BUAN/MKT 6337.001 – Predictive Analytics using SAS
Group 4 – Project report

Scanner Data Analysis

Fiza Gupta
Priyanka Savant
Kaustubha Madhu

Introduction

Doritos is an American brand of tortilla chips produced by Frito-Lay, which became a wholly owned subsidiary of PepsiCo in 1965. Doritos sells products internationally and is one of the top ranked tortilla chips brand in the United States. It offers a wide breadth of flavors and multiple product lines to satisfy consumer needs and promote market growth. Data we analyzed is based on sales of consumer goods obtained by 'scanning' the bar codes for individual products at electronic points of sale in retail outlets.

Objective

In this project, our focus is to perform brand analysis for Doritos and interpret the models that we are running on the panel data set. The goal is to understand the customers and come up with marketing recommendations for the brand. Also, give actionable insights and provide recommendations to the manager in taking actions to increase the sales and market share.

We performed three analysis to come up with the recommendation for the improvement in the market share
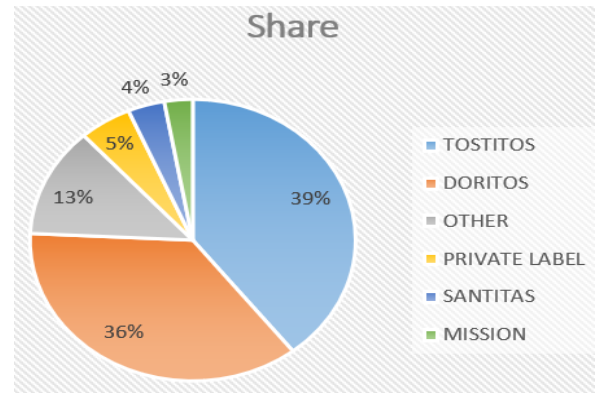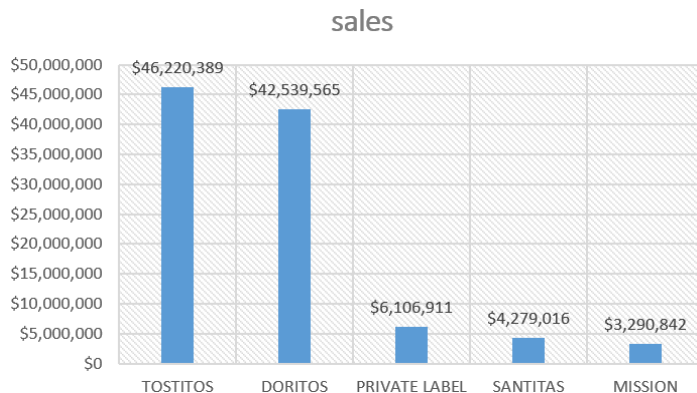1. Fixed Effects Model
2. Multinomial Logit
3. Segmentation through Recency, Frequency and Monetary

Exploratory Data Analysis

***We are analyzing tortilla chips as Doritos is belong to that category

Sales of top 6 Brands
Other brands include all the remaining brands that are not in the top 5 selling brands

The above bar chart displays the sales of top 5 brands in grocery stores. According to this chart, Tostitos leads the market with total sales of approx. $46M in a year 2001 followed by Doritos with total sales of around $42M. There is a $3.6M difference between the sales of Tostitos and Doritos. After that there is sudden drop in the sales of around 85%. Which means in the Tortilla chips category these two are the leading brands. Remaining 3 brands have almost same kind of sales with around $4M.

The pie chart above gives us a visual representation of the percentage of market share for the top 6 brands. Tostitos has the highest market share of 39% followed by Doritos with a market share of about 36%. The total % of market share for all other remaining brands comes around 25%.

So the point that we get from here it is important for the Doritos brand to sustain its position and follow the strategies which help them and secondly there is a room for them to grow as it is the second highest selling brand so it can come to first brand if taken proper marketing strategies.

