed Two-Way Estimates			
Dependent Variable: tot_units			
Model Description			
Estimation Method		FixTwo	
Number of Cross Sections		772	
Time Series Length		52	
Fit Statistics			
SSE	7739585.305	DFE	35033
MSE	220.9227	Root MSE	14.8635
R-Square	0.8756		
F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
822	35033	276.74	<.0001

HUGGIES

The SAS System			
The PANEL Procedure Fixed Two-Way Estimates			
Dependent Variable: tot_units			
Model Description			
Estimation Method		FixTwo	
Number of Cross Sections		772	
Time Series Length		52	
Fit Statistics			
SSE	4878708.418	DFE	35033
MSE	139.2604	Root MSE	11.8009
R-Square	0.6882		
F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
822	35033	74.39	<.0001

LUVS

The SAS System			
The PANEL Procedure Fixed Two-Way Estimates			
Dependent Variable: tot_units			
Model Description			
Estimation Method		FixTwo	
Number of Cross Sections		772	
Time Series Length		52	
Fit Statistics			
SSE	7146184.926	DFE	35033
MSE	203.9844	Root MSE	14.2823
R-Square	0.8622		
F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
822	35033	235.53	<.0001

PAMPERS

The SAS System			
The PANEL Procedure Fixed Two-Way Estimates			
Dependent Variable: tot_units			
Model Description			
Estimation Method		FixTwo	
Number of Cross Sections		772	
Time Series Length		52	
Fit Statistics			
SSE	12386267.47	DFE	35033
MSE	353.5600	Root MSE	18.8032
R-Square	0.8130		
F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
822	35033	160.62	<.0001

OTHER BRANDS

BETA Estimates:

Regression equation:

$$\begin{aligned}
 Sales_{HUGGIES} = & \beta_0 + \beta_1 Price_{HUGGIES} + \beta_2 Price_{LUVS} + \beta_3 Price_{PAMPERS} + \beta_4 Price_{OTHER} \\
 & + \beta_5 Display_{HUGGIES} + \beta_6 Display_{LUVS} + \beta_7 Display_{PAMPERS} + \beta_8 Display_{OTHER} \\
 & + \beta_9 Feature_{HUGGIES} + \beta_{10} Feature_{LUVS} + \beta_{11} Feature_{PAMPERS} \\
 & + \beta_{12} Feature_{OTHER} + \beta_{13} PR_{HUGGIES} + \beta_{14} PR_{LUVS} + \beta_{15} PR_{PAMPERS} \\
 & + \beta_{16} PR_{OTHER} + \beta_{17} Price_{HUGGIES} * PR_{HUGGIES} + \beta_{18} Price_{LUVS} * PR_{LUVS} \\
 & + \beta_{19} Price_{PAMPERS} * PR_{PAMPERS} + \beta_{20} Price_{OTHER} * PR_{OTHER} + \beta_{21} Price_{HUGGIES} \\
 & * Feature_{HUGGIES} + \beta_{22} Price_{LUVS} * Feature_{LUVS} + \beta_{23} Price_{PAMPERS} \\
 & * Feature_{PAMPERS} + \beta_{24} Price_{OTHER} * Feature_{OTHER} + \beta_{25} PR_{HUGGIES} \\
 & * Feature_{HUGGIES} + \beta_{26} PR_{LUVS} * Feature_{LUVS} + \beta_{27} PR_{PAMPERS} \\
 & * Feature_{PAMPERS} + \beta_{28} PR_{OTHER} * Feature_{OTHER}
 \end{aligned}$$

Self-Price Elasticity: A measure used in economics to show the responsiveness, or elasticity, of the quantity demanded of a good or service to a change in its price when nothing but the price changes.

$$SelfPrice\ Elasticity = \frac{\%Change\ in\ Sales_{HUGGIES}}{\%Change\ in\ Price_{HUGGIES}}$$

which can be written as,

$$SelfPrice\ Elasticity = \frac{\Delta Sales_{HUGGIES}}{Sales_{HUGGIES}} * \frac{Price_{HUGGIES}}{\Delta Price_{HUGGIES}} \approx \frac{d Sales_{HUGGIES}}{d Price_{HUGGIES}} * \frac{Price_{HUGGIES}}{Sales_{HUGGIES}}$$

Therefore,

$$SelfPrice\ Elasticity\ of\ HUGGIES = (\beta_1 + \beta_{17} PR_{HUGGIES} + \beta_{21} Feature_{HUGGIES}) * \frac{Price_{HUGGIES}}{Sales_{HUGGIES}}$$

Cross-Price Elasticity: A measures the responsiveness of the quantity demanded for a good to a change in the price of another good.

$$CrossPrice\ Elasticity = \frac{\%Change\ in\ Sales_{HUGGIES}}{\%Change\ in\ Price_{LUVS}}$$

which can be written as,

$$CrossPrice\ Elasticity = \frac{\Delta Sales_{HUGGIES}}{Sales_{HUGGIES}} * \frac{Price_{LUVS}}{\Delta Price_{LUVS}} \approx \frac{d Sales_{HUGGIES}}{d Price_{LUVS}} * \frac{Price_{LUVS}}{Sales_{HUGGIES}}$$

Therefore,

$$CrossPrice\ Elasticity\ of\ Brand1 = (\beta_2 + \beta_{18} PR_{LUVS} + \beta_{22} Feature_{LUVS}) * \frac{Price_{LUVS}}{Sales_{HUGGIES}}$$

Following are the β -estimates for the regression:

HUGGIES						
	Variable	Estimate	Standard Error	t Value	Pr > t	Mean
b0	Intercept	31.24	3.67	8.52	<.0001	
b1	wt_price_brand1	-5.98	5.67	-1.06	0.2912	0.31
b2	wt_price_brand2	-25.91	5.45	-4.75	<.0001	0.26
b3	wt_price_brand3	0.99	4.86	0.2	0.8393	0.31
b4	wt_price_brand4	1.58	5.72	0.28	0.7826	0.22
b5	disp_wt_brand1	29.42	1.23	23.9	<.0001	0.02
b6	disp_wt_brand2	0.72	1.04	0.69	0.4892	0.01
b7	disp_wt_brand3	-4.25	1.06	-4.02	<.0001	0.02
b8	disp_wt_brand4	-3.79	0.94	-4.03	<.0001	0.03
b9	Feature_wt_brand1	-4.50	5.10	-0.88	0.3777	0.13
b10	Feature_wt_brand2	-2.65	4.88	-0.54	0.5868	0.06
b11	Feature_wt_brand3	12.27	4.34	2.83	0.0047	0.11
b12	Feature_wt_brand4	-0.42	0.52	-0.8	0.421	0.10
b13	PR_wt_brand1	92.07	5.31	17.34	<.0001	0.20
b14	PR_wt_brand2	-0.16	0.38	-0.43	0.6637	0.16
b15	PR_wt_brand3	-3.98	0.52	-7.63	<.0001	0.17
b16	PR_wt_brand4	-1.07	0.49	-2.19	0.0285	0.21
b17	price_PR1	-360.39	14.36	-25.1	<.0001	0.06
b18	price_PR2	119.01	10.73	11.09	<.0001	0.05
b19	price_PR3	22.45	9.24	2.43	0.0151	0.06
b20	price_PR4	-35.97	10.38	-3.46	0.0005	0.05
b21	price_F1	63.01	16.44	3.83	0.0001	0.04
b22	price_F2	-1.95	18.57	-0.11	0.9163	0.01
b23	price_F3	-41.55	11.87	-3.5	0.0005	0.03
b24	price_F4	-1.36	12.26	-0.11	0.9116	0.03
b25	PR_F1	-5.61	1.27	-4.43	<.0001	0.08
b26	PR_F2	1.22	1.52	0.8	0.4227	0.04
b27	PR_F3	0.58	1.21	0.48	0.6307	0.07
b28	PR_F4	2.40	1.31	1.84	0.0656	0.02

mean Sales	
tot_units HUGGIES	43.86
tot_units LUVS	19.75
tot_units PAMPERS	38.58
tot_units OTHER	45.07

Self Price Elasticity	
if PR	-2.561
if Feature and PR	-2.120
if no PR and no Feature	-0.042

Cross Price Elasticity with LUVS	
if PR	0.554
if Feature and PR	0.542

Cross Price Elasticity with PAMPERS	
if PR	0.164
if no PR and no Feature	0.007

Cross Price Elasticity with OTHER	
if Feature	0.001
if no PR and no Feature	0.008

From the results we observe that if HUGGIES does a reduction in price by 1% then there will be 0.042% increase in its sales. But if it does a reduction in price alongside a discount sale, for every 1% reduction in price, the sales will increase by 2.56%.

