BUAN 6337.005

Predictive Analytics using SAS

Project:
Make Blades Sharp Again

Group 3

Group Members:

1. Weiyang Sun
2. Harsh Gupta
3. Gongzhan Hua
4. Mengyun Xia

Summary

X) Business Recommendation

Effects of Advertising on Sales Price

- Minor Displays are associated with a **310% increase** in unit sales as compared to no displays, controlling for other variables.
- Major Displays are associated with a **412% increase** in unit sales as compared to no displays, controlling for other variables.
- Large features are associated with a **242% increase** in unit sales as compared to no features, controlling for other variables.
- Small features are associated with a **283% increase** in unit sales as compared to no features, controlling for other variables.

Effects of Advertising on Switching

- For every unit dollar increase in unit selling price, there is an associated **1% increase** in switching occurring, controlling for other variables.
- Major displays are associated with a **54% decrease** in switching occurring as compared to no displays, controlling for all other variables.
- Features with coupons are associated with a **98% increase** in switching occurring as compared to no features, controlling for all other variables.

Overall Business Recommendation:

Dear BIC,
Your battle for new market shares would be with Schick.

Reducing price would not entice an individual to switch. The usage of features with coupons and minor displays would however retain and attract customers. Advertisements should be focused on multi-blade capable of cartridge replacement for unisex.

Your target audience will be on households with income below $12000 per year.

Using RFM, the criteria enforced was:
1) RFM Score >= 12
2) Each component of the score must be >= 4
3) Both condition 1 & 2 must be satisfied

Demographic Traits of a Loyal Customer

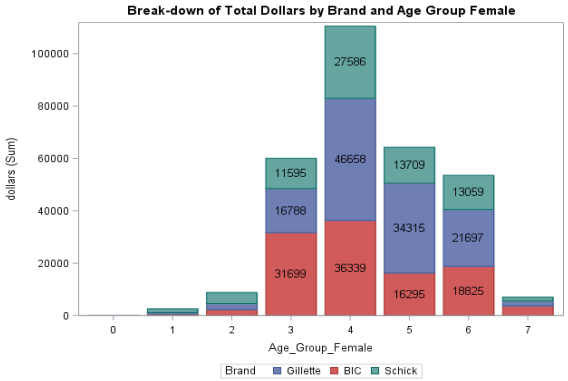
- Households with female heads between the age of 35-44 years are **2.72 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with female heads between the age of 45-54 years are **3.61 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with female heads between the age of 55-64 years are **4.32 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with combined income between $0 - 9.9k per year are **1/0.436 = 2.29 times more likely** to be loyal customers as compared to households with combined income between $10 - 11.9k per year.

Effects of Product on Brand selection for Loyal Customers

- Individuals are (1/0.067) = **14.9 times more likely** to choose a cartridge over disposables for Schick relative to BIC, controlling for all other variables.
- Individuals are **9.56 times more likely** to choose triple blades over singles for Schick relative to BIC, controlling for all other variables.
- Individuals are **13.66 times more likely** to choose twin blades over singles for Schick relative to BIC, controlling for all other variables.
- Individuals are (1/0.099) = **10.1 times more likely** to choose unisex shavers over female for Schick relative to BIC, controlling for all other variables.

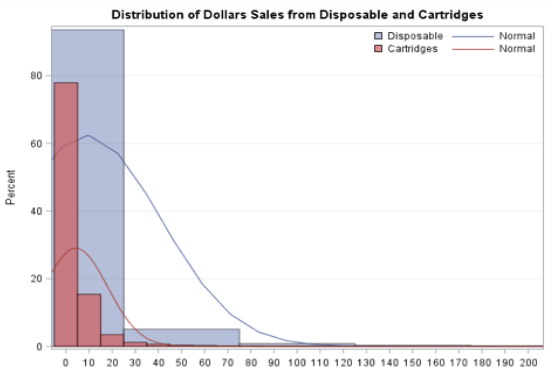
Part 1 - Descriptive Analytics

A) Breaking down dollars sales by various factors



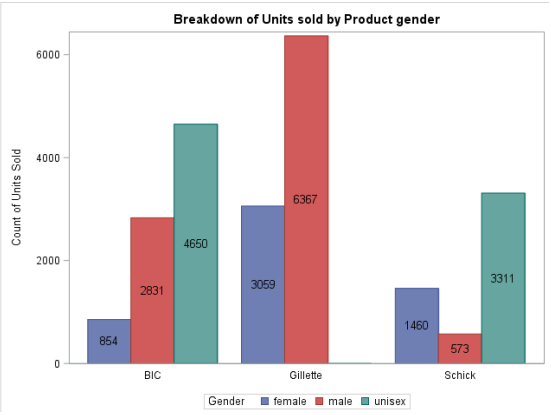
In general, households with females in the age group of 45-54 (group 4) spent the most on blades as compared to other groups.

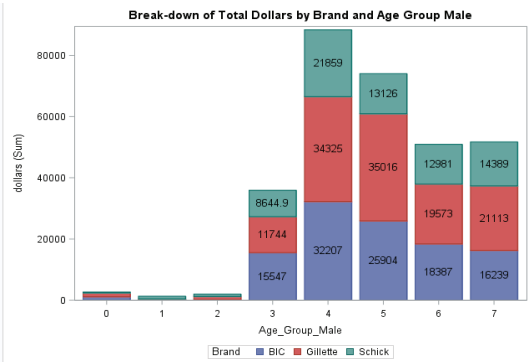
The brand with the highest dollars sales in this group is Gillette. This pattern remains true for households with females with age of 55 and above. This trend does not hold for the younger groups.



Regardless of whether a shaver is disposable or cartridge form, the majority of sales (~90%) is made within the range of $1 to $20. Based on the normal plot, we conclude that the mean of the disposable and cartridge dollars sales are similar. No t-test is further required.

B) Breaking down dollars sales by various factors





For households with males, Gillette is the top sales generator from the age group of 25 onwards. This trend holds steady throughout.

Similar to females, group 4 exhibits the highest sales.

Business Insight: Gillette is a best seller for groups over 45 years of age, while BIC seems to dominate the younger markets.

Since 90% of sales occurred within $1- $20. Therefore, by performing a boxplot of the distribution of dollars sales generated by brand and whether price reduction event occurred within the range of $1-20, we can conclude that price reduction does not seem to have an effect on sales distribution.

Business Insight: Regardless of brand or form (i.e disposable), price reduction does not change sales distribution and most sales occurs within the range of $1-$20.

By searching google based on the different product names per brand, we were able to derive additional gender information. Count of units refers to the total units sold.

