



EMORY
GOIZUETA
BUSINESS
SCHOOL

GOIZUETA
BEYOND

Marketing

Analytics

Pernalonga

EDA

404



hello!

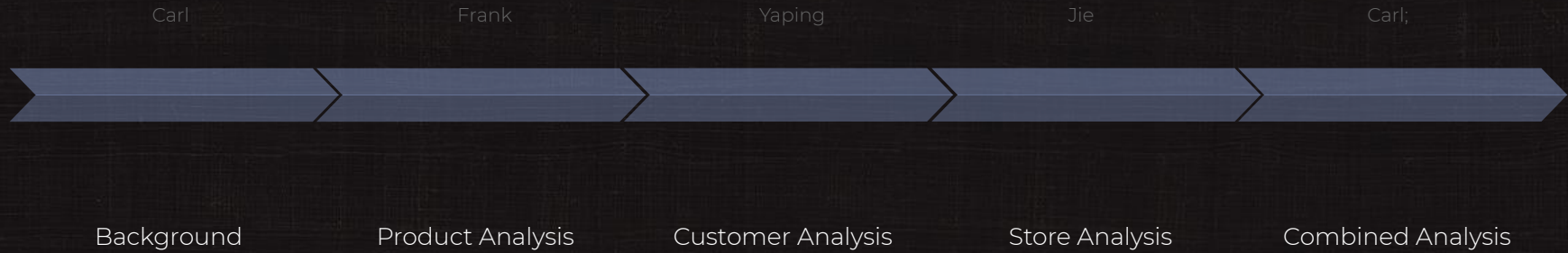
WE ARE
TEAM

404

We have conducted
comprehensive
marketing analysis of
Pernalonga's operations

1. Background

Presentation Overview



“

With over 10,000 products in 400+ categories, Pernalonga is an undisputed leader in the Lunitunia retail space. We have conducted segmentation analysis on Pernalonga's customers, products and stores for marketing analytics purposes.

Our Value-Add

- Regularly partners with suppliers to fund promotions (source of ~30% of sales)
 - Mostly in-store
 - Recently started partnering with select suppliers to experiment on personalized promotions.

We are here to offer insight on your customers, stores and products because:

- Personalized promotions are more **efficient**
 - Offers are **only** made to **targeted** individuals who **need** one to make the purchase
- Most in-store promotions make **temporary** price **reductions** on a product available to **all customers** whether or not a customer **actually needs** the incentive to make the purchase
- This **boost in efficiency** comes at the cost of **prior additional analysis** on customer transaction data
 - Determine the **most likely purchasers** on offers
 - **Maximize opportunity** for incremental sales and profits

10767 obs. x 7 vars.

Transaction Data

2,961,785 obs. x 12 vars.

Product Data

Customer & Store

Derived from Transaction Data

2. Product Analysis

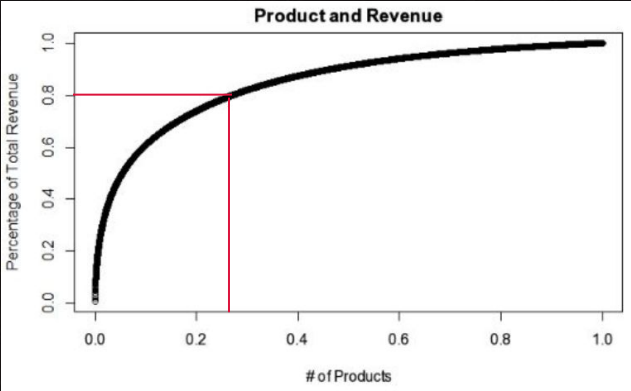
Which products are generating the most revenue?