Top Flavor chosen By People
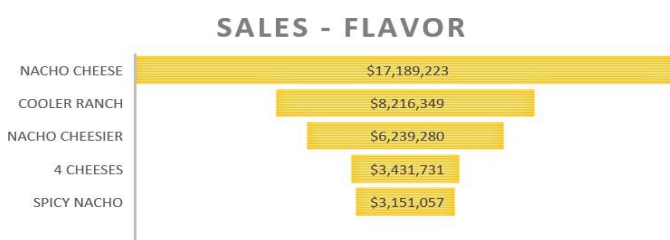
Nacho cheese flavor scent is most demanding and has the highest sales of $17.2 M and then around 50% drop of sales in the second flavor scent that is cooler ranch and has sales of $8.2 M. All the remaining brands have a difference not as large as the first two flavors. Third highest sold flavor is Nacho cheese and the remaining with the almost same sales of $3M are 4 cheeses and Spicy Nacho. The difference in all the sales of all flavor are statistically significant

From the above analysis we can judge the cheese scent is liked more and sells most

Top Selling Locations

In Los Angeles, Doritos has the highest sale of $2.8 Million. From the graph, one can notice there is not a sudden drop but a gradual decrease in sales with <$ 1M. The second highest sales happened in New York of $2 Million and then Chicago with $1.6 Million.

All the differences in sales are statistically different from each other. The difference in sales is not because of population. After checking the population in 2001 Los Angeles has lower population than New York.

Sales in Millions

Los Angeles people preferred Doritos because for the Tostitos this place is the 5th place. Otherwise all countries have the same order. So I am sure Los Angeles is the place where my loyal customers are more.
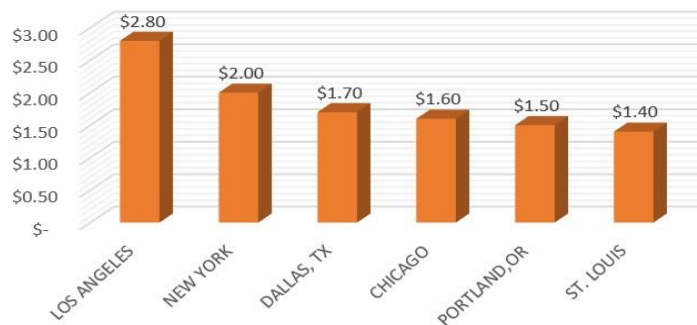
Sales and Prices per Unit over Time
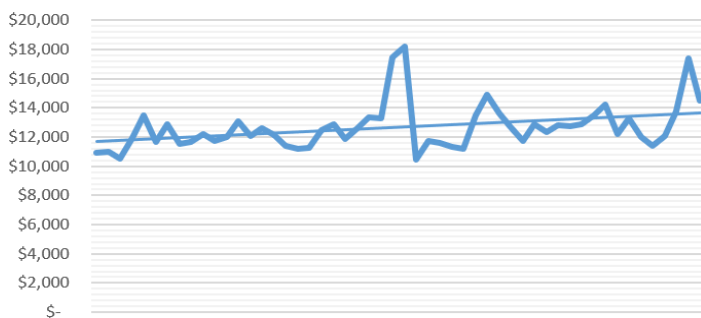

Mean Price Per Unit Fluctuations

Drug Store: Prices Per Unit Decreases overtime(-ve trend)

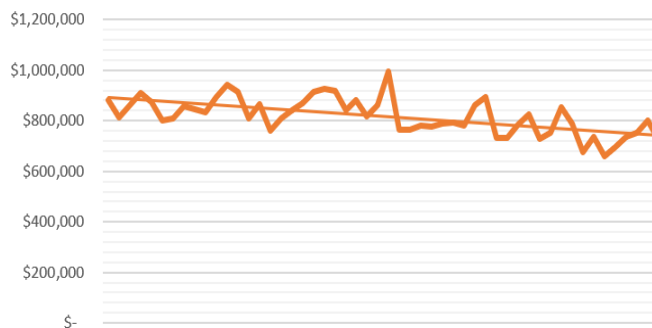Drug Store: Total Sales Increases overtime(+ve trend)

Grocery Store: Prices Per Unit Increases overtime(+ve trend)