Based on this, Gillette is the best seller for male and female shavers as compared to the other brands. However, if an individual wishes to purchase a generic unisex shaver, BIC is the choice.

It is interesting to note also that Schick female shavers are sell approximately twice more than Schick male shavers.

Business Insight: While Gillette covers gender specific products, BIC and Schick covers primarily unisex products.

Part 2 - Effects of Advertising on Unit Sales

C) Calculating Weighted Components

In order to aggregate the various variables down to a brand level, we had to apply this weighted formula across features, brands, displays, forms and packaging.

$$Total\ Weighted\ Price = \frac{Total\ Dollars\ Sales * \frac{Units\ of\ selected\ Brand}{Total\ Units\ Bought}}{Volume\ Equivalent}$$

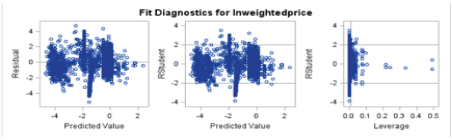
Note: This above formula is only for unit sales. The same concept applies for the other variables.

D) Regressing Log(Unit Sales) on Advertising

Base-levels are: Cartridges, BIC, No price reduction, No display, No feature, Single blade

Some product characteristics were added in to control for the effects of brand, packaging and forms.

- The adjusted R Square of the model is 0.4732.
- There are no variables with VIF > 10 or condition index > 100 thus indicating no multicollinearity.



As the residual plots from the fit diagnostic did not indicate a constant error variance, white-standard correction was applied.

F) Performing Panel Data Regressions

	Parameter Estimates				
Variables	OLS(1)	FixOne	FixTwo	RanOne	RanTwo
Intercept	-1.58	-4.51	-4.54	-1.59	-1.61
total_weightedtype	-0.73	-0.74	-0.73	-0.74	-0.73
total_weightedbrand2	-0.67	0.37	0.37	-0.62	-0.62
total_weightedbrand3	-0.16	0.68	0.68	-0.06	-0.06
total_weightedpr	0.02	-0.47	-0.40	-0.15	-0.12
total_weighteddmin	3.45	2.71	2.85	3.12	3.10
total_weighteddmaj	4.47	4.27	4.32	4.01	4.12
total_weightedfa	2.55	3.35	3.57	2.42	2.42
total_weightedfa1	-1.27	-1.37	-1.78	-1.71	-1.72
total_weightedfb	-0.01	-0.08	-0.21	0.25	0.18
total_weightedfc	2.87	1.61	1.57	2.80	2.83
total_weightedform2	3.56	3.18	3.20	3.55	3.58
total_weightedform3	4.06	3.90	3.94	4.03	4.05
brand3*fc	-6.84	-3.04	-2.77	-6.53	-6.56
brand2*fc	-5.24	-2.89	-1.95	-4.84	-4.85
brand2*fb	0.49	1.12	1.87	0.70	0.70
brand2*dmin	-4.38	-5.96	-6.00	-4.01	-3.88
brand3*dmin	-3.56	-2.90	-3.01	-3.45	-3.34
R-Squared	0.476	0.837	0.845	0.456	0.457

Note: Since heteroscedasticity was found in OLS, we have perform Newey-West94 correction on the panel data as well.

Hausman Test for Random Effects			
Coefficients	DF	m Value	Pr > m
17	17	25.77	0.0787

Since p-value > 0.05 for Hausman Test only for RanTwo model. Therefore, we will adopt that estimator.

E) Regressing Log(Unit Sales) on Advertising with Interactive Effects

lnweightedprice = total_weightedtype + total_weightedbrand2 + total_weightedbrand3 + total_weightedpr + total_weighteddmin + total_weighteddmaj + total_weightedfa + total_weightedfa1 + total_weightedfb + total_weightedfc + total_weightedform2 + total_weightedform3 + brand3*fc + brand2*fc + brand3*dmin + brand2*dmin

Parameter Estimates									
						Heteroscedasticity Consistent			
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	-1.58370	0.05713	-27.72	<.0001	0.08792	-18.01	<.0001	0
total_weightedtype	1	-0.73140	0.02912	-25.12	<.0001	0.03623	-20.19	<.0001	-0.38550
total_weightedbrand2	1	-0.66937	0.20767	-3.22	0.0013	0.16304	-4.11	<.0001	-0.08450
total_weightedbrand3	1	-0.15639	0.21685	-0.72	0.4708	0.17473	-0.90	0.3708	-0.01691
total_weightedpr	1	0.01673	0.28801	0.06	0.9537	0.19634	0.09	0.9321	0.00090465
total_weighteddmin	1	3.44689	0.81109	4.25	<.0001	0.71233	4.84	<.0001	0.09238
total_weighteddmaj	1	4.47291	1.26988	3.52	0.0004	0.63779	7.01	<.0001	0.04790
total_weightedfa	1	2.54579	0.77052	3.30	0.0010	0.82663	3.08	0.0021	0.04642
total_weightedfa1	1	-1.26680	1.88021	-0.67	0.5005	0.76418	-1.66	0.0975	-0.00890
total_weightedfb	1	-0.00584	0.65767	-0.01	0.9929	0.70137	-0.01	0.9934	-0.00014647
total_weightedfc	1	2.87125	0.60171	4.77	<.0001	0.42088	6.82	<.0001	0.10525
total_weightedform2	1	3.56241	0.20579	17.31	<.0001	0.20333	17.52	<.0001	0.46726
total_weightedform3	1	4.05711	0.27130	14.95	<.0001	0.24377	16.64	<.0001	0.31571
inte21	1	-6.83772	1.81417	-3.77	0.0002	1.02404	-6.68	<.0001	-0.06693
inte20	1	-5.23589	1.70684	-3.07	0.0022	1.21398	-4.31	<.0001	-0.05754
inte17	1	0.48787	2.54253	0.19	0.8478	2.43010	0.20	0.8409	0.00293
inte5	1	-4.38297	2.25268	-1.95	0.0518	1.70120	-2.58	0.0100	-0.03718
inte6	1	-3.56072	2.61327	-1.36	0.1731	2.13542	-1.67	0.0955	-0.02309

After trying out 21 different interactive terms which aims to capture the relationship of brand reputation on advertising, only 4 terms were signficiant at a ~95% confidence level.

