Predictive Analytics Project

Group 11

Shilpa Parameshwara Bhat
Meghana Koteshwara Prahlad
Trang Pham
Neeraj Agarwala
Ankita Anilkumar Lodhavia

Objective

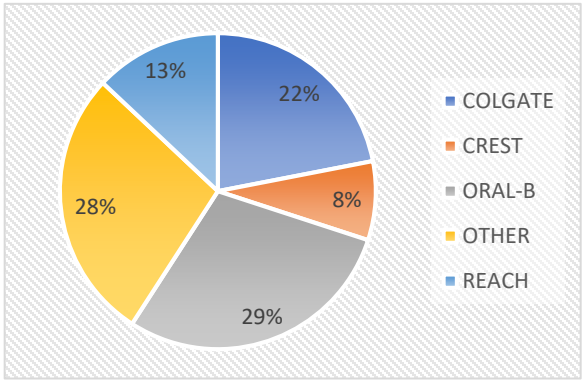
Colgate is an umbrella brand principally used for oral hygiene products such as toothpastes, toothbrushes, mouthwashes and dental floss.

In this project the focus is o find insights from data that would assist the brand manager of Colgate in taking actions to increase their market share.

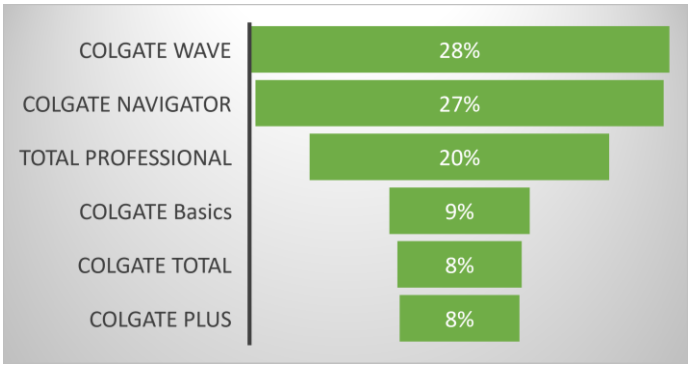
The following analysis were made to come out with the recommendation for improvement in the market share

1. Segmentation through Recency Frequency and Monetory
2. Price Elasticity
3. Preference of customer of brand based on price reduction flag, display etc.

Exploratory Data Analysis

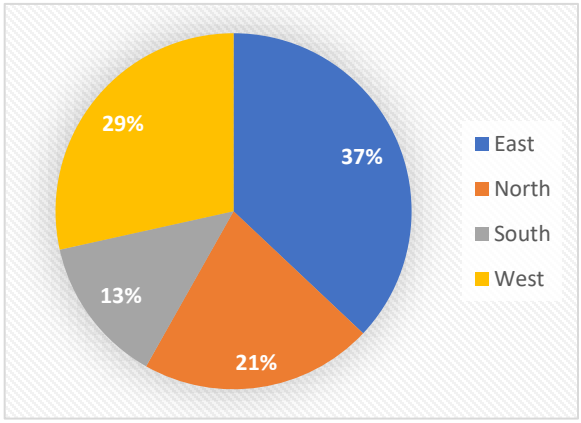


Oral B is the market leader, followed by Colgate, Reach and crest. Colgate should adopt to certain pricing strategy, advertisement and promotions to compete with Oral-B

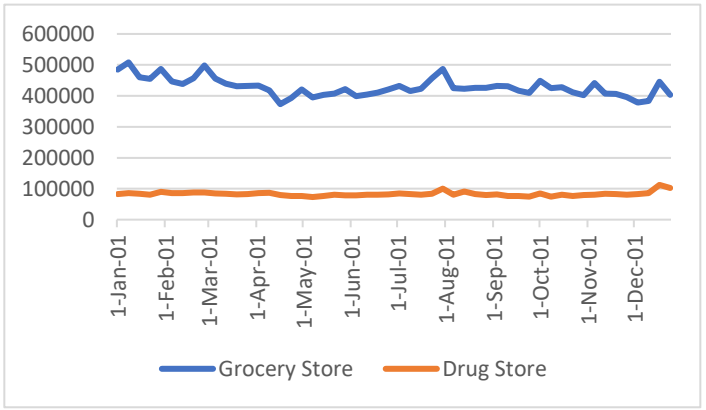


Colgate wave is the most preferred followed by Colgate Navigator and then Colgate total professional, these three contribute to 75% of the total Colgate variant sales