- ◎ Purposes: identify the products which generate the most revenue
- ◎ Motivation: generate business recommendations for store layout, target advertisement, etc.
- ◎ Analysis Level: individual product, subcategory, category

Products that generated most revenue

Top 5 Products by Revenue	
999749469	602109.4
999956795	546554.8
999749894	530179.2
999455829	482758.4
999649801	416159.9
Bottom 3 Products by Revenue	
999244932	502.50
999749706	502.18
999556061	501.61

- Top products generate significantly more revenue than the bottom products.
- Is there a revenue gap between top selling items and bottom selling items?

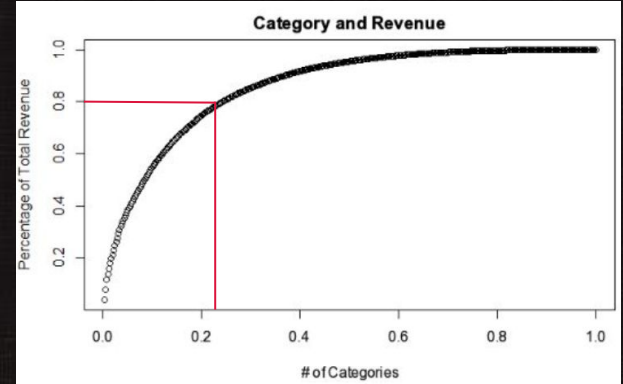


Pernalonga should spend the majority of time and money on the small percentage of products that bring the most revenue.

Revenue of categories and subcategories

Top 4 Subcategories by Revenue (after translation)	
PORK STANDARD FR	1975470
STEAK FR	1537885
NAC BEER WITH ALCOHOL	1025159
CAFE CAPSULES	923422
Top 4 Categories by Revenue	
FRESH PORK	2483963
FRESH BEEF	2401663
FRESH POULTRY MEAT	2379408
DRY SALT COD	1344207

- By taking a closer look at categories and subcategories, we found that revenue generators are: fine meat and alcohol.



- Similar trends exist for product categories: about 21% of categories contributed to 80% of total revenue.



Product Further Breakdown

Revenue Drivers	Customer Magnets	Most Transactions	Always Promoted
Top 5 Categories: Fresh Pork Fresh Beef Fresh Poultry Dry Salt Cod Fine Wines	Top 5 Categories: Fine Wafers Fresh Pork Fine Wines Yogurt Health Fruit Juices	Top 5 Categories: Bags Fine Wafers Fresh Pork Mineral Waters Fresh Poultry Meat	Top 5 Categories: Fine Wafers Yogurt Health Banana Fresh Uht Milk Fine Wines
Top 5 Brands: PrivateLabel NoLabel Frutas Vegetais Perciveis Carne Sourcing	Top 5 Brands: Private Label No Label Frutas Vegetais Perciveis Carne Mimosa	Top 5 Brands: Private Label No Label Frutas Vegetais Perciveis Carne Mimosa	Top 5 Brands: Private Label Frutas Vegetais No Label Mimosa Activia

- © Based on descriptive analysis, we found that fresh foods like pork, beef and poultry products generated the most revenue. They have two features: 1. They are necessities that people need on a frequent basis, 2. They tend to be expensive.
- © Compared to the revenue drivers, customer magnets like yogurts and fruit are products that customer purchase on a daily basis but tend to be less expensive. So they generate less revenue than revenue drivers.
- © As for most transactions, people tend to buy them in batches, for example shopping bags and mineral waters. We believe it is a less important metric compared to customer base and revenue.
- © The last class is called "always promoted". Fine wafers, yogurt, bananas, and milk are products that are promoted the most. This is because these products expire quickly, and need to be sold before the expiration date.