- Gillette - Minor Display (inte5)
- Schick - Minor Display (inte6)
- Gillette - Small Feature (inte20)
- Schick - Small Feature (inte21)

G) Interpreting Significant Coefficients

- Disposable razors are associated with a **73% decrease** in unit sales as compared to cartridges, controlling for other variables.
- Gillette products are associated with a **62% decrease** in unit sales as compared to BIC, controlling for other variables.
- Minor displays are associated with a **310% increase** in unit sales as compared to no displays, controlling for other variables.
- Major displays are associated with a **412% increase** in unit sales as compared to no displays, controlling for other variables.
- Large features are associated with a **242% increase** in unit sales as compared to no features, controlling for other variables.
- Small features are associated with a **283% increase** in unit sales as compared to no features, controlling for other variables.

Part 2 - Effects of Advertising on Unit Sales

G) Interpreting Significant Coefficients

- Twin blades are associated with a **358% increase** in unit sales as compared to single blade, controlling for other variables.
- Triple blades are associated with a **405% increase** in unit sales as compared to single blade, controlling for other variables.

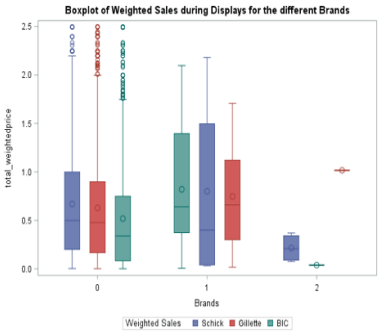
Business Insights: On a whole, displays regardless of size tend to be associated with a higher increase of sales as compared to no displays. This relationship does not hold true for other modes of advertisement. Therefore, displays is a safer advertising option.

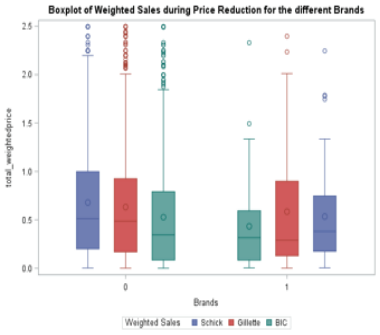
H) Further Exploration

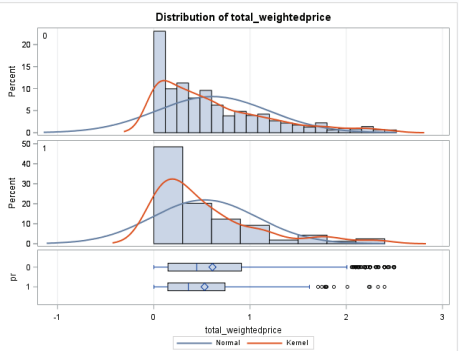
Contrary to intuitive reasoning, price reduction seem to have turned up insignificant in the random effect model. Therefore, further investigation was done, specifically on price reduction.

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2724	1.74	0.0816
Satterthwaite	Unequal	186.33	1.85	0.0663

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2562	162	1.14	0.2690







The mean unit sales between brands are significantly different between having no and some displays, after aggregating sales into unit sales.

A t-test was performed to see if there is significant difference on the mean unit sales before and after price reduction. Since p-value > 0.05 for test of equality of variances, the pooled method for equal variance is adopted. The associated p-value = 0.08, therefore, the means are signficiantly different at a ~92% confidence level.

Business Insight: The effects of price reduction and displays were hidden because the dollars and units recorded were representative of the whole purchase trip and not specific to the brand or product bought. After aggregating it down to unit sales, the difference in sales generated becomes more readily apparent.

I) Running ANOVA test with Tukey's Pairwise Comparison

ANOVA: Form vs Total Weighted Price

The GLM Procedure

Tukey's Studentized Range (HSD) Test for total_weightedprice

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	3018
Error Mean Square	3.220563
Critical Value of Studentized Range	3.31613

Comparisons significant at the 0.05 level are indicated by ***.

FORM Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
TRIPLE BLADE - TWIN BLADE	0.22921	-0.00675	0.46518	
TRIPLE BLADE - SINGLE BLADE	0.59252	0.32169	0.86335	***
TWIN BLADE - TRIPLE BLADE	-0.22921	-0.46518	0.00675	
TWIN BLADE - SINGLE BLADE	0.36331	0.17467	0.55195	***
SINGLE BLADE - TRIPLE BLADE	-0.59252	-0.86335	-0.32169	***
SINGLE BLADE - TWIN BLADE	-0.36331	-0.55195	-0.17467	***

A price parity exist in between single blades and all other kind of blades. The price difference can be notably attributed to extra cost of metal and difficulty in manufacturing, as the number of blades increases.

Business Insight: To fully understand the effects of advertisement, multi-variate regression should be used to control for other significant effects.

Comparisons significant at the 0.05 level are indicated by ***.

F Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
B - A	0.54839	-0.81330	1.91009	
B - C	0.82195	-0.26073	1.90464	
B - NONE	0.91082	0.03689	1.78475	***
B - A+	1.74666	-0.61769	4.11100	
A - B	-0.54839	-1.91009	0.81330	
A - C	0.27356	-0.95753	1.50465	
A - NONE	0.36243	-0.68977	1.41462	
A - A+	1.19826	-1.23761	3.63414	
C - B	-0.82195	-1.90464	0.26073	
C - A	-0.27356	-1.50465	0.95753	
C - NONE	0.08887	-0.56314	0.74087	
C - A+	0.92471	-1.36691	3.21632	
NONE - B	-0.91082	-1.78475	-0.03689	***
NONE - A	-0.36243	-1.41462	0.68977	
NONE - C	-0.08887	-0.74087	0.56314	
NONE - A+	0.83584	-1.36485	3.03653	
A+ - B	-1.74666	-4.11100	0.61769	
A+ - A	-1.19826	-3.63414	1.23761	
A+ - C	-0.92471	-3.21632	1.36691	
A+ - NONE	-0.83584	-3.03653	1.36485	

Only considering the effects of feature on unit sales, it seems that there is a difference in average unit sales between having a medium size feature and all other kinds of features inclusive of not even having any features.

Part 3 - Effects of Advertising on Switching

J) Regressing Weighted Switches on Advertising with non-linear price term

The effect of buyer's switching behaviour can only be accurately observed when the buyer has shopped for at least a few times. Therefore, the focus here will be on individuals who have shopped 3 or more times over the last year. There are only 279 individuals who have satisfied this criteria with 1314 shopping instances.

weighted_switches = total_weightedprice + total_weightedtype + total_weightedbrand2 + total_weightedbrand3 + total_weightedpr + total_weighteddmin + total_weighteddmaj + total_weightedfa + total_weightedfa1 + total_weightedfb + total_weightedfc + total_weightedform2 + total_weightedform3 + brand3*fc + brand2*fc + brand3*dmin + brand2*dmin

Furthermore, we hypothesized that the unit selling price would have a non-linear effect on switching. Once the selling price goes beyond a person's perceived value, switching occurs.