Toothbrush sales is highest 37% in the East followed by 29% in the West



Weekly trend of Colgate sales in Grocery Store and Drug Store



Section 1: Segmentation using RFM

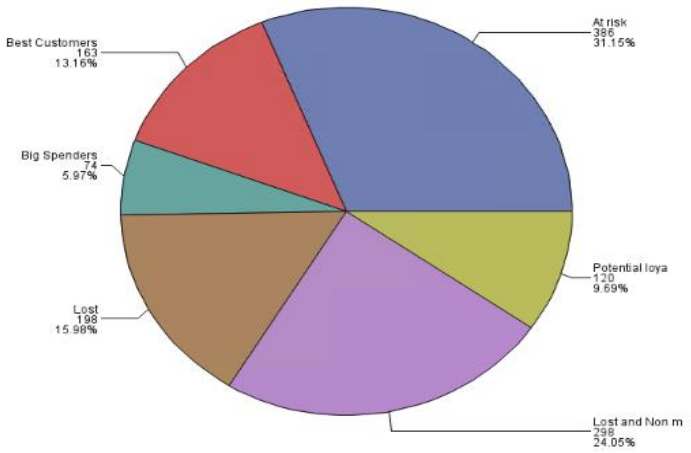
Data Preparation and cut-off selection

On the merged panel data with product details, we query and select only our brand which is 'Colgate' to run our RFM model and segment the customers who we need to concentrate on for marketing.

After we find our RFM ranking, we find that the correlation between F and M is high (0.82). Hence, we remove F and keep only R and M in the model.

We rank the customers based on **Recency and Monetary** status and group them into **4 groups** where **4** is the most recent/frequent/highest spending customer and **1** being the least.

Obs	PANID	Rec	Freq	Mon	R	F	M	RFMScore	groups
1	1102012	1165	1	1.79	4	0	1	41	Potential loya
2	1151456	1165	1	2.99	4	0	2	42	Potential loya
3	1201103	1165	4	17.94	4	3	4	44	Best Customers
4	1306522	1165	1	6.58	4	0	4	44	Best Customers
5	3104976	1165	3	10.76	4	3	4	44	Best Customers
6	3356030	1165	2	3.84	4	2	3	43	Potential loya
7	3808394	1165	1	2.89	4	0	2	42	Potential loya
8	3817387	1165	1	1.99	4	0	1	41	Potential loya
9	3836320	1165	1	2.99	4	0	2	42	Potential loya
10	1113696	1164	2	5.98	4	2	3	43	Potential loya



Insights:

GROUPS	Description	Percentage of Customers	Marketing Strategy
Best Customers	These are customers who have been frequently buying and have had recent purchases along with higher monetary buyings	7.4	No additional strategy required as these are customers who are already loyal and they will stick to the brand without any additional offer.
Big Spenders	These are customers who have been frequently buying along with higher monetary buyings but have had fewer recent purchases	5.99	Market the more expensive products or the new products that come in the market to them and they are highly likely to buy them
Potential Loyals	These are customers who have had recent purchases but not that frequent or with lower monetary buyings	31.42	These are customers who are mostly still undecided about the brand. Hence giving them offers excessively at this point would turn them into a loyal customer.
Customers at Risk	These are customers who have had fewer recent purchases but not that frequent or with lower monetary buyings comparitively	28.89	These are customers who are likely to churn faster and with lower monetary values, hence advertize excessively with price incentives or buy one get one offers.
Lost and Non Monetary	These are customers who have had very less recent purchases but not that frequent or with very low monetary buyings comparitively	23.83	These are customers that are mostly lost or have churned to other brand and they have a higher percentage of 23.83%, so do not waste money for advertising on them. Instead concentrate on the customers who are likely to churn