Grocery Store: Total Sales decreases overtime(-ve trend)

### Sum sales in Drug Strore

Sum sales in Grocery Store

The Average Price per unit per week in Drug store is slighly reducing as we can see the -ve slop and for the grocery store it's increasing. But the gap between the prices of the two stores are reducing. Might be they want to have same price range for people because huge price difference lead people to buy from cheaper stores. That might be the reason why sales of grocery stores increases and drug stores decrease. Over the year price is continously fluctuating.
This is a very tricky point here thinking of price reduction or giving discount might attract people in that store but at the same time we can notice that though the people move to drug store but the average price of the drug store still higher than the grocery store. Leading company to still make more profit
Overall the sales are maximum in the summer season in both stores and has lowest prices. Again the sales peakes in Decemeber during the festival season and prices are reduced during this period .

This shows how much price is important because decreasing price in drug store is increasing the sales and opposite is seen in grocery store (customer are price sentitive).

Doritos/Tostitos
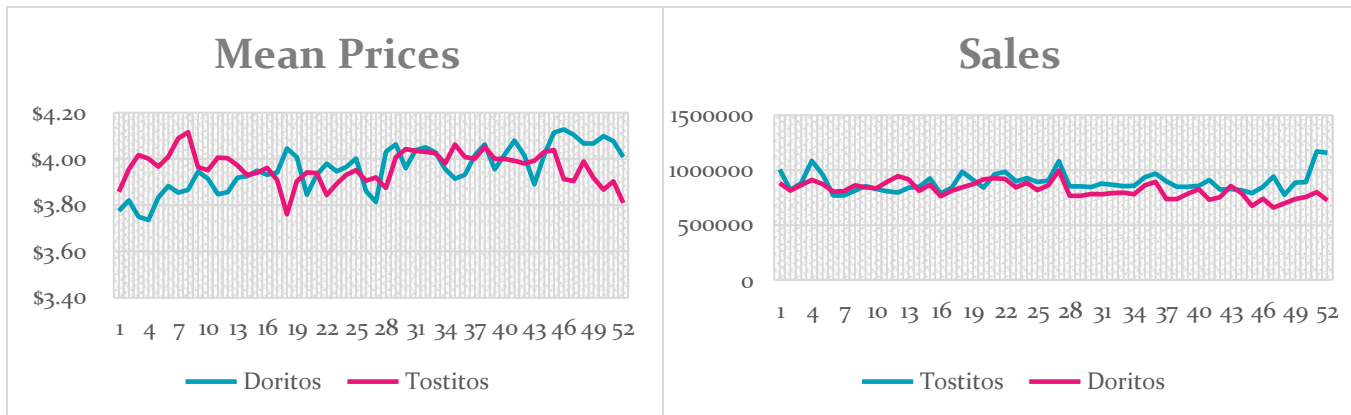
We can see the competition in the prices going on most of the time if one is increasing other are decreasing the prices. But the sales are kind of same for both except at the end the sales dropped for the doritos migh be the reason described above. In both casses we can see customers do care about the price lowering the prices effect the sales. So doritos has to take care about the prices of its competiotion

Problem Statement I

Understand the effect of Features, Price Reduction and Display on the sales of Doritos in grocery stores.

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	1448
Time Series Length	52

Fit Statistics			
SSE	5480194259	DFE	66522
MSE	82381.6821	Root MSE	287.0221
R-Square	0.6723		

Hausman Test for Random Effects			
Coefficients	DF	m Value	Pr > m
7	7	62.29	<.0001

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
1498	66522	79.68	<.0001

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	616.8742	60.6887	10.16	<.0001	Intercept
AvgPrice	1	-39.0436	5.7567	-6.78	<.0001	
AvgDisplay	1	377.7069	6.6965	56.40	<.0001	
Avg_f_plus	1	177.8393	30.5277	5.83	<.0001	
Avg_f_large	1	82.62196	25.4528	3.25	0.0012	
Avg_f_none	1	-229.672	25.4275	-9.03	<.0001	
Avg_f_medium	1	-1.59126	25.6971	-0.06	0.9506	
avg_Price_Reduced	1	-7.34542	9.4714	-0.78	0.4380	

