BUAN/MKT 6337.001 – Predictive Analytics using SAS

Group 3 – Project report

Shalini Singh
Shivani Shreyas
Sanket Pachade

OBSERVATION

Part 1: Exploratory Analysis

We began our exploratory analysis by finding out the top 6 brands in terms of their market shares irrespective of whether it is a liquid detergent or a powder detergent. We got to know that TIDE is a clear winner here with market share of 56%. ALL, PUREX, WISK, GAIN and CHEER are in the top 6 categories after TIDE. But there is a huge gap in the dollar sales of TIDE and second ranked ALL. TIDE has a firm grip on market and majority of the customers seem to prefer TIDE when it comes to their laundry detergent
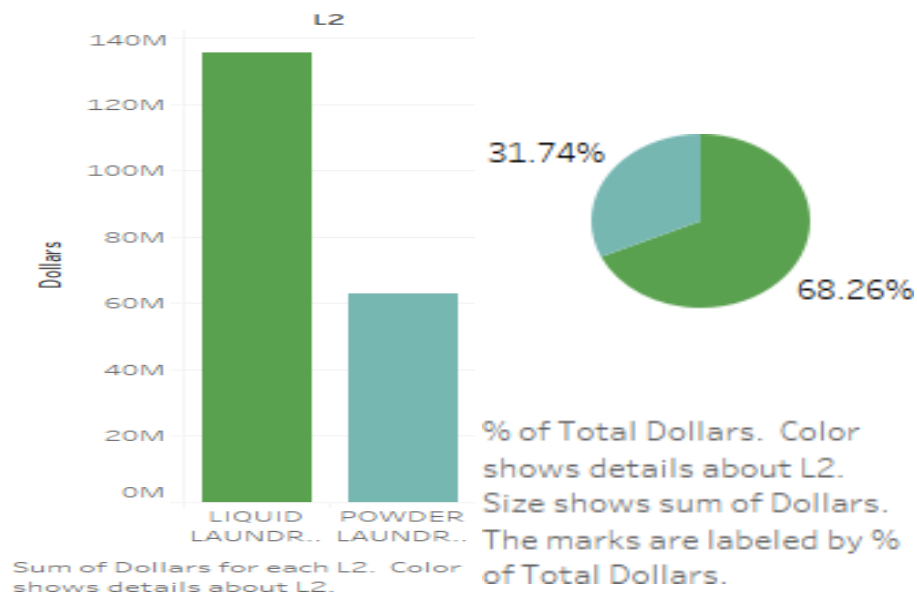
Market share of top 6 brands

Brand	Total sales	% of Total Total sales ..
TIDE	75,923,992	56.08%
ALL	15,116,464	11.17%
PUREX	12,333,547	9.11%
WISK	12,253,481	9.05%
GAIN	10,343,370	7.64%
CHEER	9,412,879	6.95%

Total sales and % of Total Total sales along Brand broken down by Brand.

In the next part of our analysis we focused on which form of detergent has more sale. The results show that Liquid detergent is prefered by the customers over powder detergent.

Market share & sales for liquid detergent and powder detergent

Sum of Dollars for each L2. Color shows details about L2.

% of Total Dollars. Color shows details about L2. Size shows sum of Dollars. The marks are labeled by % of Total Dollars.

Next we have to do our analysis on powder detergents. So first we would like to know which brands are market leaders in this category. The stats show that, again TIDE is the dominant brand with share of 61% followed by GAIN, ARM & HAMMER FABRICARE, CHEER & SURF. There is no comeptition as far as number 1 spot is concerned. Remaining part of the market share is somewhat equally distributed with small diiferences.

Market share of Top 5 brands(powder deteregents)

Brand	Dollars	% of Total Dollars alo..
TIDE	29,340,605	61.08%
GAIN	5,781,992	12.04%
ARM & HAMMER FABRICA..	5,159,799	10.74%
CHEER	4,162,414	8.66%
SURF	3,594,688	7.48%

Dollars and % of Total Dollars along Brand broken down by Brand. The view is filtered on Brand, which has multiple members selected.