Section 2: Price Elasticity

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	902
Time Series Length	52

Fit Statistics			
SSE	24296941.80	DFE	42130
MSE	576.7135	Root MSE	24.0149
R-Square	0.6255		

F Test for No Fixed Effects			
Num DF	Den DF	F Value	Pr > F
952	42130	52.01	<.0001

Studying Price Elasticity is important, as it plays an important role in creating an effective price structure.
It is important to understand own price elasticity to know how the sales varies with change in price, cross price elasticity is important, to understand the price effects on sales in relation to competitors.

Since the p-value is less than 0.05, we reject null hypothesis and use Fixed effect model

Colgate						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Mean
Intercept	1	-13.5915	4.5619	-2.98	0.0029	
wt_price_brand1	1	7.966792	0.5783	13.78	<.0001	2.737
wt_price_brand2	1	0.77297	0.5034	1.54	0.1247	3.047
wt_price_brand3	1	-0.03312	0.6253	-0.05	0.9578	2.977
wt_price_brand4	1	0.718982	0.5375	1.34	0.1811	2.360
disp_wt_brand1	1	41.26783	1.2082	34.16	<.0001	0.037
disp_wt_brand2	1	-4.40695	1.076	-4.1	<.0001	0.048
disp_wt_brand3	1	1.709446	1.2831	1.33	0.1828	0.028
disp_wt_brand4	1	-9.49044	1.5838	-5.99	<.0001	0.032
Feature_wt_brand1	1	162.2567	4.9352	32.88	<.0001	0.114
Feature_wt_brand2	1	-27.1861	4.0606	-6.7	<.0001	0.110
Feature_wt_brand3	1	15.77671	6.1314	2.57	0.0101	0.048
Feature_wt_brand4	1	0.107426	0.9581	0.11	0.9107	0.083
PR_wt_brand1	1	40.49938	4.9798	8.13	<.0001	0.232
PR_wt_brand2	1	-2.76138	0.7595	-3.64	0.0003	0.207
PR_wt_brand3	1	0.543821	0.6976	0.78	0.4357	0.161
PR_wt_brand4	1	-6.36806	0.8691	-7.33	<.0001	0.211
price_PR1	1	-16.6624	1.4796	-11.26	<.0001	0.589
price_PR2	1	-3.42726	0.9251	-3.7	0.0002	0.709
price_PR3	1	4.581394	1.0856	4.22	<.0001	0.688
price_PR4	1	6.902912	1.0195	6.77	<.0001	0.547
price_F1	1	-47.591	1.71	-27.83	<.0001	0.283
price_F2	1	6.792355	1.1853	5.73	<.0001	0.289
price_F3	1	0.185458	1.8084	0.1	0.9183	0.124
price_F4	1	-6.05318	1.6823	-3.6	0.0003	0.117
PR_F1	1	-22.3944	2.296	-9.75	<.0001	0.067
PR_F2	1	8.918627	2.3986	3.72	0.0002	0.062
PR_F3	1	-5.07548	2.8359	-1.79	0.0735	0.030
PR_F4	1	3.940534	3.3817	1.17	0.2439	0.011

Self-Price Elasticity = %Change in $Sales_{Colgate}$/%Change in $Price_{Colgate}$

Cross Price Elasticity = %Change in $Sales_{Colgate}$/%Change in $Price_{Oral\text{-}B}$