On the other hand, if LUVS plans to do a discount sale, then with every 1% reduction of LUVS price, there will be a 0.55% decrease in sales of HUGGIES diapers.

And if Pampers plans to do a discount sale, then with every 1% reduction of Pampers price, there will be a 0.164% decrease in sales of HUGGIES diapers.

LUVS						
	Variable	Estimate	Standard Error	t Value	Pr > t	Mean
b0	Intercept	23.60	2.91	8.11	<.0001	
b1	wt_price_brand1	9.61	4.50	2.13	0.0328	0.31
b2	wt_price_brand2	-54.75	4.33	-12.64	<.0001	0.26
b3	wt_price_brand3	-3.83	3.86	-0.99	0.3208	0.31
b4	wt_price_brand4	-14.14	4.54	-3.11	0.0018	0.22
b5	disp_wt_brand1	-1.62	0.98	-1.65	0.0984	0.02
b6	disp_wt_brand2	20.36	0.83	24.59	<.0001	0.01
b7	disp_wt_brand3	0.01	0.84	0.01	0.9924	0.02
b8	disp_wt_brand4	-0.31	0.75	-0.42	0.6773	0.03
b9	Feature_wt_brand1	8.44	4.05	2.08	0.0372	0.13
b10	Feature_wt_brand2	195.50	3.88	50.45	<.0001	0.06
b11	Feature_wt_brand3	-8.53	3.44	-2.48	0.0132	0.11
b12	Feature_wt_brand4	-0.63	0.41	-1.51	0.1298	0.10
b13	PR_wt_brand1	-25.57	4.21	-6.07	<.0001	0.20
b14	PR_wt_brand2	3.44	0.30	11.43	<.0001	0.16
b15	PR_wt_brand3	-1.57	0.41	-3.79	0.0001	0.17
b16	PR_wt_brand4	-1.37	0.39	-3.53	0.0004	0.21
b17	price_PR1	17.41	11.40	1.53	0.1267	0.06
b18	price_PR2	63.00	8.52	7.4	<.0001	0.05
b19	price_PR3	14.45	7.33	1.97	0.0487	0.06
b20	price_PR4	-9.09	8.24	-1.1	0.2699	0.05
b21	price_F1	-34.07	13.05	-2.61	0.009	0.04
b22	price_F2	-685.53	14.74	-46.51	<.0001	0.01
b23	price_F3	29.06	9.43	3.08	0.0021	0.03
b24	price_F4	-4.07	9.74	-0.42	0.6758	0.03
b25	PR_F1	0.93	1.01	0.92	0.3572	0.08
b26	PR_F2	-23.57	1.21	-19.53	<.0001	0.04
b27	PR_F3	0.99	0.96	1.03	0.3046	0.07
b28	PR_F4	-0.21	1.04	-0.2	0.843	0.02

mean Sales	
tot_units HUGGIES	43.86
tot_units LUVS	19.75
tot_units PAMPERS	38.58
tot_units OTHER	45.07

Self Price Elasticity	
if Feature	-9.774
if Feature and PR	-8.942
if no PR and no Feature	-0.723

Cross Price Elasticity with HUGGIES	
if PR	0.419
if no PR and no Feature	0.149

Cross Price Elasticity with PAMPERS	
if PR	0.165
if Feature	0.392
if PR and Feature	0.616

From the above results we observe that if LUVS does a reduction in price by 1% then there will be 0.72% increase in its sales. But if it does a reduction in price alongside a feature-display and discount sale, for every 1% reduction in price, its sales will increase by 9.77% and 8.92% respectively.

On the other hand, if HUGGIES plans to do a discount sale, then with every 1% reduction of HUGGIES price, there will be a 0.42% decrease in sales of LUVS diapers.

PAMPERS						
	Variable	Estimate	Standard Error	t Value	Pr > t	Mean
b0	Intercept	26.78	3.52	7.6	<.0001	
b1	wt_price_brand1	13.95	5.45	2.56	0.0104	0.31
b2	wt_price_brand2	11.60	5.24	2.21	0.0269	0.26
b3	wt_price_brand3	-60.79	4.67	-13.01	<.0001	0.31
b4	wt_price_brand4	-16.46	5.49	-3	0.0027	0.22
b5	disp_wt_brand1	-1.81	1.18	-1.53	0.1255	0.02
b6	disp_wt_brand2	-1.27	1.00	-1.26	0.2068	0.01
b7	disp_wt_brand3	32.00	1.01	31.54	<.0001	0.02
b8	disp_wt_brand4	-3.04	0.90	-3.36	0.0008	0.03
b9	Feature_wt_brand1	9.14	4.90	1.86	0.0623	0.13
b10	Feature_wt_brand2	1.05	4.69	0.22	0.8234	0.06
b11	Feature_wt_brand3	68.46	4.17	16.43	<.0001	0.11
b12	Feature_wt_brand4	-2.52	0.50	-5.05	<.0001	0.10
b13	PR_wt_brand1	4.83	5.10	0.95	0.3435	0.20
b14	PR_wt_brand2	-2.37	0.36	-6.5	<.0001	0.16
b15	PR_wt_brand3	11.95	0.50	23.86	<.0001	0.17
b16	PR_wt_brand4	-0.95	0.47	-2.02	0.0433	0.21
b17	price_PR1	-13.40	13.79	-0.97	0.3313	0.06
b18	price_PR2	-53.25	10.31	-5.17	<.0001	0.05
b19	price_PR3	42.83	8.87	4.83	<.0001	0.06
b20	price_PR4	-11.75	9.98	-1.18	0.239	0.05
b21	price_F1	-35.71	15.79	-2.26	0.0238	0.04
b22	price_F2	-14.46	17.84	-0.81	0.4176	0.01
b23	price_F3	-191.33	11.41	-16.77	<.0001	0.03
b24	price_F4	8.66	11.78	0.73	0.4625	0.03
b25	PR_F1	0.44	1.22	0.36	0.7191	0.08
b26	PR_F2	1.35	1.46	0.92	0.3566	0.04
b27	PR_F3	-2.78	1.16	-2.39	0.0167	0.07
b28	PR_F4	2.34	1.25	1.87	0.0617	0.02

mean Sales	
tot_units HUGGIES	43.86
tot_units LUVS	19.75
tot_units PAMPERS	38.58
tot_units OTHER	45.07

Self Price Elasticity	
if Feature	-2.005
if PR	-0.143
if Feature and PR	-1.665
if no PR and no Feature	-0.483

Cross Price Elasticity with HUGGIES	
if PR	0.004
if no PR and no Feature	0.111

Cross Price Elasticity with LUVS	
if no PR and no Feature	0.078

From the above results we observe that if Pampers does a reduction in price by 1% then there will be 0.48% increase in its sales. But if it does a reduction in price alongside a feature-display, for every 1% reduction in price, its sales will increase by 2%.

