PREDICTIVE ANALYTICS USING SAS

BUAN 6337.003

FINAL PROJECT REPORT

GROUP 9

NAVIN KARTHIK MURUGIAH (nxm171030)

DEEPAK ANTONY (dxa171830)

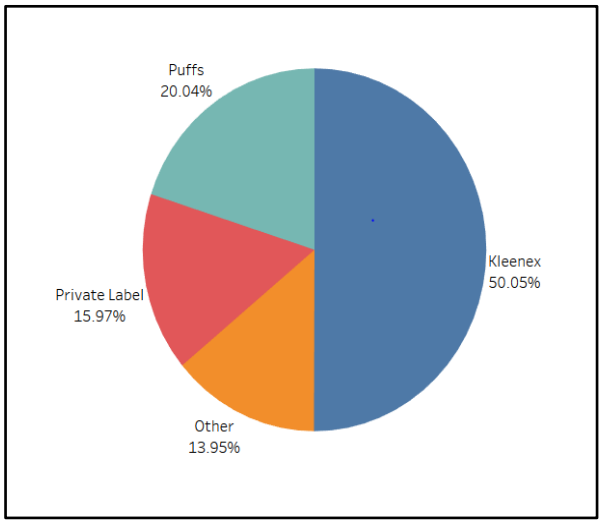
ARAVINDH SESHADRI (axs170332)

MAHAARAJAN MURUGAN GUNASEKARAN (mxm176530)

SWAMINATHAN SADANANDAM MULLAIVANANATHA (sxs172932)

INTRODUCTION

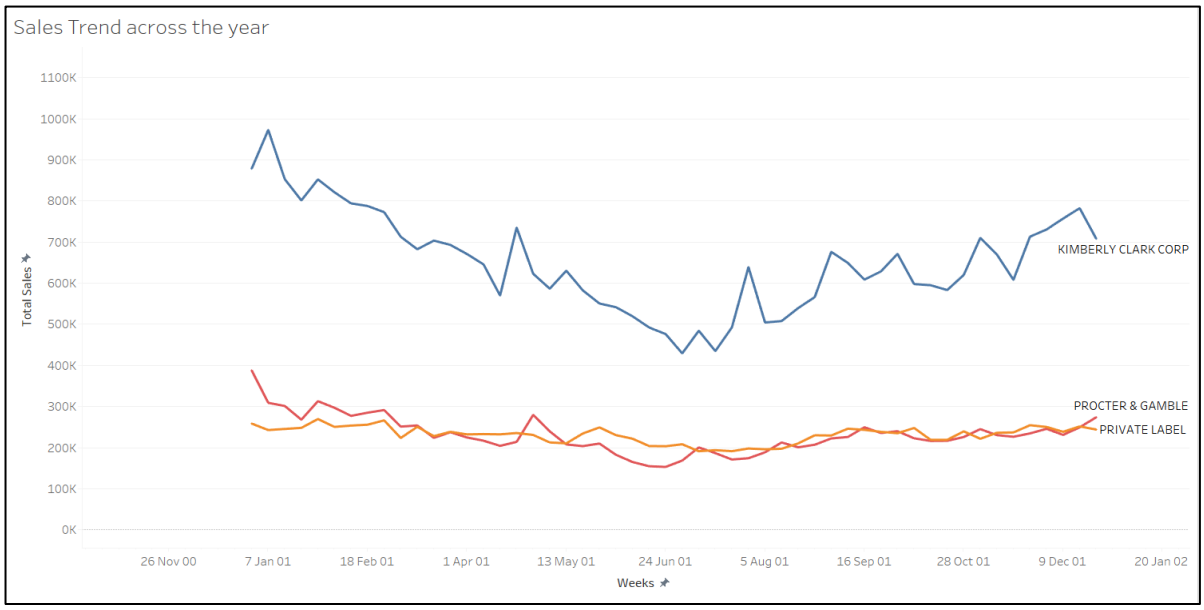
Facial tissue is one of the most commonly used products in the United States. Among all brands of facial tissues in the market, Kleenex is a pioneer with a relatively high market share. From the perspective of the brand manager of Kleenex, we have captured various insights in this report which would help Kleenex further consolidate its position in the market. We have also utilized analytical techniques to understand the prevailing market conditions of Kleenex facial tissues.