As our earlier analysis gave us the top brands, we would expect the same stats in terms of number of units sold. On doing the check we got the expected results. TIDE has sold the greatest number of units than any other brand and the difference is very significant compared to other brands. Next comes GAIN followed by ARM & HAMMER FABRICARE where the

difference is considerably less. CHEER has sold lesser number of units than SURF but its dollar sale is still more than SURF.

Number of Units sold

Brand	Units	Dollars
TIDE	3,298,900	29,340,605
GAIN	887,365	5,781,992
ARM & HAMMER FABRICA..	801,877	5,159,799
CHEER	496,401	4,162,414
SURF	558,770	3,594,688

Units and Dollars broken down by Brand. The view is filtered on Brand, which keeps ARM & HAMMER FABRICARE, CHEER, GAIN, SURF and TIDE.

Now we know what are the top brands in powder detergent category let's find out who owns what?

The top brand TIDE along with the second ranked brand GAIN, 4th in this category CHEER are owned by PROCTOR & GAMBLE. CHURCH & DWIGHT owns ARM & HAMMER FABRICARE. 5th brand SURF is owned by UNILEVER

Top 5 Brands & respective ownership

| Brand	Company	
TIDE	PROCTER & GAMBLE	29,340,605
GAIN	PROCTER & GAMBLE	5,781,992
ARM & HAMMER FABRICA..	CHURCH & DWIGHT CO INC	5,159,799
CHEER	PROCTER & GAMBLE	4,162,414
SURF	UNILEVER	3,594,688

Sum of Dollars broken down by Brand and Company. The view is filtered on Brand, which keeps ARM & HAMMER FABRICARE, CHEER, GAIN, SURF and TIDE.

Moving further we are interested in knowing more about sales of these top brands. So we drill down a bit see how does the sales over weeks look like each of these brands. As it can be seen in the line chart below the sales patterns over the weeks for these brands. The sales for top brand TIDE start at quite high above 800k in the first week and then come down to 600k after 2nd week and remains along the mean value of 600k for most of the remaining weeks. For all other brands the chart follows the same trend and the sales remain along the mean value of approximately 100k
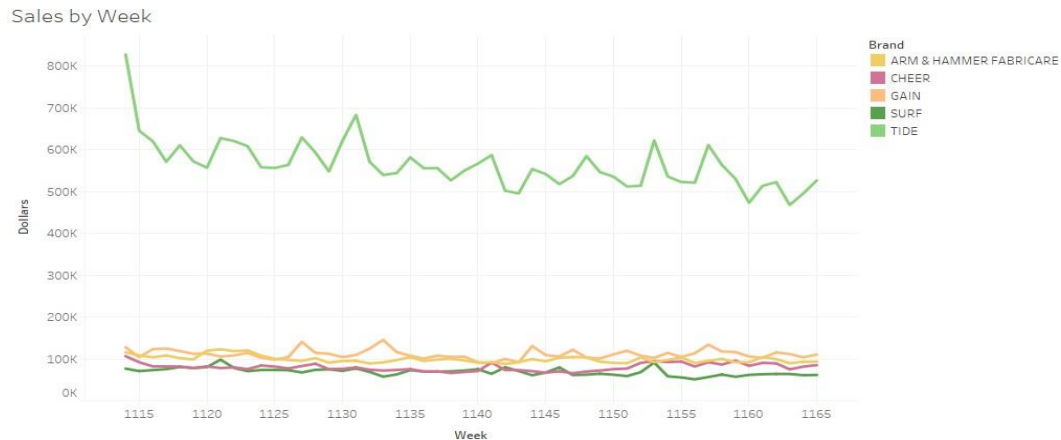
Sales by Week