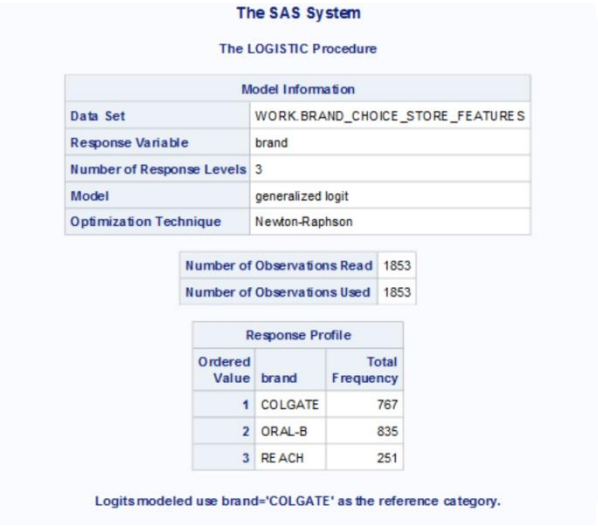
From the result we see that,

- If Colgate reduces price by 1% and there is a discount sale, then sales increases by 0.66%
- If Colgate reduce price by 1% and have a feature display, then sales increases by 3.03%
- If Colgate reduce price by 1% and have a discount sale along with feature display, then sales increases by 4.3%

- If Oral B reduces price by 1% along with a feature display, then sales of Colgate will decrease by 0.64%
- If Reach reduces price by 1% along with a discount sale, then sales of Colgate will decrease by 0.37%
- If Reach reduces price by 1% along with a feature display, then sales of Colgate will decrease by 0.013%

Section 3: Study of preference of customer based on price reduction flag, display etc.

Proc Logistic is used with brand as the dependent variable and Colgate being the reference group and regressed on weighted price, weighted display etc. for week for brand



The SAS System

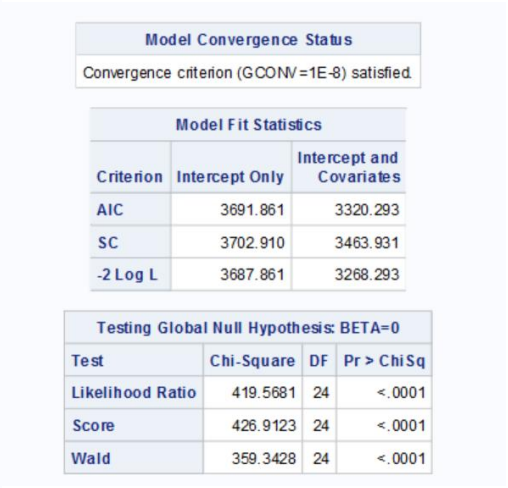
The LOGISTIC Procedure

Model Information	
Data Set	WORK.BRAND_CHOICE_STORE_FEATURES
Response Variable	brand
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	1853
Number of Observations Used	1853

Response Profile		
Ordered Value	brand	Total Frequency
1	COLGATE	767
2	ORAL-B	835
3	REACH	251

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



Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	3691.861	3320.293
SC	3702.910	3463.931
-2 Log L	3687.861	3268.293

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.5681	24	<.0001
Score	426.9123	24	<.0001
Wald	359.3428	24	<.0001