Note: inte5, inte6 and inte20 are the same interactive effects previously considered.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	7.16568	0.42151	6.00	<.0001
Error	1296	91.03542	0.07024		
Corrected Total	1313	98.20110			

Root MSE	0.26503	R-Square	0.0730
Dependent Mean	0.22868	Adj R-Sq	0.0608
Coeff Var	115.89975		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Standardized Estimate
Intercept	1	0.23581	0.01878	12.55	<.0001	0.01912	12.33	<.0001	0
total_weightedprice	1	0.00970	0.00612	1.59	0.1131	0.00559	1.74	0.0827	0.07159
total_weightedprice2	1	0.00004652	0.00019534	0.24	0.8118	0.00012324	0.38	0.7059	0.01028
total_weightedtype	1	-0.01607	0.00868	-1.85	0.0645	0.00847	-1.90	0.0580	-0.05984
total_weightedbrand2	1	-0.46959	0.06897	-6.81	<.0001	0.07328	-6.41	<.0001	-0.40502
total_weightedbrand3	1	-0.20105	0.06991	-2.88	0.0041	0.07407	-2.71	0.0067	-0.15618
total_weightedpr	1	0.03779	0.08066	0.47	0.6395	0.07106	0.53	0.5949	0.01384
total_weighteddmin	1	0.17413	0.23842	0.73	0.4653	0.15918	1.09	0.2742	0.03035
total_weighteddmaj	1	-0.53743	0.44155	-1.22	0.2238	0.11147	-4.82	<.0001	-0.03274
total_weightedfa	1	-0.01981	0.32483	-0.06	0.9514	0.23123	-0.09	0.9317	-0.00176
total_weightedfa1	1	0.98892	0.52239	1.89	0.0586	0.13070	7.57	<.0001	0.05071
total_weightedfb	1	-0.02197	0.17827	-0.12	0.9019	0.19861	-0.11	0.9119	-0.00355
total_weightedfc	1	-0.00753	0.13941	-0.05	0.9569	0.14195	-0.05	0.9577	-0.00181
total_weightedform2	1	0.31846	0.06985	4.56	<.0001	0.07351	4.33	<.0001	0.27969
total_weightedform3	1	0.32311	0.08996	3.59	0.0003	0.09404	3.44	0.0006	0.15873
inte6	1	-1.29275	0.72046	-1.79	0.0730	0.42069	-3.07	0.0022	-0.06513
inte20	1	-0.48637	0.46897	-1.04	0.2999	0.32321	-1.50	0.1326	-0.03460
inte5	1	-1.06025	0.80765	-1.31	0.1895	0.36056	-2.94	0.0033	-0.04628

However, after adding it to the model, it would seem that unit selling price does not have a non-linear effect on switching because its pivalue > 0.05. Therefore we will remove that and re-run the regression.

Root MSE	0.26494	R-Square	0.0729
Dependent Mean	0.22868	Adj R-Sq	0.0615
Coeff Var	115.85759		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent Standard Error	Heteroscedasticity Consistent t Value	Heteroscedasticity Consistent Pr > |t|	Standardized Estimate
Intercept	1	0.23521	0.01861	12.64	<.0001	0.01897	12.40	<.0001	0
total_weightedprice	1	0.01084	0.00380	2.85	0.0044	0.00262	4.14	<.0001	0.08002
total_weightedtype	1	-0.01589	0.00865	-1.84	0.0664	0.00843	-1.88	0.0597	-0.05918
total_weightedbrand2	1	-0.46946	0.06894	-6.81	<.0001	0.07331	-6.40	<.0001	-0.40491
total_weightedbrand3	1	-0.20153	0.06986	-2.88	0.0040	0.07404	-2.72	0.0066	-0.15655
total_weightedpr	1	0.03844	0.08058	0.48	0.6334	0.07108	0.54	0.5888	0.01408
total_weighteddmin	1	0.17278	0.23826	0.73	0.4685	0.15936	1.08	0.2785	0.03011
total_weighteddmaj	1	-0.54310	0.44075	-1.23	0.2181	0.10755	-5.05	<.0001	-0.03309
total_weightedfa	1	-0.02520	0.32392	-0.08	0.9380	0.22937	-0.11	0.9125	-0.00224
total_weightedfa1	1	0.98969	0.52219	1.90	0.0583	0.13069	7.57	<.0001	0.05075
total_weightedfb	1	-0.02205	0.17820	-0.12	0.9015	0.19855	-0.11	0.9116	-0.00357
total_weightedfc	1	-0.00704	0.13934	-0.05	0.9597	0.14206	-0.05	0.9605	-0.00169
total_weightedform2	1	0.31701	0.06956	4.56	<.0001	0.07313	4.33	<.0001	0.27842
total_weightedform3	1	0.32080	0.08941	3.59	0.0003	0.09337	3.44	0.0006	0.15759
inte6	1	-1.29357	0.72019	-1.80	0.0727	0.42001	-3.08	0.0021	-0.06517
inte20	1	-0.48587	0.46880	-1.04	0.3002	0.32323	-1.50	0.1330	-0.03456
inte5	1	-1.05819	0.80731	-1.31	0.1902	0.36109	-2.93	0.0034	-0.04619

Some product characteristics were added in to control for the possibility that an individual switches because of product effects (brand, packaging and forms) rather than because of advertising.

- The adjusted R Square of the model is 0.0615.
- There are no variables with VIF > 10 or condition index > 100 thus indicating no multicollinearity.

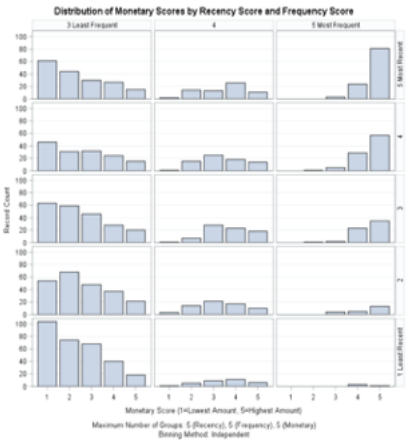
K) Interpreting Significant Coefficients

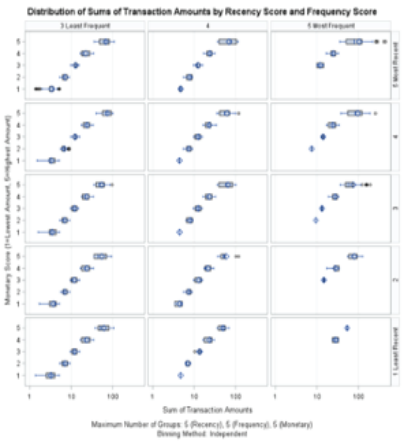
- For every unit dollar increase in selling price, there is an associated **1% increase** in switching occurring, controlling for all other variables.
- Gillette users are associated with a **47% decrease** in switching occurring as compared to BIC, controlling for all other variables.
- Schick users are associated with a **20% decrease** in switching occurring as compared to BIC, controlling for all other variables.
- Major displays are associated with a **54% decrease** in switching occurring as compared to no displays, controlling for all other variables.
- Features with coupons are associated with a **98% increase** in switching occurring as compared to no features, controlling for all other variables.
- Twin blade products are associated with a **31.7% increase** in switching occurring as compared to single blades, controlling for all other variables.
- Triple blade products are associated with a **32% increase** in switching occurring as compared to single blades, controlling for all other variables.