Brand
- ARM & HAMMER FABRICARE
- CHEER
- GAIN
- SURF
- TIDE

The trend of sum of Dollars for Week. Color shows details about Brand. The view is filtered on Brand, which keeps ARM & HAMMER FABRICARE, CHEER, GAIN, SURF and TIDE.
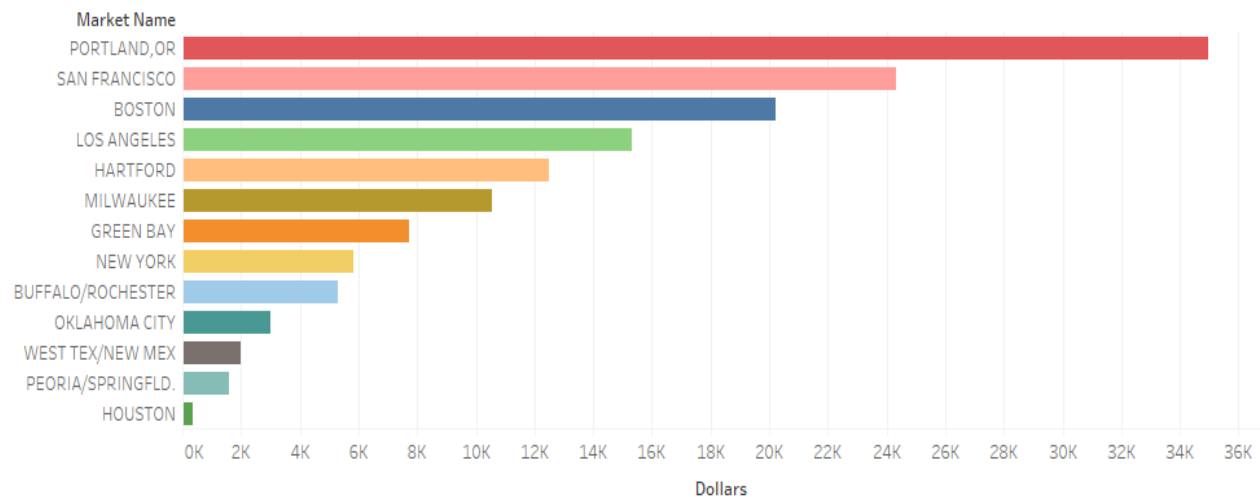
Brand Arm & HAMMER (Powder) is our brand of interest. So we analyze the weekly sales trend for it. There are two variants of the brand and out of the two ARM & HAMMER FABRICARE is preferred more by the buyers. In the first week the sales are approximately 120k and it comes down to 100k after 4th week. There is again a increase in sales to 120k in the fifth week until 2nd week. After that it remains along the mean value of 100k with slight ups and downs. For the second variant Arm & Hammer the weekly sales pretty low, the range is between 5k to 10k per week
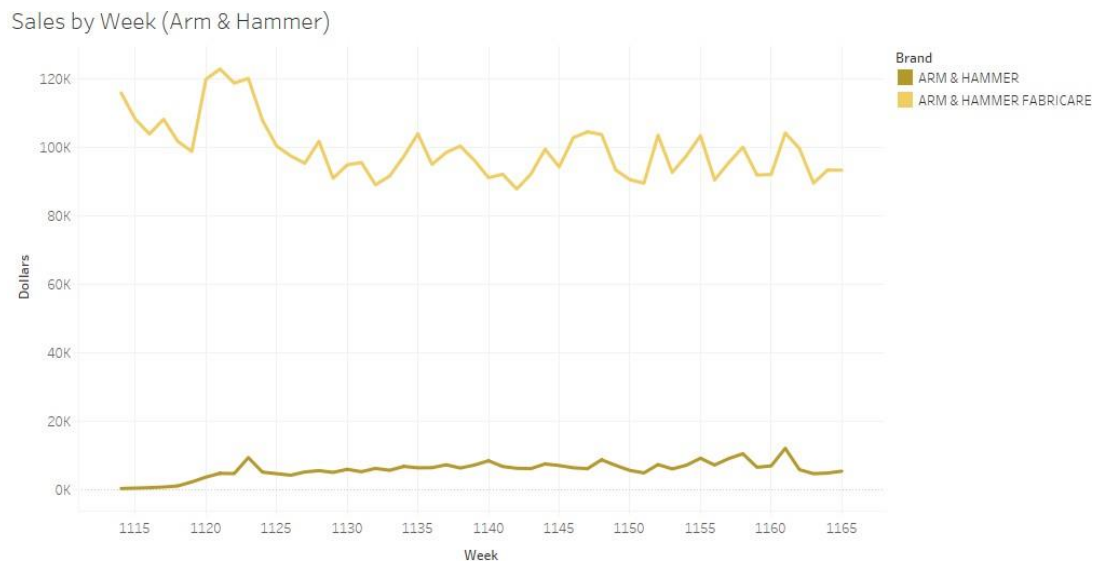
Sales by Week (Arm & Hammer)

Brand
- ARM & HAMMER
- ARM & HAMMER FABRICARE

The trend of sum of Dollars for Week. Color shows details about Brand. The view is filtered on Brand, which keeps ARM & HAMMER and ARM & HAMMER FABRICARE.

