Customer Segmentation & Marketing Strategy Advice

✓ Goal: Use data to advise business operations to better meet customer preferences

✓ Data: Customer personal information, product purchase information, promotional history

Conclusion 1:

There are 3 segments in the market, with Group 2 having the most purchasing power.

Hierarchical Clustering

1. **Use elbow finding** to determine the ideal number of segments.

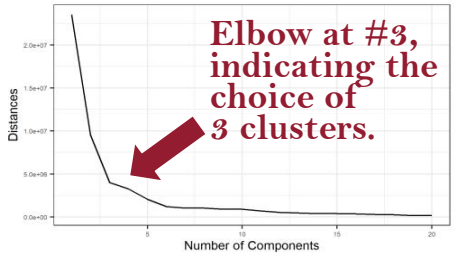


Figure1. Elbow finding

2. Select the **Ward.D linkage method** as the best clustering method due to well-separated clusters of similar size.

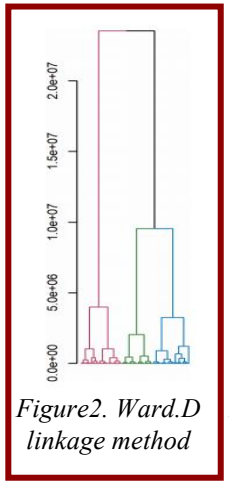


Figure2. Ward.D linkage method

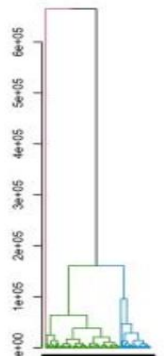


Figure3. Complete linkage method

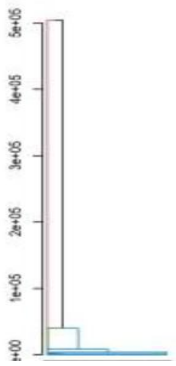


Figure4. Single linkage method

Deep dive into customer behavior

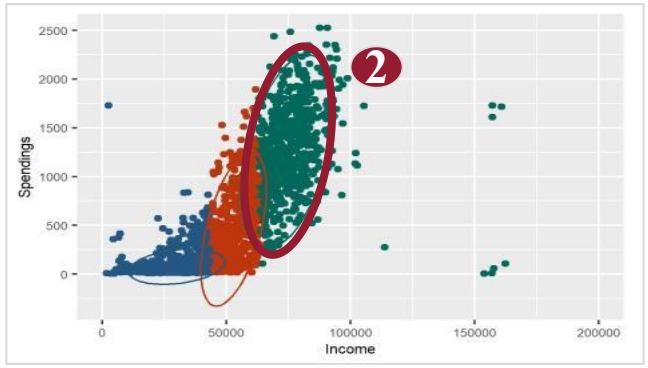


Figure5. Clusters on Spendings and Income

Take a closer look at Segment 2:

- Highest income & highest spend
- More likely to shop by catalog and in-store than online
- Highly likely to accept promotions
- Most do not have children

Figure6. Boxplots of shopping channels

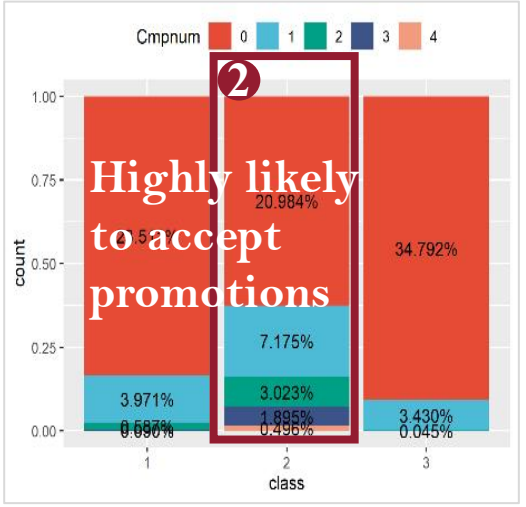


Figure7. # times accepting promotional activities

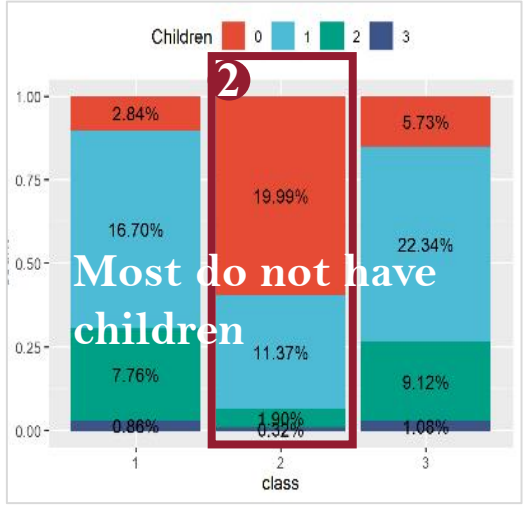


Figure8. Num of Children

Conclusion 2: Merchants should focus on improving online operations.