On the other hand, if HUGGIES plans to do a discount sale, then with every 1% reduction of HUGGIES price, there will be a 0.11% decrease in sales of Pampers diapers.

OTHER						
	Variable	Estimate	Standard Error	t Value	Pr > t	Mean
b0	Intercept	16.65	4.64	3.59	0.0003	
b1	wt_price_brand1	33.46	7.17	4.67	<.0001	0.31
b2	wt_price_brand2	28.18	6.90	4.08	<.0001	0.26
b3	wt_price_brand3	18.45	6.15	3	0.0027	0.31
b4	wt_price_brand4	-82.02	7.23	-11.34	<.0001	0.22
b5	disp_wt_brand1	-3.60	1.56	-2.31	0.0208	0.02
b6	disp_wt_brand2	-1.81	1.32	-1.37	0.1702	0.01
b7	disp_wt_brand3	-0.92	1.34	-0.69	0.4912	0.02
b8	disp_wt_brand4	34.99	1.19	29.4	<.0001	0.03
b9	Feature_wt_brand1	-19.16	6.45	-2.97	0.003	0.13
b10	Feature_wt_brand2	-5.43	6.18	-0.88	0.3791	0.06
b11	Feature_wt_brand3	69.27	5.49	12.62	<.0001	0.11
b12	Feature_wt_brand4	15.26	0.66	23.21	<.0001	0.10
b13	PR_wt_brand1	92.31	6.72	13.75	<.0001	0.20
b14	PR_wt_brand2	-2.03	0.48	-4.22	<.0001	0.16
b15	PR_wt_brand3	-0.79	0.66	-1.19	0.2322	0.17
b16	PR_wt_brand4	2.02	0.62	3.25	0.0011	0.21
b17	price_PR1	-25.50	18.16	-1.4	0.1603	0.06
b18	price_PR2	-43.90	13.57	-3.23	0.0012	0.05
b19	price_PR3	-29.79	11.68	-2.55	0.0108	0.06
b20	price_PR4	-289.03	13.13	-22.01	<.0001	0.05
b21	price_F1	59.51	20.79	2.86	0.0042	0.04
b22	price_F2	8.27	23.49	0.35	0.7249	0.01
b23	price_F3	-22.02	15.02	-1.47	0.1427	0.03
b24	price_F4	-284.33	15.51	-18.33	<.0001	0.03
b25	PR_F1	7.11	1.60	4.43	<.0001	0.08
b26	PR_F2	3.43	1.92	1.78	0.0745	0.04
b27	PR_F3	1.05	1.53	0.69	0.492	0.07
b28	PR_F4	9.80	1.65	5.93	<.0001	0.02

mean Sales	
tot_units HUGGIES	43.86
tot_units LUVS	19.75
tot_units PAMPERS	38.58
tot_units OTHER	45.07

Self Price Elasticity	
if Feature	-1.796
if PR	-1.819
if Feature and PR	-3.214
if no PR and no Feature	-0.402

Cross Price Elasticity with HUGGIES	
if Feature	0.632
if PR	0.054
if Feature and PR	0.459
if no PR and no Feature	0.228

Cross Price Elasticity with LUVS	
if Feature	0.211
if no PR and no Feature	0.163

Cross Price Elasticity with PAMPERS	
if no PR and no Feature	0.126

From the above results we observe that if Brands other than the top-3 do a reduction in price by 1% then there will be 0.4% increase in its sales. But if they do a reduction in price alongside a feature-display, for every 1% reduction in price, its sales will increase by 1.8%.

On the other hand, if HUGGIES plans to do a discount sale, then with every 1% reduction of HUGGIES price, there will be a 0.6% decrease in sales of other diapers.

7. BRAND CHOICE PREFERENCE

7.1 STORE FEATURE STUDY

To understand the impact of store level features like display, price reduction and product feature on brand choice selection by customers the following activities were carried out

7.1.1 Data Preparation

1. The store level data for Grocery and Drug Store were merged and converted to a single data extract.
2. All store level data were grouped into four brands – Huggies, Pampers, Luvs and Others.
3. The store level data were filtered for customers available in the customer panel data – around 84% customer data was found at store level.
4. Weighted Price for a product is calculated for a store for a particular week (i):

$$WeightedPrice_i = \sum Price \text{ per Diaper}_i * \frac{Sales \text{ of Diaper}_i}{Total \text{ Sales of the Brand}_i}$$

Similarly, weighted display, feature and price reduction score are calculated.

5. The final dataset had all the weighted metrics of four brands and choices made by the customers within four brands for each purchase made by them in the year 2010.

7.1.2 SAS Results

To regress the favorability for Huggies among all other brands against store level promotions done across week by each of the brands a PROC LOGISTIC regression was run with choice of brands as dependent variable and the following variables as independent variable

Independent Variable	Type
Weighted price of Huggies	Continuous
Weighted price of Pampers	Continuous
Weighted price of Luvs	Continuous
Weighted price of Other Brands	Continuous
Weighted feature of Huggies	Continuous
Weighted feature of Pampers	Continuous
Weighted feature of Luvs	Continuous
Weighted feature of Huggies	Continuous
Weighted price reduction of Pampers	Continuous
Weighted price reduction of Luvs	Continuous
Weighted price reduction of Huggies	Continuous
Weighted price reduction of Other Brands	Continuous
Weighted store display of Huggies	Continuous
Weighted store display of Pampers	Continuous

Weighted store display of Luvs	Continuous
Weighted store display of Other Brands	Continuous

The SAS System		
The LOGISTIC Procedure		
Model Information		
Data Set	WORK.BRAND_CHOICE_STORE_FEATURES	
Response Variable	brand	
Number of Response Levels	4	
Model	generalized logit	
Optimization Technique	Newton-Raphson	
Number of Observations Read	3674	
Number of Observations Used	3674	
Response Profile		
Ordered Value	brand	Total Frequency
1	HUGGIES	825
2	LUVS	737
3	OTHER	1351
4	PAMPERS	761

Logits modeled use brand='HUGGIES' as the reference category.

total_num_panel_records
4430

From a total of 4430 records in panel data, we consider 83% of the data for performing logistic regression for brand choice selection. Since, these purchases are made from drug and grocery stores.