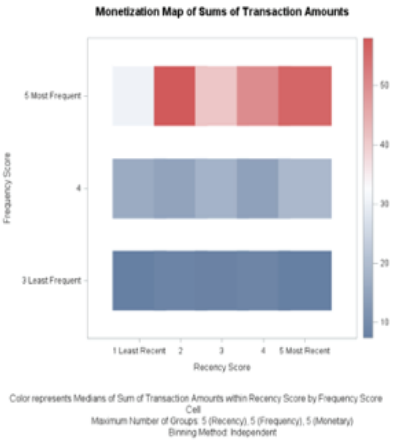
Business Insight: As seen from the low R-squared, an individual's decision to switch is influenced by more than advertising and the product characteristics. However, the most importance advertising variable are features with coupons. This seems to be the best incentive to promote switching.

Part 4 - Recency, Frequency & Monetary

L) Perform RFM modeling to identify loyal customers







Correlation between R, F and M

The CORR Procedure

3 Variables: recency_score frequency_score monetary_score

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
recency_score	1659	2.99699	1.42885	4972	1.00000	5.00000	Recency Score (1=Least Recent, 5=Most Recent)
frequency_score	1659	3.53225	0.77000	5860	3.00000	5.00000	Frequency Score (3=Least Frequent, 5=Most Frequent)
monetary_score	1659	2.99940	1.41400	4976	1.00000	5.00000	Monetary Score (1=Lowest Amount, 5=Highest Amount)

Pearson Correlation Coefficients, N = 1659
Prob > |r| under H0: Rho=0

	recency_score	frequency_score	monetary_score
recency_score Recency Score (1=Least Recent, 5=Most Recent)	1.00000	0.34025 <.0001	0.21553 <.0001
frequency_score Frequency Score (3=Least Frequent, 5=Most Frequent)	0.34025 <.0001	1.00000	0.57973 <.0001
monetary_score Monetary Score (1=Lowest Amount, 5=Highest Amount)	0.21553 <.0001	0.57973 <.0001	1.00000

Based on the correlation between R, F and M, we have determined that all 3 components are required.

Looking at the distribution of monetary score by recency and frequency score, individuals are clustered into the 2 extremes (left and right). On the extreme left lies the majority of the sample with low frequency and low monetary value. On the extreme right are individuals who have a high frequency and high monetary value. This trend stays the same regardless of recency.

The monetization map shows sum of transactions. There is a particular group who has high monetary score but a poor recency score. This trend stays the same regardless of recency.

To account for all of these trends, a loyal customer will be determined based on:

1. The total RFM score must be greater than or equal to 12
2. Each component of RFM must be greater than or equal to 4

With this criteria in place, we have:

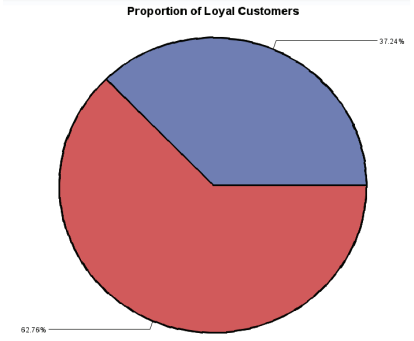
1. 1896 out of 3021 individuals classified as not loyal (62.76%)
2. 1125 out of 3021 individuals classified as loyal (37.24%)

By breaking down the loyal customers by brand, Gillette is observed to have the most customers who are both frequent and high-spenders. It is also the brand with the least frequent customers.

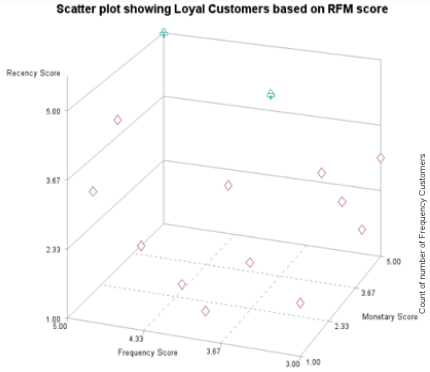
Proportion of Loyal Customers

The FREQ Procedure

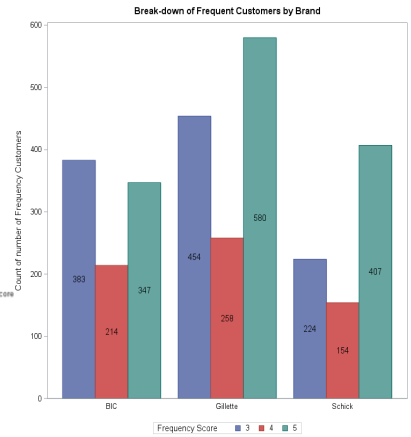
loyal	Frequency
0	1896
1	1125

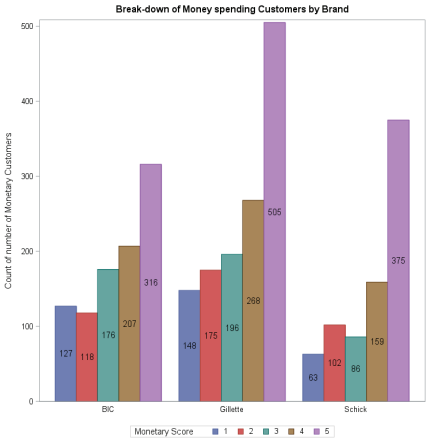


Business Insight: Gillette sells good quality shavers that are durable but expensive. To find the low-hanging fruits, with demographics similar to loyal customers, we will need to proceed to the next section.



The green spades represents the clusters of loyal customers based on RFM scores, while the red diamonds represents non-loyal customers.