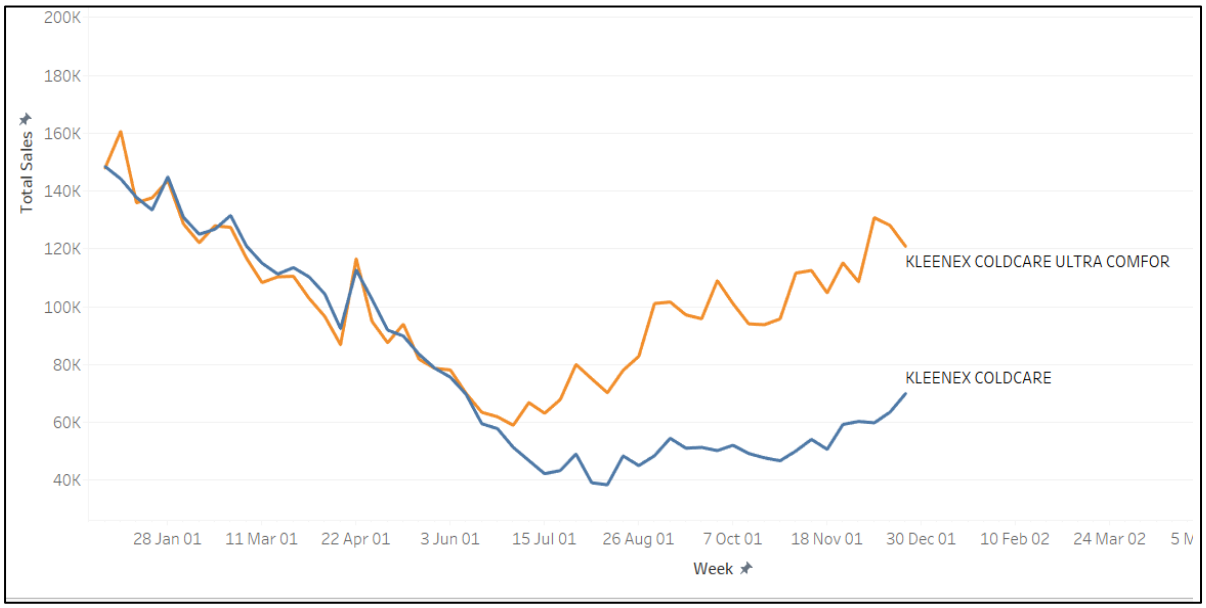
SALES TREND ACROSS WEEKS FOR THE YEAR 2001

To observe the sales trend across weeks for the year 2001 in both drug and grocery stores, we considered the top three brands in the market which are as follows,

- Kleenex
- Puffs
- Private Label



It can be clearly observed that Kleenex has the highest market share. Sales trend seems to increase at the start of winter season for Kleenex tissues. This may be because people are more prone to cold/flu during the winter season.



This is reiterated by the fact that sales of Kleenex variants - “Coldcare” and “Coldcare Ultra Comfort” increases from August and slowly drops as summer season approaches.

BUSINESS OBJECTIVES

- Recommend promotional strategies based on price, display and feature elasticities.
- To segment customers based on monetary value.
- To study brand choice preference of customers and understand customer characteristics.

PROBLEM STATEMENT - 1

Apply panel regression

- To determine the effects of price, display, feature and price reduction on the sales of Kleenex tissues.
- To determine price elasticity, display elasticity and feature elasticity of Kleenex tissues.
- To determine the effects of cross price elasticity and cross display elasticity of Puffs tissues on sales of Kleenex tissues.

Data Preparation:

Data of 1217 stores which sold Kleenex tissues for all 52 weeks were considered to get a balanced panel dataset.

Since the count of tissues may vary with respect to different UPCs across Kleenex brand, number of sheets was considered instead of number of units to ensure standardization. The number of sheets in a unit is calculated as follows.

Number of sheets in a unit = (Volume Equivalent corresponding to that unit) * 100

Average price per sheet was calculated for store, week and UPC combination.

To run panel regression, data had to be aggregated for store, week combination. For this purpose, the categorical variables such as display, feature and price reduction had to be made continuous by multiplying these variables with a weight.

Weighted Price is calculated as follows

Weighted Price =

$$\left(\frac{\text{Number of sheets of a Kleenex product sold in a week at a store}}{\text{Number of sheets of all Kleenex products sold in the same week at the same store}}\right) \times \text{Average Price Per Sheet of that product}$$

Weighted Display is calculated as follows

Weighted Display =

$$\left(\frac{\text{Number of sheets of a Kleenex product sold in a week at a store}}{\text{Number of sheets of all Kleenex products sold in the same week at the same store}}\right) \times \text{Display Indicator of that product}$$

Weighted Feature is calculated as follows

Weighted Feature =

$$\left(\frac{\text{Number of sheets of a Kleenex product sold in a week at a store}}{\text{Number of sheets of all Kleenex products sold in the same week at the same store}}\right) \times \text{Feature Indicator of that product}$$

Weighted Price Reduction is calculated as follows

Weighted Price Reduction =

$$\left(\frac{\text{Number of sheets of a Kleenex product sold in a week at a store}}{\text{Number of sheets of all Kleenex products sold in the same week at the same store}}\right) \times \text{Price Reduction Indicator of that product}$$

After these calculations, data was aggregated for store, week combination.