A closer look : K-Means Clustering

Apart from examining the product brands, categories and sub-categories, we also used K-means clustering to identify natural groupings based on product attributes. Here we took the following steps:

- ◎ Step 1: for each product id identify key features
 - Total Revenue
 - Total Transactions
 - Total Distinct Customers
 - Total Stores
 - Average Discount Rate
 - Percentage of Discount Products

	Total Revenue	Total Transactions	Total Distinct Customers	Total Stores	Average Discount Rate	Percentage of Discount Products
Minimum	500.0	3.0	2.0	1.0	0.0000432	0.0005537
1st Quartile	939.5	307	189.0	124.0	0.0227892	0.0362067
Median	1898.0	757	405.5	220.0	0.1095774	0.0715926
Mean	5802.4	2764	725.8	221.9	0.1587617	0.1185600
3rd Quartile	4591.7	2061	882.8	324.0	0.2787683	0.1564843
Maximum	602109.4	769890	7862.0	419.0	0.6895196	1.0000000

A few things can be observed from the above summary table:

The top quarter of products attracts far more customers and contributes much more revenue than the bottom three quarters of products, which again confirms the 80/20 rule; average discount rate for all products is about 10% while the max discount rate is as high as 69%; for most of the products, only a small percentage of times the products are on discount, but there are certain products have discounts all the time.

A closer look : K-Means Clustering

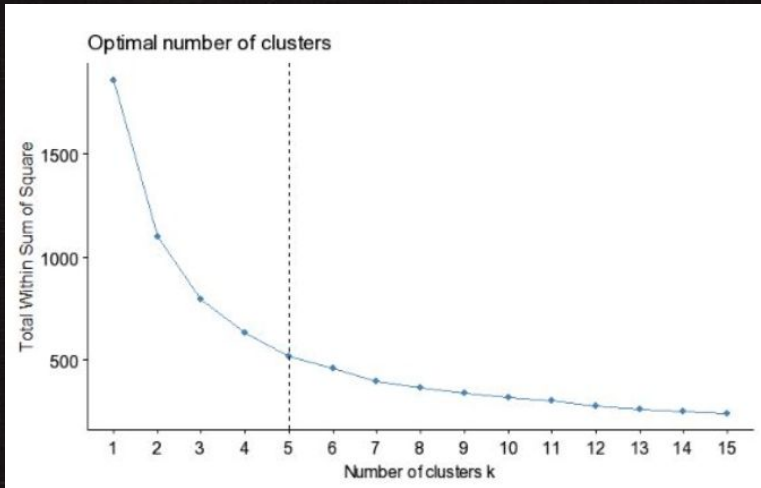


Figure: Elbow Graph

We chose $k=5$ as our cluster number

- ◎ Step 2: Normalize every key attribute above using min-max standardization, before we use distance-based clustering technique.
- ◎ Step 3 : Use both elbow graph and average silhouette graph to choose the optimal number of k for clustering. While the elbow method looks at the percentage of variance explained as a function of the number of clusters, the silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). Both methods are useful for determining the number of k .

A closer look : K-Means Clustering

◎ Step 4: Get the cluster results and Interpret

Cluster	Total Revenue	Total Transactions	Total Distinct Customers	Total Stores	Average Discount Rate	Percentage of Discount Products	Example
1	High	High	High	High	Low	Low	"Apple"
2	Middle	Middle	High	High	Middle	Middle	"Tea"
3	Low	Low	Low	Low	High	High	"Toys"
4	Middle	Middle	Middle	Middle	Low	Low	"Soaps"
5	Low	Low	Low	Low	Low	Low	"DVD"

	Total Revenue	Total Transactions	Total Distinct Customers	Total Stores	Average Discount Rate	Percentage of Discount Products
1	0.035112	0.0155	0.281966	0.914481	0.123283	0.031123
2	0.008388	0.002594	0.115067	0.723884	0.4954	0.109235
3	0.001473	0.000348	0.023003	0.291503	0.479064	0.305507
4	0.004511	0.001809	0.073499	0.619741	0.094985	0.064234
5	0.001839	0.000733	0.0236	0.232097	0.074576	0.097621