Methodology Adopted: Panel Regression

We have the data based on types of salty snacks purchased during different interval of time, there can correlation between same stores. In data we have the information of promotion types i.e. features(F) and display(D) which is going on at a given time. Therefore, we decided to use panel regression to analyze the promotion type that are most effective in determining sales

Output:
1) Since the Hausman test is significant as P value is <0.05, we rejected the null hypothesis for the random effects model and ran fixed effect model
2) Price Elasticity: 1% increase in the price can reduced the average sales by 5.7%
3) Among all the different kind of features, Feature A+ (coupon or retable) has most effect on the sales. Order based on effectiveness of features is A+ (coupons) > A large advertisement > B = C (medium and small advertisements) > None

Interpretations:
1) Price reduction do not play a significant role in determining the sales of Doritos
2) When compared to small advertisement, a product that has large advertisement in a store then that store is expected to increase their total sales by $82.62
3) When compared to small advertisement, a product that has A+, use coupons, then that store is expected to increase their average sales by $177.83

4) When compared to small advertisement, a product that has no advertisements that store is expected to decrease their average sales by $229.67
5) If a product is displayed at the store, we see that the total sales are expected to increase by $377 in a week as compared to when it is not displayed during that week
6) When we increase the price of a product by one dollar, the sales go down by $39. Therefore, increasing the price of the product would result in reduction of sales.

We ran the same model for the TOSTITOS and found that all coefficients effects sales more for Tostitos for example doing a large advertisement will affect Tostitos sales more than the Doritos sales except for the Display and the feature None.
So, having a Doritos chips on display effect sales more for Doritos then for Tostitos and not have feature will drop sales more for Doritos then for Tostitos. Doritos band having feature and display is most important to beat Tostitos

Problem Statement II

Find the attributes that affect the choice of the customer(panelists) to choose brands

In this we are trying to analyze how the feature, display, price reduction and their interaction can affect chances of choosing brand.
We are choosing top brand DORITOS, TOSTITOS, SANTITAS, PRIVATE LABEL and OTHERS (taken as reference brand).

Feature	DORITOS	1	2.1363	0.2237	91.2247	<.0001
Feature	PRIVATE LABEL	1	1.8496	0.1885	96.3158	<.0001
Feature	SANTITAS	1	-12.8238	157.0	0.0067	0.9349
Feature	TOSTITOS	1	1.5157	0.2282	44.1243	<.0001
D	DORITOS	1	0.7731	0.0361	459.7125	<.0001
D	PRIVATE LABEL	1	-0.3587	0.0453	62.5637	<.0001
D	SANTITAS	1	-0.6318	0.0654	93.2593	<.0001
D	TOSTITOS	1	0.5616	0.0373	226.6503	<.0001
PR	DORITOS	1	1.6571	0.0754	482.5924	<.0001
PR	PRIVATE LABEL	1	-2.3966	0.0974	605.5530	<.0001
PR	SANTITAS	1	-2.7540	0.1218	510.8808	<.0001
PR	TOSTITOS	1	1.2038	0.0768	245.5836	<.0001
PPU	DORITOS	1	8.2829	0.1520	2967.7357	<.0001
PPU	PRIVATE LABEL	1	-3.3231	0.3126	112.9934	<.0001
PPU	SANTITAS	1	219.4	8.3361	692.8832	<.0001
PPU	TOSTITOS	1	8.3568	0.1545	2926.7159	<.0001
F_pr	DORITOS	1	1.4901	0.2333	40.8060	<.0001
F_pr	PRIVATE LABEL	1	-1.4384	0.2298	39.1740	<.0001
F_pr	SANTITAS	1	3.8305	157.6	0.0006	0.9806
F_pr	TOSTITOS	1	0.9254	0.2398	14.8987	0.0001
Price_sq	DORITOS	1	-0.9030	0.0199	2059.6581	<.0001
Price_sq	PRIVATE LABEL	1	0.2135	0.0696	9.4127	0.0022
Price_sq	SANTITAS	1	-61.6438	2.3351	696.9033	<.0001
Price_sq	TOSTITOS	1	-0.9088	0.0200	2066.5864	<.0001