Part 5 - Determining Loyal Customer Traits

M) Perform Logistic Regression of Loyal Customers on Demographics

loyal_customer = outlet + age group of head houshold male (agm) + age group of head household female (agf) + income status + education level male (elm) + education level female (elf) + marital status + residential status

Note: Reference levels are as such - outlet = GR, agm ='18 - 24 years old', agf = '18 - 24 years old', income status ='$00,000 to $ 9,999 per yr', elm ='Some grade school or less', elf = 'Some grade school or less', marital = 'Single', status = 'Renter'

Number of Observations Read	3021
Number of Observations Used	3021

Response Profile

Ordered Value	loyal_customer	Total Frequency
1	Yes	1125
2	No	1896

Probability modeled is loyal_customer='Yes'.

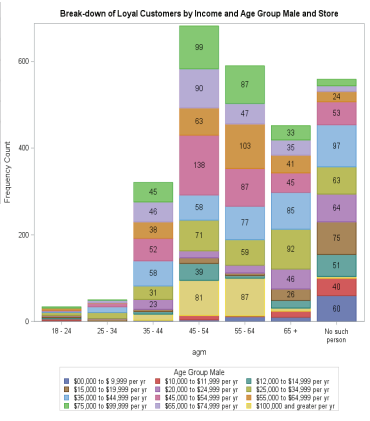
Analysis of Maximum Likelihood Estimates

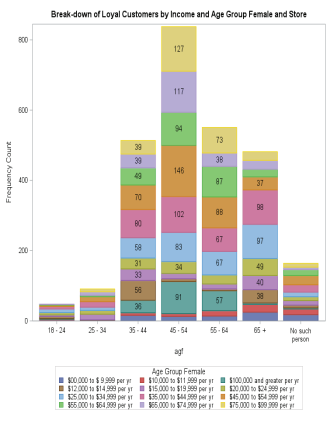
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-10.1545	172.2	0.0035	0.9530	0.000
outlet	DR	1	-0.7485	0.1402	28.5209	<.0001	0.473
agm	25 - 34	1	0.6846	0.5173	1.7518	0.1857	1.983
agm	35 - 44	1	-0.4934	0.4789	1.0614	0.3029	0.611
agm	45 - 54	1	-0.0970	0.4769	0.0414	0.8388	0.908
agm	55 - 64	1	-0.5723	0.4796	1.4248	0.2326	0.564
agm	65 +	1	-0.4472	0.4887	0.8374	0.3602	0.639
agm	No such person	1	-2.7076	0.7952	11.5948	0.0007	0.067
agf	25 - 34	1	0.8452	0.4731	3.1919	0.0740	2.328
agf	35 - 44	1	1.0019	0.4573	4.7997	0.0285	2.724
agf	45 - 54	1	1.2824	0.4629	7.6738	0.0056	3.605
agf	55 - 64	1	1.4633	0.4712	9.6448	0.0019	4.320
agf	65 +	1	1.3042	0.4826	7.3014	0.0069	3.685
agf	No such person	1	1.6106	0.7631	4.4552	0.0348	5.006
inc_status	$10,000 to $11,999 per yr	1	-0.8296	0.4220	3.8646	0.0493	0.436
inc_status	$100,000 and greater per yr	1	0.4953	0.3203	2.3915	0.1220	1.641
inc_status	$12,000 to $14,999 per yr	1	0.9231	0.3194	8.3505	0.0039	2.517
inc_status	$15,000 to $19,999 per yr	1	0.8900	0.3074	8.3814	0.0038	2.435
inc_status	$20,000 to $24,999 per yr	1	0.00674	0.3114	0.0005	0.9827	1.007
inc_status	$25,000 to $34,999 per yr	1	0.3025	0.2866	1.1141	0.2912	1.353
inc_status	$35,000 to $44,999 per yr	1	0.5792	0.2833	4.1805	0.0409	1.785
inc_status	$45,000 to $54,999 per yr	1	0.3906	0.2898	1.8165	0.1777	1.478
inc_status	$55,000 to $64,999 per yr	1	0.3943	0.2990	1.7397	0.1872	1.483
inc_status	$65,000 to $74,999 per yr	1	0.3307	0.3048	1.1769	0.2780	1.392
inc_status	$75,000 to $99,999 per yr	1	0.2303	0.3046	0.5723	0.4494	1.259
elm	Completed grade school	1	10.3240	172.2	0.0036	0.9522	30454.67
elm	Graduated from college	1	9.5947	172.2	0.0031	0.9556	14686.22
elm	Graduated high school	1	9.5724	172.2	0.0031	0.9557	14363.42
elm	No such head of household	1	11.6211	172.2	0.0046	0.9462	111423.5
elm	Post graduate work	1	10.2063	172.2	0.0035	0.9527	27073.13
elm	Some college	1	9.4675	172.2	0.0030	0.9562	12932.00
elm	Some high school	1	9.4870	172.2	0.0030	0.9561	13187.50
elm	Technical school	1	10.2610	172.2	0.0036	0.9525	28594.78
elf	Completed grade school	1	-1.0566	0.9905	1.1377	0.2861	0.348
elf	Graduated from college	1	-1.3811	0.8328	2.7503	0.0972	0.251
elf	Graduated high school	1	-0.9139	0.8253	1.2264	0.2681	0.401
elf	No such head of household	1	-1.7665	1.0213	2.9916	0.0837	0.171
elf	Post graduate work	1	-1.4613	0.8419	3.0128	0.0826	0.232
elf	Some college	1	-1.5897	0.8300	3.6678	0.0555	0.204
elf	Some high school	1	-1.9286	0.8640	4.9826	0.0256	0.145
elf	Technical school	1	-1.0780	0.8317	1.6801	0.1949	0.340
marital	Divorced	1	0.3235	0.1952	2.7471	0.0974	1.382
marital	Married	1	0.1313	0.1944	0.4564	0.4993	1.140
marital	Separated	1	0.6051	0.3446	3.0842	0.0791	1.831
marital	Widowed	1	0.0919	0.2583	0.1266	0.7219	1.096
status	Owner	1	-0.2169	0.1201	3.2639	0.0708	0.805

Business Insight: As managers, our target audience will therefore be households with female heads in the household over the age of 35 and who shops at the grocery store. Less focus should be paid to households with income between $10 - 11.9k per yr, besides this group accounts only for a small segment of the market.