Model fit Stats – Based on AIC, SC and -2log L output – Intercept and Covariates model is the best fit model for the data

Model Convergence Status			
Convergence criterion (GCONV=1E-8) satisfied.			
Model Fit Statistics			
Criterion	Intercept Only	Intercept and Covariates	
AIC	9937.851	9233.883	
SC	9956.478	9476.035	
-2 Log L	9931.851	9155.883	
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	775.9687	36	<.0001
Score	877.0930	36	<.0001
Wald	745.7140	36	<.0001

reference = HUGGIES

Analysis of Maximum Likelihood Estimates						
Parameter	brand	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	LUVS	-0.09	0.1037	0.6837	0.41	0.918
Intercept	OTHER	0.78	0.0873	79.3205	<.0001	2.176
Intercept	PAMPERS	-0.20	0.1042	3.6123	0.06	0.82
disp_wt_brand1	LUVS	-4.14	1.8796	4.8407	0.03	0.016
disp_wt_brand1	OTHER	-1.94	1.24	2.4408	0.12	0.144
disp_wt_brand1	PAMPERS	-5.62	2.7496	4.1817	0.04	0.004
disp_wt_brand2	LUVS	-1.03	0.7277	2.0184	0.16	0.356
disp_wt_brand2	OTHER	-0.65	0.6315	1.0655	0.30	0.521
disp_wt_brand2	PAMPERS	0.32	0.576	0.3142	0.58	1.381
disp_wt_brand3	LUVS	-1.64	0.6674	6.066	0.01	0.193
disp_wt_brand3	OTHER	-1.27	0.5622	5.0992	0.02	0.281
disp_wt_brand3	PAMPERS	-0.81	0.543	2.246	0.13	0.443
disp_wt_brand4	LUVS	-0.54	0.6831	0.6168	0.43	0.585
disp_wt_brand4	OTHER	2.38	0.4963	23.0214	<.0001	10.817
disp_wt_brand4	PAMPERS	-1.54	0.7994	3.6961	0.05	0.215
Feature_wt_brand1	LUVS	-1.57	0.3551	19.5729	<.0001	0.208
Feature_wt_brand1	OTHER	-1.81	0.3019	35.779	<.0001	0.164
Feature_wt_brand1	PAMPERS	-1.46	0.3653	15.8952	<.0001	0.233
Feature_wt_brand2	LUVS	1.51	0.2473	37.3032	<.0001	4.53
Feature_wt_brand2	OTHER	0.22	0.2543	0.7153	0.40	1.24
Feature_wt_brand2	PAMPERS	0.42	0.286	2.1587	0.14	1.522
Feature_wt_brand3	LUVS	-0.81	0.3889	4.3648	0.04	0.444
Feature_wt_brand3	OTHER	-1.38	0.3444	15.992	<.0001	0.252
Feature_wt_brand3	PAMPERS	0.42	0.3522	1.4264	0.23	1.523
Feature_wt_brand4	LUVS	0.07	0.3111	0.0488	0.83	1.071
Feature_wt_brand4	OTHER	0.85	0.2533	11.3562	0.00	2.348
Feature_wt_brand4	PAMPERS	0.04	0.3038	0.0154	0.90	1.038
PR_wt_brand1	LUVS	0.53	0.3507	2.3167	0.13	1.705
PR_wt_brand1	OTHER	0.24	0.2991	0.6278	0.43	1.267
PR_wt_brand1	PAMPERS	0.40	0.3607	1.2226	0.27	1.49
PR_wt_brand2	LUVS	0.39	0.1572	6.0089	0.01	1.47
PR_wt_brand2	OTHER	-0.38	0.149	6.3385	0.01	0.687
PR_wt_brand2	PAMPERS	-0.05	0.1643	0.0809	0.78	0.954
PR_wt_brand3	LUVS	0.81	0.3751	4.7114	0.03	2.257
PR_wt_brand3	OTHER	1.00	0.3377	8.8281	0.00	2.727

PR_wt_brand3	PAMPERS	1.24	0.3494	12.5232	0.00	3.443
PR_wt_brand4	LUVS	-0.23	0.2589	0.7595	0.38	0.798
PR_wt_brand4	OTHER	-0.42	0.2252	3.4383	0.06	0.659
PR_wt_brand4	PAMPERS	0.28	0.2498	1.2399	0.27	1.321

Most of the parameters are statistically significant - disp_wt_brand1_{Huggies}, disp_wt_brand4_{Other}, disp_wt_brand3_{Pampers} Feature_wt_brand1_{Huggies}, Feature_wt_brand2_{Luvs}, Feature_wt_brand3_{Pampers}, Feature_wt_brand4_{Other}, PR_wt_brand2_{Luvs}, PR_wt_brand3_{Pampers}, PR_wt_brand4_{Other}.

INTERPRETATION:

Given other variables are held constant, the likelihood of choice selection between brands are listed out below

1. In comparison with brand Huggies, if there is a high incidence of store display for Huggies products people will less likely choose Luvs products.
2. In comparison with brand Huggies, if there is a high incidence of store display for Huggies products people will less likely choose Pampers products.
3. In comparison with brand Huggies, if there is a high incidence of store display for Pampers products people will less likely choose Luvs products.
4. In comparison with brand Huggies, if there is a high incidence of store display for Pampers products people will less likely choose Other brand products.
5. In comparison with brand Huggies, if there is a high incidence of store display for Other brand products people will more likely choose Other brand products.
6. In comparison with brand Huggies, if there is a high incidence of store display for Other brand products people will less likely choose Pampers products. (90%)
7. In comparison with brand Huggies, if there is a high incidence of store level promotion for features of Huggies products people will less likely choose Luvs or Pampers or Other brand products.
8. In comparison with brand Huggies, if there is a high incidence of store level promotion for features of Luvs products people will more likely choose Luvs products.
9. In comparison with brand Huggies, if there is a high incidence of store level promotion for features of Pampers products people will less likely choose Luvs products.
10. In comparison with brand Huggies, if there is a high incidence of store level promotion for features of Pampers products people will less likely choose Other brand products.
11. In comparison with brand Huggies, if there is a high incidence of store level promotion for features of Other brand products people will more likely choose Other brand products.
12. In comparison with brand Huggies, if there is a high incidence of store level price reduction of Luvs products people will more likely choose Luvs products.
13. In comparison with brand Huggies, if there is a high incidence of store level price reduction of Luvs products people will less likely choose Other brand products.
14. In comparison with brand Huggies, if there is a high incidence of store level price reduction of Pampers products people will more likely choose Luvs or Pampers or Other brand products.
15. In comparison with brand Huggies, if there is a high incidence of store level price reduction of Other products people will less likely choose Other brand products. (90%)

7.1.3 Insight

Based on the results of Multinomial Logistic Regression on store level metrics, the following preference matrix has been created. The matrix explains effect of brand promotions of Huggies compared to all other brands and their impact on Huggies. The following insight can be used to strategize an effective store promotion strategy so that losses can be minimized on store promotions.

-  **Preference for Huggies**
-  **Preference for Other Brands**

DFP Brands \	LUVS	PAMPERS	OTHERS
Display- HUGGIES	✓	✓	---
Display- PAMPERS	✓	---	✓
Display - OTHERS	---	---	✓
Feature - HUGGIES	✓	✓	✓
Feature - LUVS	✗	---	---
Feature- PAMPERS	✓	---	✓
Feature - Others	---	---	✗
Price Reduction- LUVS	✗	---	✓
Price Reduction- PAMPERS	✗	✗	✗

7.2 DEMOGRAPHIC STUDY

To understand more about the demographics of customer characteristics favoring for a brand, a LOGISTIC regression was run on customer demographic data.

7.2.1 Data Preparation

1. The customer purchase data across stores are joined with their demographic data as the first step in the data preparation process
2. Different groupings are created based upon education, age of male head, age of female head, family size, number of children and occupation level
3. The customers that had missing demographic information were filtered out of the data.
4. Finally, dummy variables were created to compare each demographic characteristic keeping a base reference characteristic. For example – for variable family size one dummy variable “fam_size_large” was created keeping regular size as reference.
5. Similarly, the step 4 was carried out to all demographic variables considered for the logistic regression.

7.2.2 SAS Results

The SAS System		
The LOGISTIC Procedure		
Model Information		
Data Set	WORK.PANEL_MDC1	
Response Variable	brand	
Number of Response Levels	4	
Model	generalized logit	
Optimization Technique	Newton-Raphson	
Number of Observations Read	2306	
Number of Observations Used	2306	
Response Profile		
Ordered Value	brand	Total Frequency
1	HUGGIES	563
2	LUVS	577
3	OTHER	671
4	PAMPERS	495

Logits modeled use brand='HUGGIES' as the reference category.