Using -2Log L we can calculate the R-squared value i.e. ~ 11.38%

Analysis of Maximum Likelihood Estimates							
Parameter	brand	DF	Estimate	StandardError	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	ORAL-B	1	-0.1083	0.0877	1.5245	0.2169	0.897
Intercept	REACH	1	-1.1171	0.1265	77.9679	<.0001	0.327
disp_wt_brand1	ORAL-B	1	-1.2335	0.492	6.2856	0.0122	0.291
disp_wt_brand1	REACH	1	-2.0073	0.7836	6.5618	0.0104	0.134
disp_wt_brand2	ORAL-B	1	1.6851	0.4134	16.6153	<.0001	5.393
disp_wt_brand2	REACH	1	-0.6159	0.6994	0.7755	0.3785	0.54
disp_wt_brand3	ORAL-B	1	1.2529	0.5489	5.2104	0.0225	3.5
disp_wt_brand3	REACH	1	0.5192	0.7223	0.5168	0.4722	1.681
disp_wt_brand4	ORAL-B	1	-0.3775	0.5628	0.45	0.5024	0.686
disp_wt_brand4	REACH	1	-1.553	0.936	2.753	0.0971	0.212
Feature_wt_brand1	ORAL-B	1	0.3805	0.3272	1.3525	0.2448	1.463
Feature_wt_brand1	REACH	1	0.0793	0.4743	0.028	0.8671	1.083
Feature_wt_brand2	ORAL-B	1	1.2993	0.3935	10.9032	0.001	3.667
Feature_wt_brand2	REACH	1	-1.3735	0.5605	6.0043	0.0143	0.253
Feature_wt_brand3	ORAL-B	1	-0.1252	0.413	0.0919	0.7617	0.882
Feature_wt_brand3	REACH	1	0.9981	0.4415	5.1102	0.0238	2.713
Feature_wt_brand4	ORAL-B	1	-0.7498	0.5293	2.0068	0.1566	0.472
Feature_wt_brand4	REACH	1	-0.3762	0.739	0.2592	0.6107	0.686
PR_wt_brand1	ORAL-B	1	-2.0602	0.3441	35.8555	<.0001	0.127
PR_wt_brand1	REACH	1	-1.572	0.4811	10.6781	0.0011	0.208
PR_wt_brand2	ORAL-B	1	0.3381	0.4006	0.7122	0.3987	1.402
PR_wt_brand2	REACH	1	1.3472	0.5373	6.2872	0.0122	3.847
PR_wt_brand3	ORAL-B	1	-0.6292	0.3828	2.7021	0.1002	0.533
PR_wt_brand3	REACH	1	1.0714	0.4196	6.5199	0.0107	2.919
PR_wt_brand4	ORAL-B	1	2.2586	0.4299	27.6041	<.0001	9.569
PR_wt_brand4	REACH	1	1.8176	0.5868	9.5931	0.002	6.157

Interpreting the significant Co-efficient:

- Customers are less likely to choose Oral-B compared to Colgate if there is high weightage on display for Colgate
- Customers are less likely to choose Reach compared to Colgate if there is high weightage on display for Colgate
- Customers are more likely to choose Oral-B compared to Colgate if there is high weightage on display for Oral-B
- Customers are more likely to choose Oral-B compared to Colgate if there is high weightage on display for Reach
- Customers are less likely to choose Reach compared to Colgate if there is high weightage on display for Reach
- Customers are more likely to choose Oral-B compared to Colgate if there is high weightage on feature for Oral-B
- Customers are less likely to choose Reach compared to Colgate if there is high weightage on feature for Oral-B
- Customers are less likely to choose Reach compared to Colgate if there is high weightage on feature for Reach
- Customers are less likely to choose Oral-B compared to Colgate if there is high weightage on price reduction for Colgate
- Customers are less likely to choose Reach compared to Colgate if there is high weightage on price reduction for Colgate
- Customers are more likely to choose Reach compared to Colgate if there is high weightage on price reduction for Oral-B
- Customers are more likely to choose Reach compared to Colgate if there is high weightage on price reduction for Reach

Recommendations

Oral-B is the market leader and the tough competitor for Colgate compared to any other brand. To be the market leader Colgate must battle with Oral-B for new market shares.

- Reducing price would not improve the sales to a greater extent but reducing the price and having right advertisement like Price Reduction tag and feature display would improve the sales.
- If Oral-B along with price reduction also has a display, this could affect the sale of Colgate, this is when manager must place the right advertisement for Colgate in line with competitor Oral-B
- From the brand choice analysis, we see that customers are less likely to choose Oral-B if there is price reduction along with display advisement by Colgate. When there is price reduction sale it's important to have display advisement, else the price reduction is not that effective
- Colgate wave, Colgate Navigator and Colgate total professional contribute to 75% of Colgate Sales, so it's important to have right advertising and Price reduction strategy for these variants particularly. This would help in the battle with Oral-B to gain new market share
- From the RFM analysis we can conclude that we need to concentrate the marketing campaign on Big Spenders, Potential Loyal and Customer at Risk. ~23% are lost and non- monetary, these people should be removed from the marketing campaign and focus must be on who are likely to churn, which is ~28% customers.

Code.txt