Another important aspect of our exploration is sales by market location. The data shows that Portland, San Francisco and Boston make our top 3 in terms of sales. Whereas West Tex/New Mex, Peoria/Springfield, Houston are at the bottom

Sales by market location

Market Name

Market Name	Dollars
PORTLAND,OR	~35K
SAN FRANCISCO	~24K
BOSTON	~20K
LOS ANGELES	~15K
HARTFORD	~12K
MILWAUKEE	~10.5K
GREEN BAY	~7.5K
NEW YORK	~5.5K
BUFFALO/ROCHESTER	~5K
OKLAHOMA CITY	~3K
WEST TEX/NEW MEX	~2K
PEORIA/SPRINGFLD.	~1.5K
HOUSTON	~0.3K

Dollars scale: 0K 2K 4K 6K 8K 10K 12K 14K 16K 18K 20K 22K 24K 26K 28K 30K 32K 34K 36K

Dollars

Sum of Dollars for each Market Name. Color shows details about Market Name. The data is filtered on Brand, which keeps ARM & HAMMER and ARM & HAMMER FABRICARE.

Problem Statement I

Find out the best group of customers for the Brand Arm & HAMMER (Powder)

Analysis Method Implemented: RFM Clustering

In this project, we studied the several attributes of the customers to unfold their buying patterns and clustered them into 5 segments on the basis of recency, frequency and monetary value of their purchase behaviour. After customer segmentation was done, further studies from the segmented data helped us to derive critical insights and make meaningful recommendations.

In this case, we identified the customer's purchase patterns for ARM & HAMMER (Powder)

Insights:

Market Share of Top 5 Brands

The FREQ Procedure

		Cluster		
CLUSTER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
3	1145937	53.74	1145937	53.74
5	705613	33.09	1851550	86.84
2	280570	13.16	2132120	100.00
1	35	0.00	2132155	100.00
4	35	0.00	2132190	100.00

1. We calculated the R, F and M scores and mapped the combined scores to form 5 clusters. On drilling down to see Cluster vs RFM Score, we figured out that **Cluster 3 and Cluster 5** has high values of R, F and M compared to cluster 1,2 and 5.

2. Cluster 3 accounts for 53.74% of the total number of customers purchasing a ARM & HAMMER product and is our area of analysis as these are the customers who have transacted most recently and visit the store most frequently and spend large amounts to contribute towards the sales of ARM & HAMMER. These are the most loyal customers towards ARM & HAMMER.

We studied the demographics and other properties of these customers to make meaningful recommendations.

Sales Analysis across Cluster:

The MEANS Procedure

		Analysis Variable : DOLLARS				
Cluster	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	35	35	10.4160000	6.7262279	5.4800000	30.5400000
2	280570	280570	29.8573552	48.4172194	0.1900000	11165.36
3	1145937	1145937	28.5550550	39.6409321	0.1000000	10159.04
4	35	35	8.8042857	4.4233247	3.9900000	21.9200000
5	705613	705613	30.9768349	43.7792270	0.1900000	3744.54

Average total Sales in each of the 5 clusters are calculated below:-
- Cluster 1 : $10.4*35=$364

- Cluster 2 : $29.86*280570=$8377820
- Cluster 3 : $28.6*1145937=$32773798
- Cluster 4 : $8.8*35=$308
- Cluster 5 : $40*705613=$28224520

Here, we can see that **Cluster 3 accounts for the most sales** in dollars in ARM & HAMMER products, along with Recency, Frequency as well as Monetary.