From a total of 4430 records in panel data, we consider 83% of the data for performing logistic regression for brand choice selection. Since, these purchases are made from drug and grocery stores.

Model fit Stats – Based on AIC, SC and -2log L output – Intercept and Covariates model is the best fit model for the data

Model Convergence Status			
Convergence criterion (GCONV=1E-8) satisfied.			
Model Fit Statistics			
Criterion	Intercept Only	Intercept and Covariates	
AIC	6372.457	6051.952	
SC	6389.686	6448.238	
-2 Log L	6366.457	5913.952	
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	452.5046	66	<.0001
Score	442.9329	66	<.0001
Wald	393.0267	66	<.0001

reference = HUGGIES

Analysis of Maximum Likelihood Estimates							
Parameter		brand	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		LUVS	0.25	0.4019	0.389	0.53	1.285
Intercept		OTHER	1.38	0.3519	15.4319	<.0001	3.985
Intercept		PAMPERS	0.42	0.3767	1.2184	0.27	1.516
fam_size	large	LUVS	0.38	0.0925	16.5177	<.0001	1.456
fam_size	large	OTHER	0.46	0.0941	23.9183	<.0001	1.584
fam_size	large	PAMPERS	0.66	0.1008	42.4965	<.0001	1.93
fam_income	high	LUVS	-0.43	0.1517	7.9293	0.00	0.652
fam_income	high	OTHER	-0.35	0.1477	5.468	0.02	0.708
fam_income	high	PAMPERS	-0.20	0.15	1.7708	0.18	0.819
fam_income	medium	LUVS	-0.12	0.145	0.6821	0.41	0.887
fam_income	medium	OTHER	0.05	0.142	0.114	0.74	1.049
fam_income	medium	PAMPERS	-0.36	0.1457	6.2311	0.01	0.695
fam_income	very_high	LUVS	-0.84	0.2699	9.7149	0.00	0.431
fam_income	very_high	OTHER	-0.59	0.2495	5.5469	0.02	0.556
fam_income	very_high	PAMPERS	0.17	0.2295	0.5236	0.47	1.181
age_male	elder	LUVS	-0.43	0.3469	1.5169	0.22	0.652
age_male	elder	OTHER	-0.20	0.2986	0.4656	0.50	0.816
age_male	elder	PAMPERS	-0.12	0.316	0.1467	0.70	0.886
age_male	mid_career	LUVS	0.54	0.2984	3.3304	0.07	1.724
age_male	mid_career	OTHER	-0.21	0.257	0.6796	0.41	0.809
age_male	mid_career	PAMPERS	0.40	0.2684	2.2635	0.13	1.498
age_female	elder	LUVS	-0.24	0.2867	0.6975	0.40	0.787
age_female	elder	OTHER	-0.07	0.2451	0.0839	0.77	0.931
age_female	elder	PAMPERS	-0.53	0.2568	4.2403	0.04	0.589
age_female	mid_career	LUVS	0.27	0.1994	1.7903	0.18	1.306
age_female	mid_career	OTHER	-0.01	0.1747	0.0044	0.95	0.988
age_female	mid_career	PAMPERS	-0.34	0.1755	3.7979	0.05	0.71
educ_male	college	LUVS	-0.41	0.135	9.3903	0.00	0.661
educ_male	college	OTHER	-0.34	0.1316	6.7702	0.01	0.71
educ_male	college	PAMPERS	-0.38	0.1406	7.1164	0.01	0.687
educ_male	graduate	LUVS	-0.06	0.1596	0.1356	0.71	0.943
educ_male	graduate	OTHER	-0.07	0.1586	0.1787	0.67	0.935
educ_male	graduate	PAMPERS	-0.14	0.168	0.6845	0.41	0.87
educ_female	college	LUVS	0.14	0.2868	0.248	0.62	1.154
educ_female	college	OTHER	-0.10	0.2722	0.1391	0.71	0.903
educ_female	college	PAMPERS	0.22	0.2825	0.6213	0.43	1.249

educ_female	graduate	LUVS	-0.29	0.2948	0.9821	0.32	0.747
educ_female	graduate	OTHER	-0.78	0.2815	7.6534	0.01	0.459
educ_female	graduate	PAMPERS	-0.80	0.2918	7.5049	0.01	0.45
child_num	1	LUVS	0.51	0.1905	7.0953	0.01	1.661
child_num	1	OTHER	-0.12	0.1537	0.5976	0.44	0.888
child_num	1	PAMPERS	0.15	0.1824	0.6787	0.41	1.162
child_num	2	LUVS	-0.72	0.1994	13.0373	0.00	0.487
child_num	2	OTHER	-1.06	0.1613	42.9266	<.0001	0.348
child_num	2	PAMPERS	-0.83	0.1859	19.8894	<.0001	0.436
child_num	3	LUVS	-0.49	0.4821	1.0144	0.31	0.615
child_num	3	OTHER	0.93	0.3378	7.5364	0.01	2.528
child_num	3	PAMPERS	-0.59	0.4259	1.9472	0.16	0.552
occu_male	blue	LUVS	-0.19	0.1395	1.7979	0.18	0.829
occu_male	blue	OTHER	-0.02	0.1329	0.0327	0.86	0.976
occu_male	blue	PAMPERS	-0.19	0.1432	1.6884	0.19	0.83
occu_male	white_high	LUVS	-0.22	0.1188	3.4744	0.06	0.801
occu_male	white_high	OTHER	-0.13	0.1126	1.2701	0.26	0.881
occu_male	white_high	PAMPERS	-0.18	0.1213	2.2072	0.14	0.835
occu_male	white_low	LUVS	-0.23	0.135	2.989	0.08	0.792
occu_male	white_low	OTHER	0.18	0.1231	2.2035	0.14	1.2
occu_male	white_low	PAMPERS	-0.04	0.1353	0.0851	0.77	0.961
occu_female	blue	LUVS	0.24	0.1675	2.0042	0.16	1.268
occu_female	blue	OTHER	0.50	0.148	11.4407	0.00	1.65
occu_female	blue	PAMPERS	0.05	0.1693	0.0745	0.78	1.047
occu_female	white_high	LUVS	-0.30	0.1297	5.1766	0.02	0.744
occu_female	white_high	OTHER	-0.42	0.1211	12.3069	0.00	0.654
occu_female	white_high	PAMPERS	-0.08	0.1308	0.3489	0.55	0.926
occu_female	white_low	LUVS	0.00	0.1457	0.001	0.98	1.005
occu_female	white_low	OTHER	-0.08	0.1363	0.3536	0.55	0.922
occu_female	white_low	PAMPERS	0.08	0.1486	0.2785	0.60	1.082
pets_total		LUVS	-0.18	0.0622	8.7626	0.00	0.832
pets_total		OTHER	0.06	0.0568	1.1406	0.29	1.063
pets_total		PAMPERS	-0.04	0.0629	0.3515	0.55	0.963

From the results of multinomial logistic regression, we can find the following variables are significant in the model - fam_size -large, fam_income -high, fam_income -medium, fam_income -veryhigh, age_male, age_female, educ_male, educ_female, child_num, occu_male, occu_female, pets_total

INTERPRETATION:

Given other variables are held constant, the likelihood of choice selection between brands are listed out below