Conclusion 3: Promotion should target consumers who have fewer children and prefer web and catalog channels.

Multiple linear regression

Step 1: Data Preparation

Select and transform variables to be put into MLR.

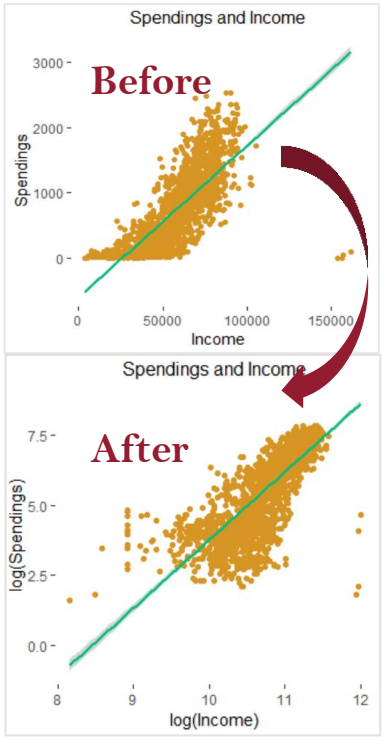


Figure1. Relationship between spendings & income

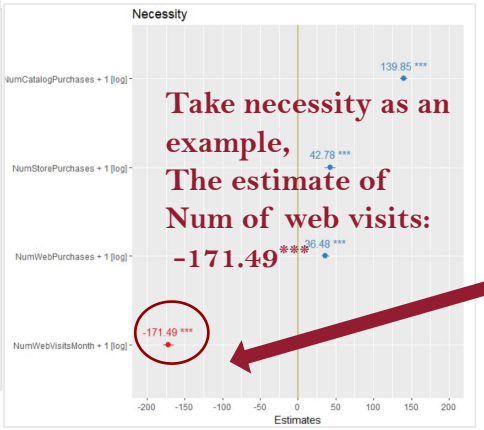
Step 2: MLR for Total Spend

The impact on total spend varies by shopping channel.

Table1. Regression Results

num	term	estimate	std.error	statistic	p-value
1	(Intercept)	-1.444008e+04	225.0652784	-64.1595344	0.000000e+00
2	Age	-1.136748e-01	0.2180082	-0.5214245	6.021903e-01
3	Child	-2.327788e+02	7.0949730	-32.8089750	2.220772e-159
4	log(Income)	1.346189e+03	20.8678385	64.5102473	0.000000e+00
5	log(NumCatalogPurchases + 1)	2.430744e+02	8.0103886	30.3448923	1.031154e-142
6	log(NumStorePurchases + 1)	-6.708839e+00	8.3055620	-0.8077526	4.194312e-01
7	log(NumWebPurchases + 1)	-7.946649e+01	11.4440853	-6.9438920	6.983111e-12
8	log(NumWebVisitsMonth + 1)	3.468809e+02	10.9429728	31.6989642	7.108005e-152

Step 3: MLR for spending on 3 different types of products



Divide products into 3 categories from necessity to luxury, and do MLR for each. For all three, there is a negative relationship between web visits and sales.

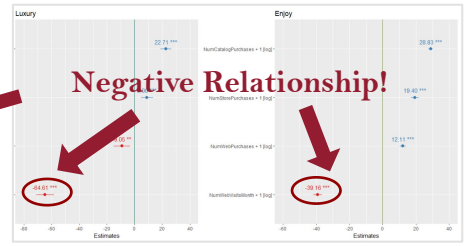


Figure2. The estimate of shopping channels for spending of 3 categories

Logistic regression

Create a logistic regression model to predict customer response to promotions

- ✓ Convert the factor variables to dummy variables
- ✓ Use stepwise search with AIC criterion to perform feature selection, 9 variables are selected

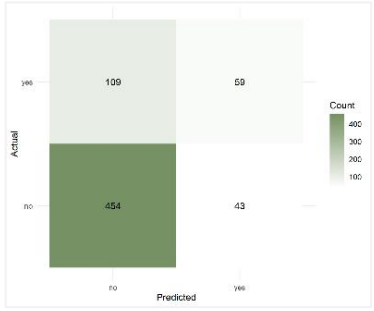


Figure3.Confusion matrix

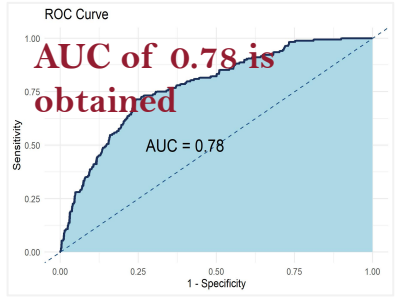


Figure4. ROC curve

Interpretation of results

We can target promotions to customers who visit the Web frequently, make more catalog purchases, make fewer in-store purchases, and have fewer children.

To summarize,

- ➢ Clustering: Basic customer characteristics
- ➢ MLR: Advice for improving shopping channels
- ➢ Logistic Regression: Advice on promotional activities