McFadden Rsq: 39.5%

Price have the non-linear effect for 2 major brands

Own Price Elasticity of Doritos: 1% increase in price of Doritos will lead to 3.81% increase in probability in choosing brand Doritos

Cross Price Elasticity of Doritos: increasing price of Doritos by 1% the probability of choosing other brand decreases by 4.79%

Own Price Elasticity of Tostitos: 1% increase in Tostitos price will lead to 4.82.% increase in probability in choosing brand Tostitos

Cross Price Elasticity of Tostitos: increasing price of Tostitos by 1% the probability of choosing other brands decreases by 1.8%

Own Price Elasticity of Private Label: 1% decrease in price private label will lead to 3.7% increase in probability in choosing brand Private Label

Cross Price Elasticity of Private label: increasing price by 1% of Private Label the probability of choosing other brand increases by .3%

1. **Feature:** all the coefficients are significant except for the SANTITAS. So, having features when compared to not having features makes it more likely to buy Doritos, Tostitos and private label as compare to Other brands. And if accompanied by price reduction will increase the chances more of preferring these brands because of significant interaction between them.
2. **Display**: all the coefficients are significant. Having a brand on display will lead to more chances of buying Doritos and Tostitos compared to when not on display. But for SANTITAS and private labels the preference decreases compared to Other brands
3. **Price Reduction:** all the coefficients are significant. Having a discount/price reduction on brands leads to more probability of buying Doritos and Tostitos compared to Other brands. But for SANTITAS and private labels the preference decreases compared to Other brands
4. **Price per Unit:** all the coefficients are significant. Having a higher price makes it more likely to buy Doritos, and Tostitos compared to Other brands. But for SANTITAS and private labels the preference decreases compared to Other brands. Though adding price to high value may be because of adding quality or certain new features that make people buy more but this effect is up to certain point as we can see there is a non-linear effect after a certain limit of price increase the chances of choosing Doritos and Tostitos will start decreasing which makes sense. After certain price limit is exceed, we opt other brands. So, if we keep on adding price we might stop buying, therefore price cannot have linear effect. So, increasing price directly not affect the preference of Doritos but valid up to certain price

In the below interpretation we will be comparing Doritos to Tostitos because these two are the leading brands. We will get the same results as above but to be more assured here is the calculations below:

	Analysis of Maximum Likelihood Estimates					
Parameter	Final_Brand	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	DORITOS	1	-0.1858	0.1007	3.4048	0.0650
Feature	DORITOS	1	0.6147	0.0924	44.2823	<.0001
D	DORITOS	1	0.2278	0.0208	119.5064	<.0001
PR	DORITOS	1	0.4533	0.0509	79.4121	<.0001
PPU	DORITOS	1	-0.0298	0.0249	1.4362	0.2308
F_pr	DORITOS	1	0.5784	0.1062	29.6675	<.0001

- If the brand has a feature, the odds of choosing DORITOS is 1.84 times the odds of choosing Tostitos
- If the brand is on display, odds of choosing Doritos is 1.22 times of choosing Tostitos
- Having a price reduction, odds of choosing Dorito is more than odds of choosing Tostitos. Because interaction term is significant between feature and price reduction the chances/probability of choosing Doritos further increases
- Increase in price per unit will have no effect on choosing brands
- Own Price Elasticity of Doritos: 1% decrease in price of Doritos will lead to .04% increase in probability in choosing brand Doritos
- Cross Price Elasticity of Doritos: if Doritos increases price by 1% the probability of choosing other brand increases by .05%

So for customers to prefer our brand to its competitive its important that our brand should be on display, have feature and having price reduction will benefit more
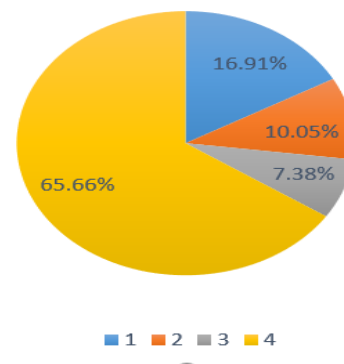
Problem Statement III
Find out who are the best group of customers for the Brand Doritos