1. In comparison with brand Huggies as alternative, large family households(members>4) will most likely prefer LUVS, Pampers and Other Brands over a regular family household.
2. In comparison with brand Huggies as alternative, families with high level income will less likely prefer products of LUVS and Other brands over a family with lower income level.
3. In comparison with brand Huggies as alternative, families with medium level income will less likely prefer Pampers products over a family with lower income level.
4. In comparison with brand Huggies as alternative, families with very high-level income will less likely prefer products of LUVS and Other brands over a family with lower income level.
5. In comparison with brand Huggies as alternative, a middle-aged man will more likely prefer Luvs products compared to a younger man.
6. In comparison with brand Huggies as alternative, an elderly woman will less likely prefer Pampers products compared to a younger woman.
7. In comparison with brand Huggies as alternative, a middle-aged woman will more likely prefer Pampers products compared to a younger woman.
8. In comparison with brand Huggies as alternative, a male with college level degree will less likely prefer Luvs, Pampers and other brand products compared to a person with high school degree.
9. In comparison with brand Huggies as alternative, a female with graduate level degree will less likely prefer Pampers and other brand products compared to a female with high school education.
10. In comparison with brand Huggies as alternative, a household with single child will more likely prefer Luvs products compared to a household expecting a kid.
11. In comparison with brand Huggies as alternative, a household with two kids will less likely prefer Luvs, Pampers and Other brand products compared to a household expecting a kid.
12. In comparison with brand Huggies as alternative, a household with three kids will more likely prefer Other brand products compared to a household expecting a kid.
13. In comparison with brand Huggies as alternative, a female carrying out blue collar work will more likely prefer Other brand products compared to an unemployed female.
14. In comparison with brand Huggies as alternative, a female carrying out white collar work will less likely prefer LUVS and Other brand products compared to an unemployed female.
15. In comparison LUVS brand Huggies as alternative, a household having high incidence of pets will less likely prefer LUVS brand products.

7.2.3 Insights

Based on the results of Multinomial Logistic Regression on customer demographics metrics, the following preference matrix has been created. The matrix shows the effect of customer demographics on brand preference over another. This insight can be used to understand our customer base as to which kind of customers will most preferably buy our brand compared to others and create effective promotion strategies targeting those customers.

-  **Preference for Huggies**
-  **Preference for Other Brands**

Demographics \ Brands	LUVS	PAMPERS	OTHERS
Family Income (High vs Low)	✓	---	✓
Medium vs Low	---	✓	---
Very High vs Low	✓	---	✓
Female Elderly vs Young	---	✓	
Male College vs School	✓	✓	✓
Female Graduate vs School	---	✓	✓
Number of child 1 vs 0	✗	---	---
Number of child 2 vs 0	✓	✓	✓
Female Occp Blue vs Unemployed	---	---	✗
Female Occp White vs Unemployed	✓	---	✓

8. MONETARY SEGMENTATION using RFM

To understand more about the characteristics of customers preferring Huggies and to study which segment of customers are more frequent, recent buyers of the product and revenue generators, an RFM analysis has been carried out.

8.1 RFM Strategy

1. The data set containing the purchase information of each panel ID across week was filtered for brand Huggies
2. From this dataset, the following metrics were calculated
 - a. Recency by Week – How recently a customer purchased? – Difference between last week of purchase and each week a customer bought an Huggies product.
 - b. Frequency – How often a customer purchased? – Count of the number of weeks of purchase available for each customer in the dataset.
 - c. Monetary – How much a customer spent for the product? – Sum of all the dollar amount paid by a customer made during purchase across weeks in the dataset.
3. Based upon the frequency distribution and univariate results, every 20th percentile of each of the RFM metrics were calculated
4. Each customer based upon their recency, frequency and total money spent on purchase were tagged with Recency, Frequency and Monetary Scores with '5' being highest and '1' being lowest.
5. For Recency, the lowest value was given the highest score of 5 and vice versa for Frequency and Monetary scores.

The SAS System						
The MEANS Procedure						
Variable	Minimum	20th Pctl	40th Pctl	60th Pctl	80th Pctl	Maximum
monetary	1.000000	11.590000	19.390000	36.540000	86.440000	929.650000
freq	1.000000	1.000000	2.000000	3.000000	7.000000	69.000000
weekcnt	0	3.000000	8.000000	18.000000	31.000000	51.000000

8.2 RFM Output

We checked the correlation between Recency, Frequency and Monetary scores and found that Monetary and Frequency had 0.93 correlation. Hence, the monetary scores explained frequency of customers buying our brand from stores.

In addition to the above, Recency doesn't impact a lot on sales, as people diapers only there is a child expected to be born or when there are kids between 1-17 months in the house. Therefore, the segmentation was carried out using Monetary scores which differentiated the most valuable and least valuable customers for our brand.

The SAS System						
The CORR Procedure						
3 Variables:	recency_score	frequency_score	monetary_score			
Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
recency_score	946	3.07294	1.45550	2907	1.00000	5.00000
frequency_score	946	2.62368	1.58366	2482	1.00000	5.00000
monetary_score	946	2.78964	1.60526	2639	1.00000	5.00000
Pearson Correlation Coefficients, N = 946 Prob > r under H0: Rho=0						
	recency_score	frequency_score	monetary_score			
recency_score	1.00000	0.49488 <.0001	0.47352 <.0001			
frequency_score	0.49488 <.0001	1.00000	0.93246 <.0001			
monetary_score	0.47352 <.0001	0.93246 <.0001	1.00000			

The customers were segmented into two groups – top customers with score of 4 and 5 and bottom customers with score of 1 and 2.

Later the demographics of these customers were studied to generate insights on which customers to target for or give discounts to increase sales.