Figure: avg. statistics for centroids

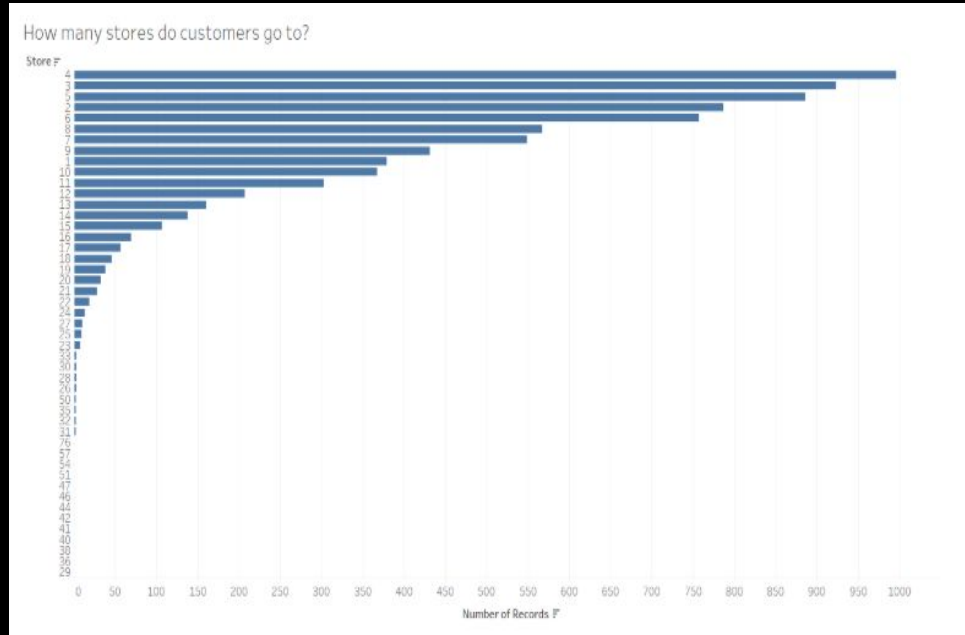
- ◎ 1. **Cash Drivers**: life essentials that contribute most revenue and attract customers
- ◎ 2. **Buy with discounts** : products with substitutes, usually purchased with discounts or by customers with fixed tastes
- ◎ 3. **Need help** : Constantly on discount, have a small audience
- ◎ 4. **Cheap necessities** : relatively cheap but very important products; don't need to be promoted
- ◎ 5. **Cold bench** : rarely purchased; customers are not sensitive to discounts either

3. Customer Analysis

Customer Exploratory Analysis and Segmentation

Customers' Loyalty to Stores

- ◎ The majority of people only went to less than 10 stores (6642), taking up 83.86% of the total customer population
- ◎ The number of people visiting over 30 stores is really small (23), taking up less than 0.3% of the whole customer population.
- ◎ And the extremely loyal customers, those who only visit a single store, takes up 4.79% of the total population.