Distribution of Age across Clusters:

The FREQ Procedure

Frequency Percent Row Pct Col Pct									
Table of CLUSTER by HH_AGE									
		HH_AGE							
CLUSTER(Cluster)	0	1	2	3	4	5	6	Total	
1	0	0	0	7	7	10	11	35	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	20.00	20.00	28.57	31.43		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
2	0	0	0	80300	39560	39344	121366	280570	
	0.00	0.00	0.00	3.77	1.86	1.85	5.69	13.16	
	0.00	0.00	0.00	28.62	14.10	14.02	43.26		
	0.00	0.00	0.00	17.76	7.43	12.08	16.40		
3	0	0	41416	286848	328293	203244	286136	1145937	
	0.00	0.00	1.94	13.45	15.40	9.53	13.42	53.74	
	0.00	0.00	3.61	25.03	28.65	17.74	24.97		
	0.00	0.00	50.69	63.43	61.66	62.41	38.66		
4	1	1	1	5	8	4	15	35	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	2.86	2.86	2.86	14.29	22.86	11.43	42.86		
	100.00	100.00	0.00	0.00	0.00	0.00	0.00		
5	0	0	40290	85085	164594	83036	332608	705613	
	0.00	0.00	1.89	3.99	7.72	3.89	15.60	33.09	
	0.00	0.00	5.71	12.06	23.33	11.77	47.14		
	0.00	0.00	49.31	18.81	30.91	25.50	44.94		
Total	1	1	81707	452245	532462	325638	740136	2132190	
	0.00	0.00	3.83	21.21	24.97	15.27	34.71	100.00	

1. Most of the customers from Cluster 3 belong to Age Band 4 which basically comprises people having age between 45 - 54' years. Hence, we can say that people within this age band has a greater tendency to choose ARM & HAMMER Powder detergent brand.

Problem Statement II

Find out the attributes that affect the brand of choice of the household panelists.

Analysis Method Implemented: Multinomial Logistic Regression
We have used the top 5 detergent brands in Powder Category in terms of market share
i.e., TIDE,GAIN,ARM&HAMMER,CHEER,SURF.We used multinomial regression to
study how the attributes like the income of the household,family size of the household,the
price per unit ounce of the brand,whether the brand was on display or featured or if it had
price reduction will affect the choice of the customer.

Insights:

Analysis of Maximum Likelihood Estimates						
Parameter	brand_final	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	ARM&HAMMER	1	4.9476	0.5551	79.4304	<.0001
Intercept	CHEER	1	-0.7413	0.5820	1.6224	0.2028
Intercept	SURF	1	1.1291	4.4932	0.0631	0.8016
Intercept	TIDE	1	-2.2045	0.4354	25.6382	<.0001
HH_Income	ARM&HAMMER	1	0.1474	0.0397	13.8151	0.0002

HH_Income	CHEER	1	0.1487	0.0427	12.1427	0.0005
HH_Income	SURF	1	0.1141	0.3618	0.0995	0.7525
HH_Income	TIDE	1	0.0831	0.0305	7.4353	0.0064
Family_Size	ARM&HAMMER	1	-0.4700	0.0990	22.5374	<.0001
Family_Size	CHEER	1	-0.4044	0.1063	14.4625	0.0001
Family_Size	SURF	1	-0.3861	0.7422	0.2706	0.6029
Family_Size	TIDE	1	-0.2292	0.0692	10.9751	0.0009
F_num	ARM&HAMMER	1	1.3559	0.5763	5.5362	0.0186
F_num	CHEER	1	0.3456	0.6869	0.2531	0.6149
F_num	SURF	1	-7.2999	338.2	0.0005	0.9828
F_num	TIDE	1	1.1116	0.5291	4.4145	0.0356
D	ARM&HAMMER	1	1.4968	183.0	0.0001	0.9935
D	CHEER	1	0.0852	229.1	0.0000	0.9997
D	SURF	1	2.4812	2057.4	0.0000	0.9990
D	TIDE	1	5.0381	131.7	0.0015	0.9695
PR	ARM&HAMMER	1	-0.9297	0.4723	3.8755	0.0490
PR	CHEER	1	0.0573	0.5064	0.0128	0.9098
PR	SURF	1	-9.3058	293.5	0.0010	0.9747
PR	TIDE	1	-0.0666	0.4087	0.0265	0.8706
PPU	ARM&HAMMER	1	-74.1932	5.7978	163.7576	<.0001
PPU	CHEER	1	-0.6699	4.4981	0.0222	0.8816
PPU	SURF	1	-110.2	73.4656	2.2483	0.1338
PPU	TIDE	1	32.1155	3.3008	94.6662	<.0001

OBSERVATIONS:

1. The preference of the customers in the brands is ARM & HAMMER> CHEER = SURF =
GAIN>TIDE (GAIN is the base brand.

2. As the income of the household increases by 10k USD,the odds of choosing ARM & HAMMER compared to GAIN brands will increase by 15.88%.