Multicollinearity Check:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	92253	1430.12498	64.51	<.0001	0
WEIGHTED_PRICE	1	-5052922	115034	-43.93	<.0001	1.34563
WEIGHTED_FEATURE	1	44405	1157.96623	38.35	<.0001	1.91269
WEIGHTED_DISPLAY	1	54648	1552.06750	35.21	<.0001	1.12929
WEIGHTED_PRICE_REUDCTION	1	-22864	974.55730	-23.46	<.0001	1.98741

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variation				
			Intercept	WEIGHTED_PRICE	WEIGHTED_FEATURE	WEIGHTED_DISPLAY	WEIGHTED_PRICE_REUDCTION
1	3.11611	1.00000	0.00192	0.00196	0.02384	0.02934	0.02355
2	0.98538	1.77830	0.00502	0.00815	0.11015	0.12254	0.02675
3	0.65254	2.18525	0.00000835	0.00013011	0.10829	0.84523	0.05300
4	0.23324	3.65516	0.00057691	0.00187	0.73104	0.00000432	0.79170
5	0.01273	15.64455	0.99247	0.98788	0.02668	0.00288	0.10499

For checking the presence of multicollinearity VIF and Condition Index were calculated. VIFs were not over 10 and Condition Index was not over 100 for any of the variables. These checks do not indicate the presence of multicollinearity among the variables.

Panel Regression:

- **Hausman Test for Random Effects**

The Hausman test is done to determine whether fixed effects model or random effects model is to be used for panel regression

NULL Hypothesis (H_0):

No Correlation between the error term (u_i) and the independent variables in the model. Random effects model is to be used

ALTERNATE Hypothesis (H_a):

Correlation is present between the error term (u_i) and the independent variables in the model. Fixed effects model is to be used

Hausman Test for Random Effects			
Coefficients	DF	m Value	Pr > m
8	8	455.28	<.0001

Since p-value of Hausman test is less than 0.05, there is convincing evidence to reject null hypothesis in favour of alternate hypothesis. Thus, fixed effects model needs to be used.

- **Panel Regression Model (Two Way Fixed Effects):**

Dependent Variable:
SHEETS SOLD PER WEEK AT A STORE

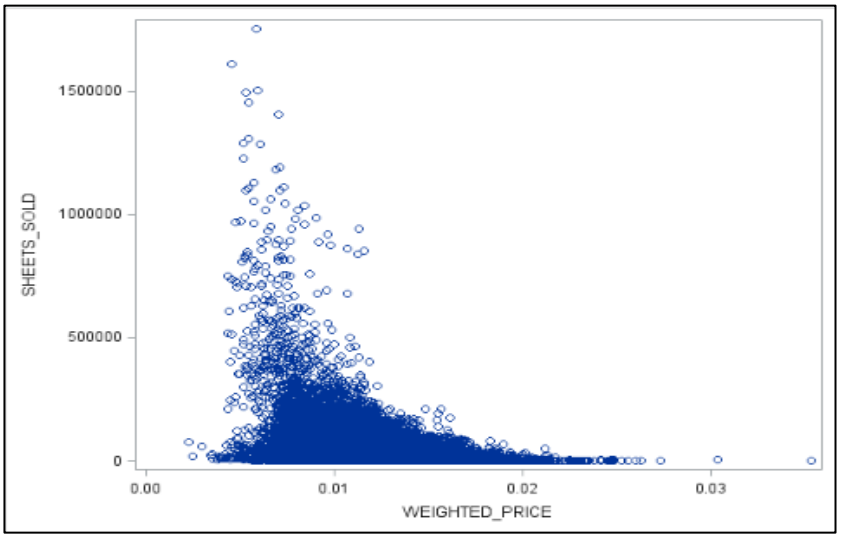
Independent Variables:
WEIGHTED PRICE PER SHEET, SQUARE OF WEIGHTED PRICE PER SHEET, WEIGHTED DISPLAY, WEIGHTED FEATURE, WEIGHTED PRICE REUDCTION, WEIGHTED FEATURE * WEIGHTED PRICE REUDCTION, WEIGHTED DISPLAY * WEIGHTED PRICE REUDCTION, WEIGHTED DISPLAY * WEIGHTED FEATURE * WEIGHTED PRICE REUDCTION

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	159494.2	6005.0	26.56	<.0001	Intercept
WEIGHTED_PRICE	1	-2.301E7	495013	-46.47	<.0001	
WT_PRICESQ	1	8.2992E8	18919024	43.87	<.0001	
WEIGHTED_DISPLAY	1	45859.38	1856.9	24.70	<.0001	
WEIGHTED_FEATURE	1	54302.72	1789.0	30.35	<.0001	
WEIGHTED_PRICE_REUDCTION	1	6006.493	811.4	7.40	<.0001	
WTFEATURE_WTPRRED	1	-45862.5	2532.8	-18.11	<.0001	
WTDISP_WTPRRED	1	-42295	3935.9	-10.75	<.0001	
WTDISP_WTFEATURE_WTPRRED	1	123099	4451.2	27.66	<.0001	