N) Interpreting Significant Coefficients

- Individuals who shops at grocery stores are 1/0.473 = **2.11 times more likely** to be loyal customers as compared to those who shops at drug stores.
- Households with female heads between the age of 35-44 years are **2.72 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with female heads between the age of 45-54 years are **3.61 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with female heads between the age of 55-64 years are **4.32 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with female heads over 65+ years are **3.69 times more likely** to be loyal customers as compared to households with female heads between 18-24 years.
- Households with combined income between $0 - 9.9k per year are **1/0.436 = 2.29 times more likely** to be loyal customers as compared to households with combined income between $10 - 11.9k per year.
- Households with combined income between $15 - 19.9k per year are **2.44 times more likely** to be loyal customers as compared to households with combined income between $0 - 9.9k per year.
- Households with combined family income between $35 - 44.9k per year are **1.79 times more likely** to be loyal customers as compared to households with combined family income between $0 - 9.9k per year.
- Households with female head having an education level of some grade school or less are **1/0.145 = 6.9 times** more likely to be loyal customers as compared to households with female head having an education level of some high school.





Part 5 - Determining Loyal Customer Traits

O) Interpret Fit of Model using -2LogL, AIC and SC

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	3991.032	3808.630
SC	3997.045	4085.244
-2 Log L	3989.032	3716.630

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	272.4016	45	<.0001
Score	256.4774	45	<.0001
Wald	229.9579	45	<.0001

Using the -2LogL, we can calculate an approximate R-squared value:

(3989.032 - 3716.63) / (3989.032) = 0.068 (6.8%)

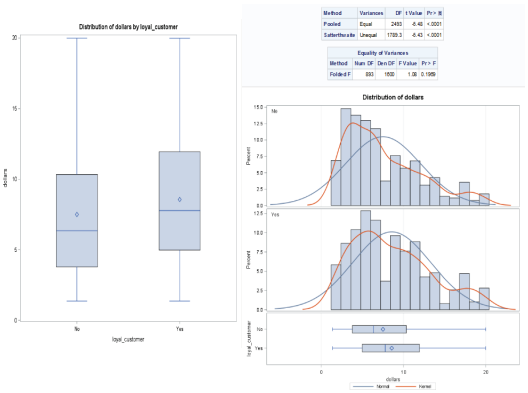
Two models are under consideration here. 1) Intercept only 2) Intercept and Covariates. Since AIC and -2LogL are all lower for the 'intercept-covariate' model as compared to the 'intercept-only' model, the additional variables helps to improve the fit of the model.

P) Percent Concordant

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	68.1	Somers' D	0.362
Percent Discordant	31.9	Gamma	0.362
Percent Tied	0.1	Tau-a	0.169
Pairs	2133000	c	0.681

The percentage concordant is 68.1%. To find the percentage concordant, we look at all possible pairs of observations and select a 'Yes' and 'No'. We will then calculate the probability using the model, and if Pr(Y) > Pr(N) then it is concordant.

R) Perform t-test to determine if mean unit sales is significantly different between loyal and non-loyal customers



The mean unit sales for loyal customer is $8.50 while the mean unit sales for non-loyal customer is $7.50. Since p-value of t-test is < 0.05, this $1 difference is significant.

Business Insight: Using this, we hypothesize that the cost spent on converting an individual should not exceed $1.

Q) Confusion Matrix of the Model

Confusion Matrix for Loyal Customers

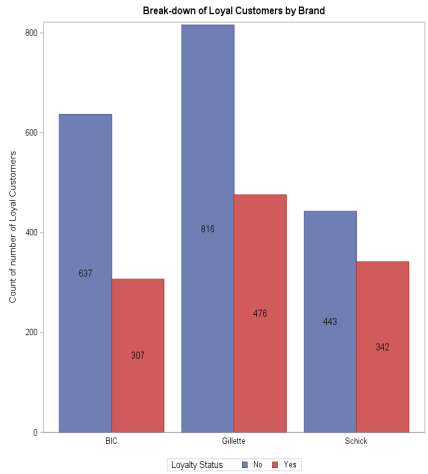
The FREQ Procedure

Frequency	Table of loyal_customer by loyal_hat		
	loyal_hat		
loyal_customer	no	yes	Total
No	1666	230	1896
Yes	806	319	1125
Total	2472	549	3021

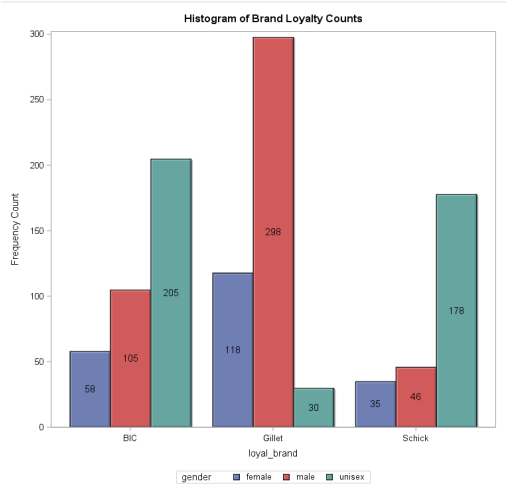
- **Accuracy:**
  (1666 + 319) / 3021 = 0.657 (65.7%)
- **Misclassification:**
  (230 + 806) / 3021 = 0.343 (34.3%)

Part 6 - Determining Choice of Loyal Customers

S) Break down of Loyal Customers using Brands



BiC has the highest ratio of non-loyal customers. For every 1 loyal BiC customer, there are 2 non-loyal customers. On the other hand, Schick has the best ratio of loyal to non-loyal customer of 1:1.2.



There are a total of 1073 loyal customers. It would seem that Gillette main focus is in the male market, since it sells the most male shavers. Whereas BIC and Schick are do not seem to have a specific targeted gender.

Business Insight:
BiC would benefit the most if it attempts a targeted marketing strategy on households with female above 35 years and older. If it attempts to capture households with males, it would be in direct competition with Gillette, which is currently the strongest player in the shavers market.