3. As the number of family members increases by 1, the odds of choosing ARM & HAMMER compared to other brands will decrease by 37.5% [(exp(0.47)-1)*100)]

4. The display does not play a role whether customer prefers ARM & HAMMER or not(insignificant value)

5. As the price per unit ounce of a brand increases, the odds of choosing ARM & HAMMER decreases.

6. Price elasticities:
Price coefficient: -74.1932 mean: 0.0482 ms(p) = 0.1074
Own price elasticity: (1-0.1074)*(-74.1932)*0.0482 = -3.189
If ARM & HAMMER decreases its price by 1 % then its share would decrease by 3.19%
Cross Price elasticity of ARM & HAMMER w.r.t. TIDE
Price coefficient = 32.1155, mean = 0.094, ms(P) = 0.6108
Cross Price elasticity = -0.6108*32.1155*0.094 = -1.84%
If TIDE were to increase prices by 1%, ARM & HAMMER's share would increase by 1.84%

Problem Statement III
Find the most effective method of Marketing Communication from Types of Features and Display and Price Reduction.

Dependent Variable: TotalSales_w

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	1446
Time Series Length	52

Fit Statistics			
SSE	175600522.5	DFE	64444
MSE	2724.8545	Root MSE	52.2001
R-Square	0.6544		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1496	64444	73.04	<.0001

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	124.1473	9.7924	12.68	<.0001	Intercept
AvgPrice_w	1	-853.417	72.4375	-11.78	<.0001	
AvgDisplay_w	1	180.392	3.1069	58.06	<.0001	
AvgFeat_coupon_w	1	105.5424	4.2679	24.73	<.0001	
AvgFeat_large_w	1	88.92109	2.2978	38.70	<.0001	
AvgFeat_medium_w	1	35.93026	2.1890	16.41	<.0001	
AvgFeat_small_w	1	10.70866	3.5023	3.06	0.0022	
AvgPriceRed_w	1	-1.44778	1.1563	-1.25	0.2105	

Analysis Method Implemented: Panel Regression

We have the data that portrays the different purchases by consumers along with the information of types of promotions (features and display) going on at that time. The panel regression thus is an apt method to determine which of the promotion methods and marketing communications were most effective to influence the consumer buying pattern.

Output:
1) We ran Hausman Test as P value was <0.05 so we rejected the null hypothesis and ran fixed effect model.
2) Here we can see that amongst the features and display promotions, Display are the most effective mode of marketing communications affecting the total sales in the most positive manner followed by Coupon, large advertisements, medium advertisements and small advertisements. We can also assess that an increase in the price reduction is insignificant on total sales.
Insights and Interpretations:

1. The regression output shows that attributes like types of advertisements for the products, and whether a Product is displayed- all these factors play a significant role in determining the sales of the product.
2. Compared to no coupons, A product advertised through coupons in a week in a store is expected to generate $ 105.54 more sales.
3. Compared to no advertisement, a product having a large advertisement in a week in a store is expected to increase total sales by $88.92.
4. Compared to no advertisement, a product having a medium advertisement in a week in a store is expected to increase total sales by $35.93.
5. Compared to no advertisement, a product having a small advertisement in a week in a store is expected to increase total sales by $10.70.
6. If a product is displayed at the store, the total sales in a week is expected to increase by $180.39 compared to the case when it is not displayed in that week in that store.
7. Effect of price reduction in a week in a store is insignificant.

Recommendations:
1. ARM & HAMMER brand should focus on Display features to be marketed to the best customers (age group 45-54, working in Professional or Technical) to get highest increase in sales.
2. The households with higher income are preferring the ARM & HAMMER brand. Hence the management should come up with The products that target this section of customers.

3. In exploratory analysis we observed that some of the market locations such as New york,Buffalo/Rochester,Oklahama city,West texas ,Houston have annual sales less than 10k so management should find out the effective ways of promoting the brand to increase the sales in these regions through display advertisement