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
SHEETS_SOLD	63284	38141.50	58608.61	45.0000000	1753023.00
WEIGHTED_PRICE	63284	0.0113843	0.0022203	0.0022976	0.0353333
WEIGHTED_DISPLAY	63284	0.0601911	0.1507541	0	1.0000000
WEIGHTED_FEATURE	63284	0.1266824	0.2629678	0	1.0000000
WEIGHTED_PRICE_REUDCTION	63284	0.2406279	0.3185022	0	1.0000000

Check for Nonlinearity



Square term of weighted price in the panel regression equation is significant indicating the presence of non-linearity. The below residual plot of weighted price against total number of sheets sold confirms the presence of non-linear relationship between weighted price and sheets sold.

- **Effects of Price, Display, Feature and Price Reduction on Sales**

  - **WEIGHTED PRICE PER SHEET:**

    Nonlinear relationship between WEIGHTED PRICE PER SHEET and SHEETS SOLD PER WEEK AT A STORE indicates that sheets sold will decrease as price increases up to a certain price. Then the sheets sold will start to increase, as the price increases.

  - **WEIGHTED DISPLAY:**

    For every 1% increase in WEIGHTED DISPLAY, SHEETS SOLD PER WEEK AT A STORE increases by 54304 keeping other explanatory variables constant.

  - **WEIGHTED FEATURE:**

    For every 1% increase in WEIGHTED FEATURE, SHEETS SOLD PER WEEK AT A STORE increases by 45859 keeping other explanatory variables constant

  - **WEIGHTED PRICE REDUCTION:**

    If WEIGHTED PRICE REDUCTION increases by 1%, SHEETS SOLD PER WEEK AT A STORE increases by 6006 keeping other explanatory variables constant

  - **Interaction between WEIGHTED FEATURE and WEIGHTED PRICE REDUCTION:**

    Interaction between WEIGHTED FEATURE and WEIGHTED PRICE REDUCTION produces a cancellation effect on SHEETS SOLD PER WEEK AT A STORE.

  - **Interaction between WEIGHTED DISPLAY and WEIGHTED PRICE REDUCTION:**

    Interaction between WEIGHTED DISPLAY and WEIGHTED PRICE REDUCTION produces a cancellation effect on SHEETS SOLD PER WEEK AT A STORE.

- **Interaction between WEIGHTED DISPLAY, WEIGHTED FEATURE and WEIGHTED PRICE REDUCTION**

  Interaction between WEIGHTED DISPLAY, WEIGHTED FEATURE and WEIGHTED PRICE REDUCTION produces a synergistic effect on SHEETS SOLD PER WEEK AT A STORE.

- **SELF ELASTICITIES:**

- **PRICE ELASTICITY:**

  Price Elasticity is given by

$$\text{Price Elasticity} = \frac{\partial(\text{SHEETS SOLD})}{\partial(\text{WEIGHTED PRICE})} * \frac{\text{Mean of WEIGHTED PRICE}}{\text{Mean of SHEETS SOLD}}$$

  = (-4113883.488) * (0.0113843/38141.50)

  = **-1.22**

  Price Elasticity of -1.22 indicates that Kleenex tissues are price elastic. Thus, one percentage increase in price per sheet may bring down the sales of facial tissues by 1.22%

- **DISPLAY ELASTICITY:**

  Considering only the individual effect of display

  Display Elasticity is given by

$$\text{Display Elasticity} = (\beta_{\text{WEIGHTED DISPLAY}}) * (\text{Mean of WEIGHTED DISPLAY}/\text{Mean of SHEETS SOLD})$$

  = (45859.38) * (0.0601911/38141.50)
  = **0.072**

  Considering the interaction effects of display and price reduction

  Display Elasticity is given by

$$\text{Display Elasticity} = (\beta_{\text{WEIGHTED DISPLAY}} + \beta_{\text{PRICE REDUCTION}}) * (\text{Mean of WEIGHTED DISPLAY}/\text{Mean of SHEETS SOLD})$$

  = (45859.38+6006.493) * (0.0601911/38141.50)
  = **0.086**

Display Elasticity is 0.072 considering the effect of only the display. On considering the effects of display and price reduction, display elasticity is 0.086. In both cases, Kleenex tissues were display inelastic. Thus, sales of Kleenex tissues is insensitive to its display.