Frequent & High Revenue Customers						Jumpers & Low Revenue Customers					
Variable	N	Mean	Std Dev	Minimum	Maximum	Variable	N	Mean	Std Dev	Minimum	Maximum
PANID	132	2206066.9	1040816.3	1118786	3842443	PANID	141	2174533.5	1047184.8	1104935	3842559
monetary	132	85.300152	70.003923	27.47	378.21	monetary	141	12.073688	2.2461225	5.99	16.99
freq	132	5.9242424	4.9940747	2	28	freq	141	1	0	1	1
weekcnt	132	12.204546	11.857126	0	50	weekcnt	141	24.411348	15.073662	0	51
monetary_score	132	4.5227273	0.501386	4	5	monetary_score	141	1	0	1	1
frequency_score	132	4.3181818	0.7238382	3	5	frequency_score	141	1	0	1	1
recency_score	132	3.6439394	1.2789104	1	5	recency_score	141	2.5035461	1.3396381	1	5
child_num	132	1.0454545	0.8548348	0	3	child_num	141	0.6595745	0.8002659	0	3
pets_total	132	0.8560606	1.0198998	0	4	pets_total	141	1	1.3680017	0	6
fam_size_L	132	0.5227273	0.501386	0	1	fam_size_L	141	0.3120567	0.4649847	0	1
fam_size_R	132	0.4772727	0.501386	0	1	fam_size_R	141	0.6879433	0.4649847	0	1
fam_size_O	132	0	0	0	0	fam_size_O	141	0	0	0	0
fam_income_L	132	0.1439394	0.3523655	0	1	fam_income_L	141	0.1489362	0.3572948	0	1
fam_income_M	132	0.4318182	0.4972164	0	1	fam_income_M	141	0.5460993	0.4996453	0	1
fam_income_H	132	0.3333333	0.4732004	0	1	fam_income_H	141	0.2624113	0.4415135	0	1
fam_income_VH	132	0.0909091	0.288575	0	1	fam_income_VH	141	0.0425532	0.2025671	0	1
fam_income_O	132	0	0	0	0	fam_income_O	141	0	0	0	0
age_mY	132	0.0075758	0.0870388	0	1	age_mY	141	0.0141844	0.1186722	0	1
age_mM	132	0.6136364	0.4887705	0	1	age_mM	141	0.4184397	0.4950617	0	1
age_mE	132	0.1515152	0.3599162	0	1	age_mE	141	0.3617021	0.4822062	0	1
age_mO	132	0.2272727	0.4206667	0	1	age_mO	141	0.2056738	0.4056341	0	1
age_fY	132	0.0454545	0.2090924	0	1	age_fY	141	0.0638298	0.245321	0	1
age_fM	132	0.719697	0.4508583	0	1	age_fM	141	0.4893617	0.5016689	0	1
age_fE	132	0.1515152	0.3599162	0	1	age_fE	141	0.3900709	0.4895048	0	1
age_fD	132	0.0933333	0.2774383	0	1	age_fD	141	0.0567376	0.2321653	0	1
educ_mS	132	0.030303	0.1720729	0	1	educ_mS	141	0.0921986	0.2903375	0	1
educ_mC	132	0.4848485	0.5016743	0	1	educ_mC	141	0.4893617	0.5016689	0	1
educ_mO	132	0.2272727	0.4206667	0	1	educ_mO	141	0.2056738	0.4056341	0	1
educ_mG	132	0.3636364	0.4828783	0	1	educ_mG	141	0.2411348	0.4292969	0	1
educ_fS	132	0.0227273	0.1496004	0	1	educ_fS	141	0.0425532	0.2025671	0	1
educ_fC	132	0.530303	0.5009821	0	1	educ_fC	141	0.6595745	0.4755416	0	1
educ_fO	132	0.0833333	0.2774383	0	1	educ_fO	141	0.0567376	0.2321653	0	1
occ_mWH	132	0.3409091	0.4758206	0	1	occ_mWH	141	0.2836879	0.4523943	0	1
occ_mWL	132	0.1287879	0.3362411	0	1	occ_mWL	141	0.1276596	0.3349001	0	1
occ_mB	132	0.1590909	0.3671542	0	1	occ_mB	141	0.1347518	0.3426756	0	1
occ_mND	132	0.219697	0.4156186	0	1	occ_mND	141	0.3617021	0.4822062	0	1
occ_fWH	132	0.4318182	0.4972164	0	1	occ_fWH	141	0.2695035	0.4452837	0	1
occ_fWL	132	0.1515152	0.3599162	0	1	occ_fWL	141	0.1843972	0.3891903	0	1
occ_fb	132	0.0681818	0.2530179	0	1	occ_fb	141	0.0851064	0.2800347	0	1
occ_fNO	132	0.1439394	0.3523655	0	1	occ_fNO	141	0.2624113	0.4415135	0	1
one_child	132	0.3636364	0.4828783	0	1	one_child	141	0.3120567	0.4649847	0	1
two_child	132	0.2954545	0.4579849	0	1	two_child	141	0.141844	0.3501339	0	1
three_child	132	0.030303	0.1720729	0	1	three_child	141	0.0212766	0.1448194	0	1
zero_child	132	0.3106061	0.4645046	0	1	zero_child	141	0.5248227	0.5011638	0	1

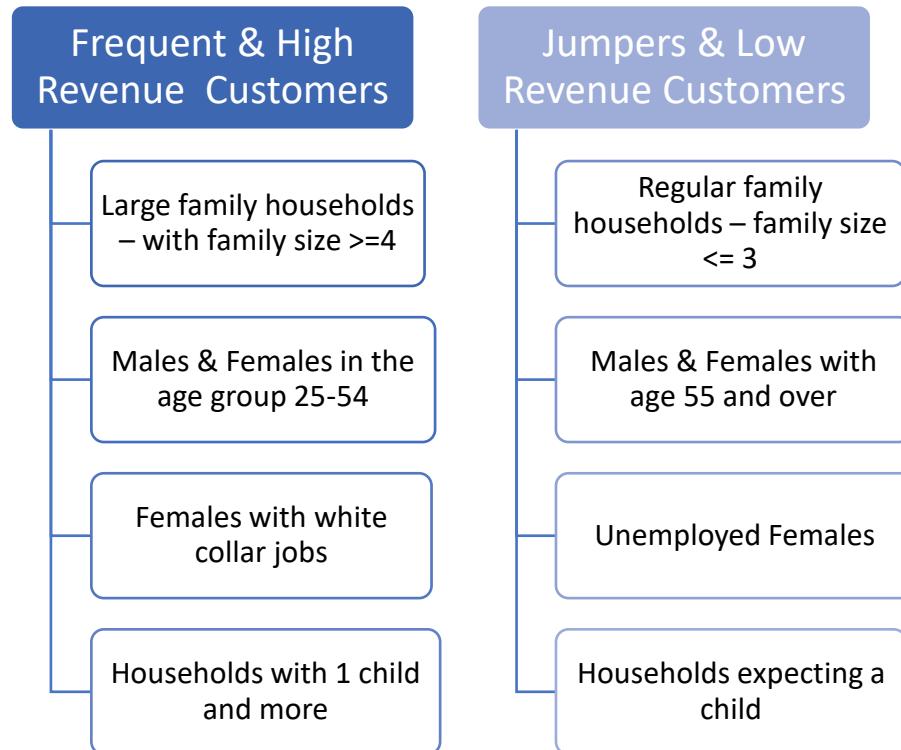
From the above output, we can see that 132 people in segment 1, have contributed up to 82.44% of the overall revenue. These are customers who visit the shop most frequently and are most recent compared to the other segment.

The segment 2 which has 141 are the low revenue customers which contribute to only 8% of the overall revenue.

1. Large family size households are high monetary customers compared to regular family size households
2. Middle-aged male and female groups in a household prefer our brand products over other age groups.
3. Our products are preferred more by household having females occupying positions in White collar jobs
4. Generally, females in the household with no occupation will more likely not favor our products.
5. Households with one or two kids will more preferably buy our product than households expecting a child.

8.3 Insights

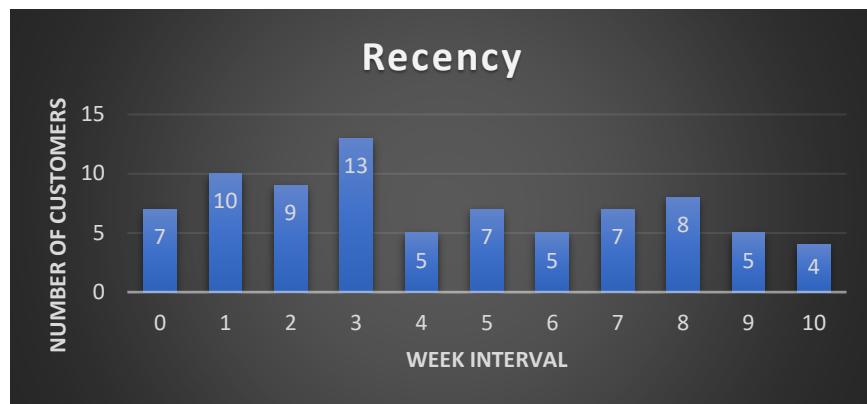
By studying the means of two clusters High and Low Monetary customers we can make out the following insights.