Part 6 - Determining Choice of Loyal Customers

T) Regressing Brand picked by Loyal Customers on Product Characterisitics

loyal_brand = L2 + Form + Gender + Package

Note: Reference levels are as such - L2 = cartridges, Form ='Single blades', Gender = 'Unisex', Package = 'Pack'

Analysis of Maximum Likelihood Estimates

Parameter		loyal_brand	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi Sq	Exp(Est)
Intercept		Schick	1	0.8155	0.6251	1.7017	0.1921	2.260
Intercept		Gillet	1	-5.7795	0.6668	75.1204	<.0001	0.003
L2	DISPOSABLE	Schick	1	-2.7084	0.3847	49.5530	<.0001	0.067
L2	DISPOSABLE	Gillet	1	0.1728	0.2910	0.3528	0.5526	1.189
FORM	TRIPLE BLADE	Schick	1	2.2580	0.5932	14.4908	0.0001	9.564
FORM	TRIPLE BLADE	Gillet	1	2.9565	0.5846	25.5768	<.0001	19.230
FORM	TWIN BLADE	Schick	1	2.6143	0.4008	42.5381	<.0001	13.657
FORM	TWIN BLADE	Gillet	1	4.0737	0.3773	116.5711	<.0001	58.775
gender	female	Schick	1	-2.3173	0.4189	30.6012	<.0001	0.099
gender	female	Gillet	1	2.4303	0.3355	52.4822	<.0001	11.363
gender	male	Schick	1	-1.9376	0.3639	28.3451	<.0001	0.144
gender	male	Gillet	1	3.4181	0.3157	117.2155	<.0001	30.511
PACKAGE	BAG	Schick	1	1.2817	1.2306	1.0848	0.2976	3.603
PACKAGE	BAG	Gillet	1	1.4821	0.7811	3.6004	0.0578	4.402
PACKAGE	BLISTER PACK	Schick	1	13.0534	579.8	0.0005	0.9820	466697.6
PACKAGE	BLISTER PACK	Gillet	1	15.3756	579.8	0.0007	0.9788	4759014
PACKAGE	CARDBOARD PEG SLEEVE	Schick	1	1.2162	1.4776	0.6775	0.4105	3.374
PACKAGE	CARDBOARD PEG SLEEVE	Gillet	1	0.1942	1.1158	0.0303	0.8618	1.214
PACKAGE	PEG BAG	Schick	1	-0.0679	0.4307	0.0249	0.8747	0.934
PACKAGE	PEG BAG	Gillet	1	0.1220	0.3837	0.1010	0.7506	1.130
PACKAGE	PEG CARD	Schick	1	-0.7618	0.3372	5.1046	0.0239	0.467
PACKAGE	PEG CARD	Gillet	1	-0.0597	0.3527	0.0286	0.8656	0.942
PACKAGE	PEGGED BOX	Schick	1	0.4302	0.6136	0.4914	0.4833	1.538
PACKAGE	PEGGED BOX	Gillet	1	2.0917	0.6012	12.1059	0.0005	8.098

Number of Observations Read	1073
Number of Observations Used	1073

Response Profile

Ordered Value	loyal_brand	Total Frequency
1	Schick	259
2	Gillet	446
3	BIC	368

Logits modeled use loyal_brand='BIC' as the reference category.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2310.976	1510.566
SC	2320.933	1630.043
-2 Log L	2306.976	1462.566

Confusion Matrix for Brand Loyalty

The FREQ Procedure

Frequency / Percent / Row Pct / Col Pct	Table of loyal_brand by _INTO_			
	INTO (Formatted Value of the Predicted Response)			
loyal_brand	BIC	Gillet	Schick	Total
BIC	260 24.23 70.65 72.22	77 7.18 20.92 14.58	31 2.89 8.42 16.76	368 34.30
Gillet	25 2.33 5.61 6.94	405 37.74 90.81 76.70	16 1.49 3.59 8.65	446 41.57
Schick	75 6.99 28.96 20.83	46 4.29 17.76 8.71	138 12.86 53.28 74.59	259 24.14
Total	360 33.55	528 49.21	185 17.24	1073 100.00

Confusion Matrix for Brand Loyalty

The MEANS Procedure

Analysis Variable : Match	
	Mean
	0.7483691

- **Accuracy:** (260 + 405 + 138) / 1073 = 0.748 (74.8%)

U) Interpret Fit of Model using -2LogL, AIC and SC

Using the -2LogL, we can calculate an approximate R-squared value:

(2306.976 - 1462.566) / (2306.976) = 0.366 (36.6%)

Overall, the additional variables seem to have significantly improved based on the AIC, SC and -2 Log L score against the intercept only model.

V) Interpreting Significant Coefficients

- Individuals are (1/0.067) = **14.9 times more likely** to choose a cartridge over disposables for Schick relative to BIC, controlling for all other variables.
- Individuals are **9.56 times more likely** to choose triple blades over singles for Schick relative to BIC, controlling for all other variables.
- Individuals are **19.23 times more likely** to choose triple blades over singles for Gillette relative to BIC, controlling for all other variables.
- Individuals are **13.66 times more likely** to choose twin blades over singles for Schick relative to BIC, controlling for all other variables.
- Individuals are **58.77 times more likely** to choose twin blades over singles for Gillette relative to BIC, controlling for all other variables.
- Individuals are (1/0.099) = **10.1 times more likely** to choose unisex shavers over female for Schick relative to BIC, controlling for all other variables.
- Individuals are **11.4 times more likely** to choose female shavers over unisex for Gillette relative to BIC, controlling for all other variables.
- Individuals are (1/0.144) = **6.9 times more likely** to choose unisex shavers over male for Schick relative to BIC, controlling for all other variables.
- Individuals are **30.5 times more likely** to choose male shavers over unisex for Gillette relative to BIC, controlling for all other variables.
- Individuals are **8.09 times more likely** to choose pegged box over pack for Gillette relative to BIC, controlling for all other variables.

W) Performing Chi-sq test to test for independence between brand selection of loyal customers and product characteristics

Frequency / Percent / Row Pct / Col Pct	Table of loyal_brand by gender			
	gender			
loyal_brand	female	male	unisex	Total
BIC	58 5.41 15.76 27.49	105 9.79 28.53 23.39	205 19.11 55.71 49.64	368 34.30
Gillet	118 11.00 26.46 55.92	298 27.77 66.82 66.37	30 2.80 6.73 7.26	446 41.57
Schick	35 3.26 13.51 16.59	46 4.29 17.76 10.24	178 16.59 68.73 43.10	259 24.14
Total	211 19.68	449 41.85	413 38.49	1073 100.00

Statistics for Table of loyal_brand by gender

Statistic	DF	Value	Prob
Chi-Square	4	343.0125	<.0001
Likelihood Ratio Chi-Square	4	391.0404	<.0001
Mantel-Haenszel Chi-Square	1	0.8108	0.3679
Phi Coefficient		0.5654	
Contingency Coefficient		0.4922	
Cramer's V		0.3998	

Based on the Chi-sq test (associated p-value < 0.05) for independence, the choice of gender-specific products are influenced by brand. This supports our prior multinomial logisitic regression results.

Business Insight: Therefore, if BIC was to begin applying targeted marketing and compete with Schick, it should focus on selling unisex shavers instead of gender-specific products like Gillette.