- **FEATURE ELASTICITY:**

Considering only the individual effect of feature

Feature Elasticity is given by

Feature Elasticity =   ($\beta_{\text{WEIGHTED FEATURE}}$) * (Mean of WEIGHTED FEATURE/Mean of SHEETS SOLD)

= (54302.72) * (0.1266824/38141.50)

= **0.180**

Considering the interaction effects of feature and price reduction

Feature Elasticity is given by

Feature Elasticity =   ($\beta_{\text{WEIGHTED FEATURE}} + \beta_{\text{PRICE REDUCTION}}$) * (Mean of WEIGHTED FEATURE/Mean of SHEETS SOLD)

= (54302.72 + 6006.493) * (0.1266824/38141.50)

= **0.2**

Feature Elasticity is 0.180 considering the effect of only the display. On considering the effects of feature and price reduction, feature elasticity is 0.2. In both cases, Kleenex tissues were feature inelastic. Thus, sales of Kleenex tissues is insensitive to its feature ads.

- **<u>CROSS ELASTICITIES</u>**

For calculating cross elasticities, data of PUFFS brand is considered as it has the second largest market share after KLEENEX.

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	172284	6291.1	27.39	<.0001	Intercept
WEIGHTED_PRICE	1	-2.428E7	530028	-45.81	<.0001	
WT_PRICESQ	1	8.8054E8	20462471	43.03	<.0001	
WEIGHTED_DISPLAY	1	46418.12	1942.9	23.89	<.0001	
WEIGHTED_FEATURE	1	55596.68	1852.5	30.01	<.0001	
WEIGHTED_PRICE_REUDCTION	1	6050.822	840.2	7.20	<.0001	
WTFEATURE_WTPRRED	1	-47575	2625.4	-18.12	<.0001	
WTDISP_WTPRRED	1	-42292.1	4108.4	-10.29	<.0001	
WTDISP_WTFEATURE_WTPRRED	1	135279.2	4674.1	28.94	<.0001	
WEIGHTED_PRICE_PG	1	-364376	77837.4	-4.68	<.0001	
WEIGHTED_FEATURE_PG	1	-15.9887	912.1	-0.02	0.9860	
WEIGHTED_DISPLAY_PG	1	-2725.79	1147.5	-2.38	0.0175	
WEIGHTED_PRRED_PG	1	830.4426	653.9	1.27	0.2041	

The MEANS Procedure

Variable	N	Mean
SHEETS_SOLD	60495	38389.71
WEIGHTED_PRICE	60495	0.0113413
WT_PRICESQ	60495	0.000133351
WEIGHTED_DISPLAY	60495	0.0578660
WEIGHTED_FEATURE	60495	0.1245820
WEIGHTED_PRICE_REUDCTION	60495	0.2396853
WEIGHTED_PRICE_PG	60495	0.0134553
WEIGHTED_FEATURE_PG	60495	0.0689938
WEIGHTED_DISPLAY_PG	60495	0.0362001
WEIGHTED_PRRED_PG	60495	0.1781003

- **CROSS PRICE ELASTICITY:**

  Cross Price Elasticity is given by

  > Cross Price Elasticity =
  > ($\beta_{\text{WEIGHTED PRICE PG}}$) * (Mean of WEIGHTED PRICE PG/Mean of SHEETS SOLD of KLEENEX)

  = (-364376) * (0.0134553/38389.71)

  = **- 0.1277**

  Cross Price Elasticity of -0.1277 indicates that sales of KLEENEX tissues is price inelastic with respect to PUFFS. Thus, sales of KLEENEX tissues is insensitive to increase or decrease in prices of PUFFs tissues.

- **CROSS DISPLAY ELASTICITY:**

  Cross Display Elasticity is given by

Cross Display Elasticity =
($\beta_{\text{WEIGHTED DISPLAY PUFFS}}$) * (Mean of WEIGHTED DISPLAY PUFFS/Mean of SHEETS SOLD of KLEENEX)

= (-2872.52) * (0.0362001/38389.71)

= **-0.0027**

Display Elasticity of -0.0027 indicates that sales of KLEENEX tissues is display inelastic with respect to PUFFS. Thus, sales of KLEENEX tissues is insensitive to display of PUFFs tissues.

PROBLEM STATEMENT – 2

- Determine the most valuable customers of Kleenex by segmenting them using RFM technique.
- Analyse the demographic characteristics of these customers to gain insights.

RFM analysis is performed on the panelists of Grocery, Drug and MA stores. From the purchase information of each panelist across weeks for Kleenex tissues, the following metrics were calculated.

- Recency – Difference between last purchase date and most recent date in the dataset for each customer.
- Frequency – Count of number of weeks of purchase made by each customer in the dataset.
- Monetary – Total dollar amount paid by each customer in the dataset.