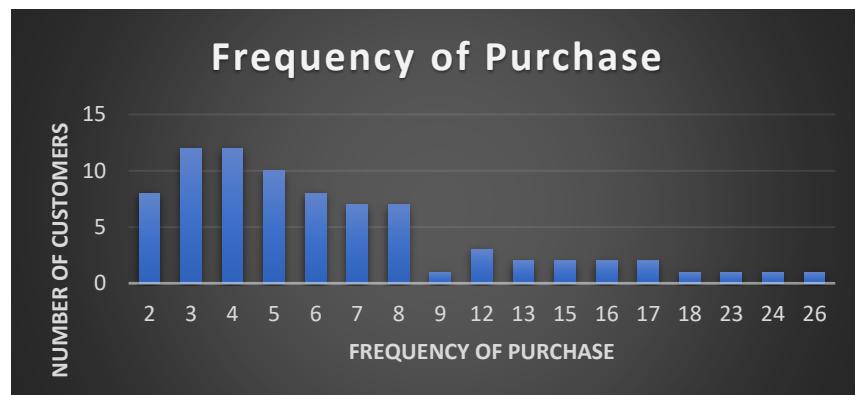
1. Generally, households where a child is expected do not have specific choice of brand as they are unaware about the advantages of one brand over another - it is most likely that their choice will be driven by price of the diaper rather than brand name or quality.
2. The family size being large has higher mean in segment1 while segment2 has higher mean for regular family size. Larger family size prefers our brand either due to experience (using the diaper for first born) and strong brand connection.
3. Age of male or female in the household also affects the choice of the brand because middle age is constantly aware about the features of a product and its offering and would be willing to go by brand name and loyalty.

Furthermore, the top customers were drilled down further to perform targeted marketing for increasing sales. The top 40% of customers based upon Monetary and Recency scores can targeted are **loyal customers**, as they have high affinity towards our products. These customers will more likely buy our products irrespective of promotions. A total of **80 (23.12%)** households are found as loyal customers.

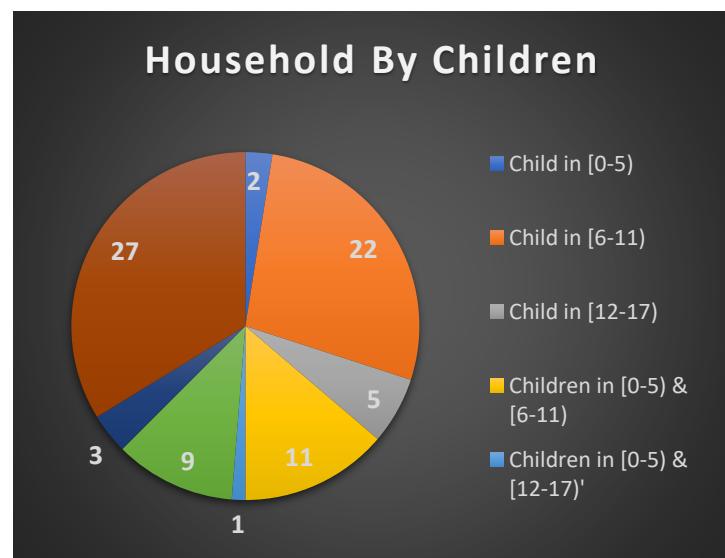
By Recency - The graph plots the most recent customers in the high monetary region who bought our products in past four weeks.



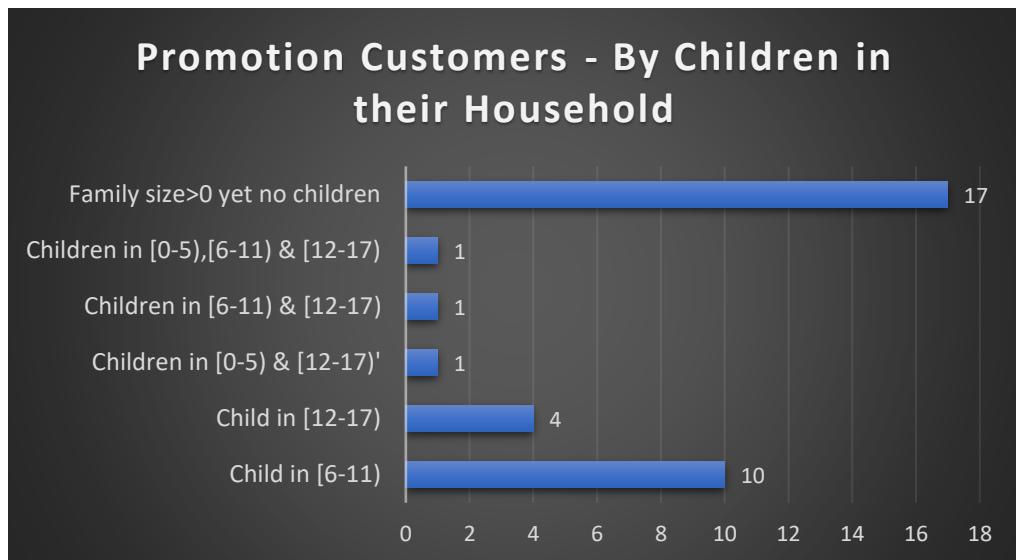
By Frequency - The graph plots the customers who have made frequent purchases more than 4 times in the high monetary region.



Around 57% percent of the revenue are gained from these loyal customers for our brand from PANEL data study. Below pie-chart shows the customer distribution by children wherein households having children between 0 to 11 months can be targeted specially.



Using the segmentation, we found out people with low monetary scores but high recency scores indicating that they may purchase diaper in future – **promotion customers**. These customers can be targeted for promotions and discounts to lure them into purchasing our brand products. The customers with Monetary score of 1 and Recency Score of 4 and 5 i.e. The customers who recently purchased our brand product. These customers can be targeted for promotions and price reductions to lure them more into buying our products. A total of **34 (9.8%)** households are found as promotion customers.



From the above two graphs, we can observe that 34 customers have made recent purchase of our product within last 10 weeks. Also, most of the customers have made within last 4 weeks. In these specific groups, we can target special promotions for household expecting children and household having kids between 0-11 months.

9. RECOMMENDATION

- I. To increase the market share, HUGGIES should target small towns where the market share is not dominated by HUGGIES or is shared with other competitors. Targeting Tier-II and Tier-III cities of South and North-East region, where there is near equal Market Share with Pampers, Price-reduction on sales will help gain the top-brand spot

South: Richmond, Raonoke, South Carolina, Knoxville,
North East: New England, Pittsfield, Syracuse
West: Spokane
Mid-West: Grand Rapids
- II. Households with 2 or more children and falling into the high-income category (>\$55,000) prefer Huggies over others. It is advisable to lower investments on advertisements or promotions targeted to this group.
- III. Unemployed females or homemakers are less favorable towards Huggies brand, so advertisements or marketing campaigns can be targeted to these specific groups. Special advertisements targeting homemakers can be telecasted over Cable Televisions.
- IV. Households expecting a child are mostly jumpers to different brand, targeted price reduction schemes can assist in turning them to frequent buyers.
 - a. Huggies can conduct special campaigns in Hospitals describing the qualities of their products along with price reduction offers
 - b. Huggies can formulate a long-term promotion scheme from birth till 17 months of a child for households expecting a child. In this manner, we can convert households expecting a child as our long term and loyal customers.
- V. The 34 customers identified through Monetary segmentation can be targeted with discount waivers or coupons to convert them into a frequent buyer. Special discounts could be given to households with children in the age group of 0-11months.
- VI. The 80 customers who were identified as loyal customers having strong brand connection can be given loyalty points without any notable price reductions. Constant communication with these customers enquiring about their child growth and quality of our product will make them talk about our brand to others and can bring new customers.

10. APPENDIX

SAS CODE	DESCRIPTION
 PRICE_ELASTICITY.sas	Data Preparation and Price Elasticity Calculation
 Logistic.sas	Brand Choice Store and Demographic Study
 RFM.sas	Monetary Segmentation using RFM technique