Analysis Method Implemented: RFM Clustering
We studied several attributes of the customers to unfold their buying patterns and clustered them into 5 segments based on recency, frequency, and monetary value of their purchase behavior. After customer segmentation was done, further studies from the segmented data helped us to derive critical insights and come up with meaningful recommendations. In this case, we identified the customer's purchase patterns for Doritos only.

	Cluster Means		
Cluster	recency_score	frequency_score	monetary_score
1	1.708061002	3.241830065	3.239651416
2	1.715116279	1.309108527	1.363372093
3	3.223602484	2.549689441	1.776397516
4	3.514736842	3.594736842	3.580000000

Pie chart: 16.91%, 10.05%, 7.38%, 65.66% — 1 2 3 4

Groups	Description	cluster	Marketing
Best Customers Cluster 4	these are the customers who frequently buy, have recent purchase, and have high monetary value	4 - 62.7%	these customers have a habit of coming to our store, no proper strategy is required. But can be given rewards once throughout a year to make them satisfied with our brand and sustain their loyalty
Future Potential Customer Cluster 3	These are the customers who have recent purchase but have less frequency and monetary value	3-7.3%	These customers are still not loyal. So, giving them offers can lead them to become a loyal customer
Risky Customers Cluster 1	These are customers who have high frequency score, high monetary value but are not recent and they might have gone for other options	1-16.9%	These might be customers who have moved to other options. Hence, advertising for them, giving them offers and telling them that the company misses them might lead them to get back
Worst Customers Cluster 2	have least frequency, recency, and monetary value	2-10.5%	We feel that we should not waste money on them, as they are least likely to come and buy our brand

So out market of loyal consumers is 62.7% but about 16% customers are risky to move to other brands, so proper action must be taken to attract them

What kind of customer comes to us (so that we can target them)?

1) Income: it has a kind of u shape structure. Consumers with income of $35k-$45k are the one who buys most (median level income buys most of our brands. It means discounts might matter to them. before this income sales decreases and even above this income that is why in u shape

2) Children age: very interesting most buyers are the one who do not have children. And second most buyers were who have children age between 12-17 and least buyers whose children age is between 0-5

3) Occupation: most buyers has profession of technical so not high-level job and that is why the income is in the median level of high buying customers

4) Customers between the age group of 45-54 buys most and this is because at this age they handle all household expenses and do the most grocery. So, might be due this reason they do the most shopping and buy more

5) Education Most buyers are Graduated high school students. Education has kind of non-linear effect. As education increases, they buy less. Might be that is the reason they do small jobs and earn less

6) House owners buy more than whose house is rental

7) Working hours: both for men and women working full times >35 hours buy most and that too many a times than any kind of job hours

So loyal customers are one who do not have children, age 45-54 earn median level pay working full time with education high school graduated. So main consumers are not high earning people. It is kind of middle earning people buy most (might be a selection bias the store only middle earn people come more)

Recommendations:

1) As most of our loyal buyers are average earning people giving offers like using coupons – most effective feature – will increase sales to a great extent and especially if they don't have children. Focus on these customers.

2) We have seen how important display and feature adding is to Doritos for its sales and to prefer customers Doritos more than Tostitos. Having no feature will heart the sales allot so to beat the Tostitos sales, display with advertisement will be very effective.

3) Customers who are of future potential or who are still not loyal to the brand can be focused by using coupon features

4) Reducing price increases sales but price should be monitored with respect to its competitor Tostitos and should be in line with it. if we increase price but less than Tostitos can still increase sales of our brand

5) Doritos should try to attract customers in the age group of 45-54 as they are the most frequent buyers of the product

6) Nacho cheese, Ranch and nacho cheesier flavour scent contribute to 82% of total Sales, so it is important to have right advertising and display strategy for these variants particularly