The CORR Procedure

3 Variables:	MONETARY FREQUENCY RECENCY

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
MONETARY	3009	13.63902	16.26091	41040	0.56000	320.66000
FREQUENCY	3009	6.28581	6.45850	18914	2.00000	73.00000
RECENCY	3009	80.06381	84.82696	240912	0	357.00000

Pearson Correlation Coefficients, N = 3009
Prob > |r| under H0: Rho=0

	MONETARY	FREQUENCY	RECENCY
MONETARY	1.00000	0.88999 <.0001	-0.30714 <.0001
FREQUENCY	0.88999 <.0001	1.00000	-0.33992 <.0001
RECENCY	-0.30714 <.0001	-0.33992 <.0001	1.00000

Since high correlation of 0.89 is observed between Monetary and Frequency metrics, the customers were segmented based on their Monetary Value.

Customers who spent more than 80th percentile of monetary value were filtered and their characteristics were examined. These top 20% customers could be categorized as the most valuable customers, as they contribute to 54% of Kleenex's revenue.

To understand the demographics of these customers, buckets were created as follows.

Cutomer Demographics	Levels	Buckets	Dummy variable
Male Education	1,2,3	School	EDUC_MS
	4,5,6	College	EDUC_MC
	7,8	Graduate	EDUC_MG
Female Education	1,2,3	School	EDUC_FS
	4,5,6	College	EDUC_FC
	7,8	Graduate	EDUC_FG
Family Size	1,2,3	Regular	FAM_SIZE_L
	4,5,6	Large	FAM_SIZE_R
Male Age	1	Young	AGE_MY
	2,3,4	Mid-Age	AGE_MM
	5,6	Elder	AGE_ME
Female Age	1	Young	AGE_FY
	2,3,4	Mid-Age	AGE_FM
	5,6	Elder	AGE_FE
Male Occupation	1,2,3	White Collar - High	OCC_MWH
	4,5	White Collar - Low	OCC_MWL
	6,7,8,9	Blue Collar	OCC_MB
	10,13	Unemployed	OCC_MNO
	0	Other	OCC_MO
Female Occupation	1,2,3	White Collar - High	OCC_FWH
	4,5	White Collar - Low	OCC_FWL
	6,7,8,9	Blue Collar	OCC_FB
	10,13	Unemployed	OCC_FNO
	0	Other	OCC_FO
Income	1,2,3,4	Low	FAM_INCOME_L
	5,6,7,8	Medium	FAM_INCOME_M
	9,10,11	High	FAM_INCOME_H
	12	Very High	FAM_INCOME_VH
Child	1,2,3	One	CHILD_1
	4,5,6	Two	CHILD_2
	7	Three	CHILD_3
	8	Zero	CHILD_0
Pets	0	No Pets	PETS_0
	>0	Pets (Dog/Cat)	PETS_GR_1

Variable	N	Mean	Std Dev	Minimum	Maximum
PANID	448	2028035.8	1039875.7	1100016	3842740
MONETARY	448	37.042701	23.942909	19.41	320.66
FREQUENCY	448	15.828125	9.189743	3	73
DAY	448	27.125	34.55874	0	308
FAM_SIZE_L	448	0.3125	0.4640306	0	1
FAM_SIZE_R	448	0.6875	0.4640306	0	1
FAM_INCOME_L	448	0.046875	0.2116074	0	1
FAM_INCOME_M	448	0.4441964	0.4974317	0	1
FAM_INCOME_H	448	0.3995536	0.4903542	0	1
FAM_INCOME_V	448	0.109375	0.3124581	0	1
AGE_MY	448	0.0044643	0.0667405	0	1
AGE_MM	448	0.3995536	0.4903542	0	1
AGE_ME	448	0.5959821	0.4912495	0	1
AGE_FY	448	0	0	0	0
AGE_FM	448	0.4464286	0.4976776	0	1
AGE_FE	448	0.5535714	0.4976776	0	1
EDUC_MS	448	0.0290179	0.1680442	0	1
EDUC_MC	448	0.59375	0.4916814	0	1
EDUC_MG	448	0.3772321	0.4852356	0	1
EDUC_FS	448	0.0245536	0.1549331	0	1
EDUC_FC	448	0.6808036	0.4666864	0	1
EDUC_FG	448	0.2946429	0.4563917	0	1
OCC_MWH	448	0.3883929	0.4879296	0	1
OCC_MWL	448	0.046875	0.2116074	0	1
OCC_MB	448	0.2165179	0.4123316	0	1
OCC_MNO	448	0.3415179	0.4747489	0	1
OCC_MO	448	0.0066964	0.0816484	0	1
OCC_FWH	448	0.4107143	0.4925135	0	1
OCC_FWL	448	0.1607143	0.3676778	0	1
OCC_FB	448	0.0691964	0.2540716	0	1
OCC_FNO	448	0.3549107	0.4790212	0	1
OCC_FO	448	0.0044643	0.0667405	0	1
CHILD_1	448	0.1852679	0.388949	0	1
CHILD_2	448	0.0803571	0.2721492	0	1
CHILD_3	448	0.0044643	0.0667405	0	1
CHILD_0	448	0.7299107	0.4445021	0	1
PETS_0	448	0.5066964	0.5005141	0	1
PETS_GR_1	448	0.4933036	0.5005141	0	1

Means of the dummy variables created were studied to get insights on customer demographics. The following characteristics were observed in the top 20% customers,

- Family with size ranging from 1 to 3.
- Family with income ranging from $20,000 to $55,000 per year.
- People with age greater than 55 (includes both male and female).
- College Students (includes both male and female).
- Professional/Manager/Administrator/Sales Person (includes both male and female).
- Family with no children and whose size is 2.

We also examined the customers having low recency (less than 20th percentile of recency value) and high monetary (greater than 60th Percentile of monetary value) values. There were 55 customers in this category. These customers may be brought into the high recency fold by targeting them with coupons, offers and special discounts.

PROBLEM STATEMENT – 3

- To study brand preferences of customers and the factors affecting their preferences.

The market for facial tissues is majorly dominated by three brands – Kleenex, Puffs and Private Label. They have a combined market share of around 86%. The multinomial logit model is used to analyse the factors that influence the brand choice of customers across these three brands across drug and grocery stores.

The dependent variable is the "decision of brand" and the independent variables are as follows:

- Weighted Price - Price
- Weighted Display - D
- Weighted Promotion of product - PR
- Weighted Feature - F
- Family Size
- Income of Head of Household
- Age group of male head of household
- Age group of female head of household
- Number of pets owned (cats and dogs)

Weighted Price, Weighted Display, Weighted Feature and Weighted Promotion were calculated for store, week combination across all the three brands.

The following dummy variables were created to study the customer demographics, which will help in getting insights about the factors affecting their preferences. Private Label brand is taken as reference.

Demographics	Dummy Variables	Meaning
Brand Choice of Customer	BC1	Kleenex
	BC2	Puffs
	BC3	Private Label
Family Size	FAM1	Family Size of customers choosing Kleenex
	FAM2	Family Size of customers choosing Puffs
	FAM3	Family Size of customers chosing Private Label
Income of Head of Household	INC1	Income of Head of Household choosing Kleenex
	INC2	Income of Head of Household choosing Puffs
	INC3	Income of Head of Household choosing Private Label
Age group of female head of household	FAGEGP1	Age group of female head of household choosing Kleenex
	FAGEGP2	Age group of female head of household choosing Puffs
	FAGEGP3	Age group of female head of household choosing Private Label
Age group of female head of household	MAGEGP1	Age group of male head of household choosing Kleenex
	MAGEGP2	Age group of male head of household choosing Puffs
	MAGEGP3	Age group of male head of household choosing Private Label
Number of pets owned (cats and dogs)	NPETS1	Number of pets owned by customers choosing Kleenex
	NPETS2	Number of pets owned by customers choosing Puffs
	NPETS3	Number of pets owned by customers choosing Private label

The SAS System

The MDC Procedure

Conditional Logit Estimates

Algorithm converged.

Model Fit Summary	
Dependent Variable	DECISION
Number of Observations	26620
Number of Cases	79860
Log Likelihood	-23641
Log Likelihood Null (LogL(0))	-29245
Maximum Absolute Gradient	6.27499E-6
Number of Iterations	4
Optimization Method	Newton-Raphson
AIC	47315
Schwarz Criterion	47446

Discrete Response Profile			
Index	CHOICE	Frequency	Percent
0	1	15899	59.73
1	2	5582	20.97
2	3	5139	19.31

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	11207	2 * (LogL - LogL0)
Upper Bound of R (U)	58490	- 2 * LogL0
Aldrich-Nelson	0.2963	R / (R+N)
Cragg-Uhler 1	0.3436	1 - exp(-R/N)
Cragg-Uhler 2	0.3866	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.3733	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.3724	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.1916	R / U
Veall-Zimmermann	0.4311	(R * (U+N)) / (U * (R+N))
N = # of observations, K = # of regressors		

According to McFadden's LRI, 19% of variation in the brand choice is explained by the independent variables. Discrete Response Profile indicates that, Kleenex is bought by the highest number of panelists followed by Puffs and Private Label.

The SAS System

The MDC Procedure

Conditional Logit Estimates

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
PRICE	1	-75.3021	4.7581	-15.83	<.0001
BC1	1	1.4867	0.1594	9.33	<.0001
BC2	1	0.9334	0.1888	4.94	<.0001
PR	1	-0.0439	0.0322	-1.36	0.1726
D	1	0.5890	0.0357	16.50	<.0001
F	1	0.6249	0.0321	19.48	<.0001
FAM1	1	-0.1699	0.0182	-9.35	<.0001
FAM2	1	-0.1560	0.0216	-7.21	<.0001
INC1	1	0.1153	0.006658	17.31	<.0001
INC2	1	0.0703	0.007865	8.94	<.0001
FAGEGP1	1	-0.0280	0.0188	-1.49	0.1355
FAGEGP2	1	-0.0604	0.0220	-2.75	0.0060
MAGEGP1	1	-0.0619	0.0182	-3.40	0.0007
MAGEGP2	1	-0.0211	0.0215	-0.98	0.3257
NPETS1	1	-0.0303	0.0158	-1.92	0.0544
NPETS2	1	-0.0310	0.0188	-1.65	0.0985

Almost all variables are statistically significant at the 95% confidence level except – PR, NPETS2, MAGEGP2, FAGEGP1.

Keeping other independent variables constant, the following insights were obtained from MDC model.

1. As the family size increases, the probability of buying private label increases; smaller families choose Kleenex while larger families choose Private Labels
2. Households in the high-income category tend to choose Kleenex over the other brands.
3. Households having older female as their head prefer Private Labels over Puffs.
4. Households having older male as their head prefer Private Label over Kleenex brand
5. Households with less number or no pets prefer Kleenex facial tissues over other brands.
6. Promotion through price reduction does not seem to have any effect on brand choice.

DEMOGRAPHICS	KLEENEX	PUFFS	PRIVATE LABEL
Small Families			
Large Families			
High Income Households			
Low Income Households			
Elder female head of households			
Younger male head of households			
Elder male head of households			
Households with less number or no pets			
Households with more pets			

Preference	
	High
	Medium
	Low
	NA

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of PREDICT by DECISION		
	DECISION		
PREDICT	0	1	Total
0	42955 53.79 80.68 80.68	10285 12.88 19.32 38.64	53240 66.67
1	10285 12.88 38.64 19.32	16335 20.45 61.36 61.36	26620 33.33
Total	53240 66.67	26620 33.33	79860 100.00

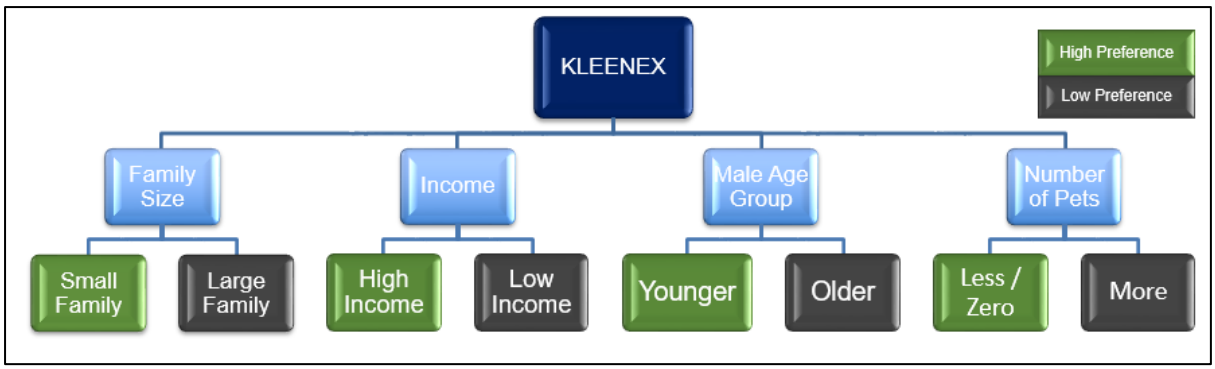
The accuracy of prediction / Prediction power of the model
= ((42955 + 16335) / 79860) * 100
= **74.24%**

RECCOMENDATIONS

- From RFM analysis, we observed that 55 panelists have high monetary and low recency values.
    - Coupons and discounts targeting these households may bring them back into Kleenex fold.

  An insight gained from RFM analysis is that the top 20% of customers contribute to 54% of Kleenex's revenue.
    - A loyalty program for these customers can be devised by Kleenex to retain its prevailing market share.

- From the Multinomial Logit model, we found out that following are the characteristics of customers preferring Kleenex.



  It can be observed that small family, high income, younger people and households with less/zero pets prefer Kleenex more than any other brand.

    - Periodic feedbacks can be obtained from these customers to understand their level of satisfaction and act appropriately in case of quality or pricing issues. This will aid Kleenex in retaining the market share.

- Performing panel regression gave us the following insights

  Sales of Kleenex tissues is sensitive to changes in its price. Decreasing price by 1% may increase sales by 1.22%.

  Sales of Kleenex tissues is insensitive to changes in display or feature percentages of Puffs - brand which is trailing Kleenex in the market. This confirms that Kleenex is the undisputed market leader.

  Sales of Kleenex tissues is insensitive to changes in its display or feature.

    - Though the sales of Kleenex tissues insensitive to changes in its display or feature percentages, Kleenex still needs do a certain level of display or feature promotions. This might help them retain their existing market share.