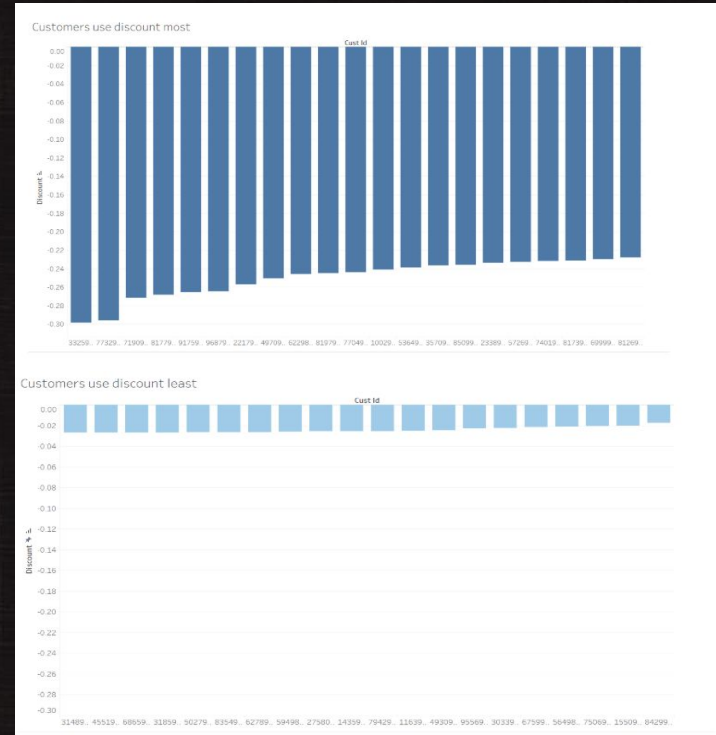


Customers' purchasing behavior on discounted products

How good a customer is at saving money=

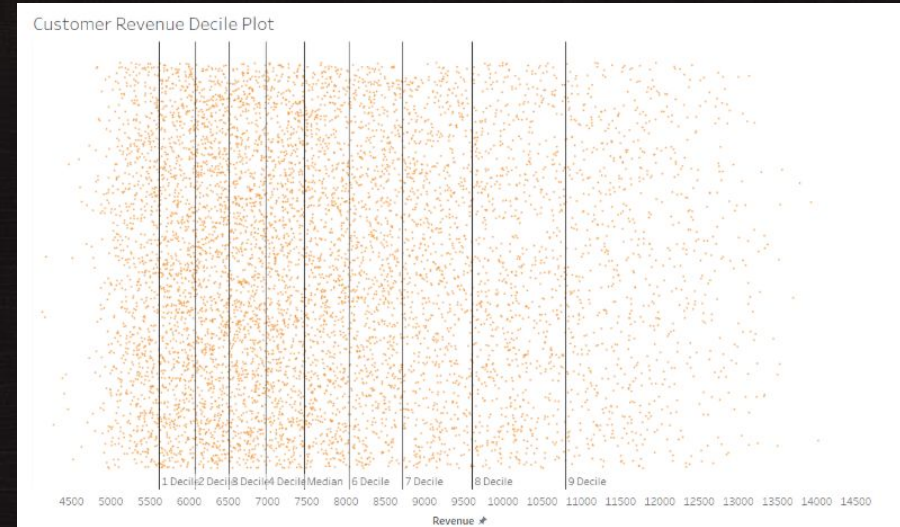
$\text{Sum} (\text{Product discount price} / \text{Product original price})$

The gap between the top 20 people who are really good at saving money and the bottom 20 customers who don't really purchase discount products is huge.



Revenue Contribution of Each Customer

- ◎ Detect revenue contribution of each customer by calculating the total money they spent on purchasing products
- ◎ Visualize customer revenue contribution distribution using decile plot
- ◎ Results: revenue contribution gap between people becomes wider and wider as the amount of revenue contribution goes up





Customer Segmentation (K-Means Model)

- ◎ Steps 1: Identify the attributes used for clustering, for each customer, we calculate:
 - Total Revenue
 - Total Transactions
 - Total Distinct products
 - Total Stores visited
 - Average Discount Rate
 - Number of Discount Products
 - Percentage of Discount Products
- ◎ Step 2: Normalize all features using min-max normalization
- ◎ Step 3: Calculate correlations between each feature and exclude those with high correlation
- ◎ Step 4: Use the elbow method to determine the best K (5)
- ◎ Step 5: Fit the k-means model with K=5 clusters
- ◎ Step 6: Get business understanding of each cluster

Attributes for each cluster centroid

Cluster	Number of Customer	Total Revenue	Total Transactions	Total categories	Total Stores	Average Discount Rate
A	1894	0.2261363	0.3806635	0.6183601	0.08059838	0.4921419
B	1889	0.2720808	0.3684890	0.5738651	0.06192342	0.2578444
C	1069	0.3949973	0.6687389	0.6233128	0.05552853	0.3076482
D	1677	0.4533388	0.3925081	0.7159229	0.09516995	0.4211302
E	1391	0.7088462	0.4709178	0.7069522	0.07672178	0.3359648

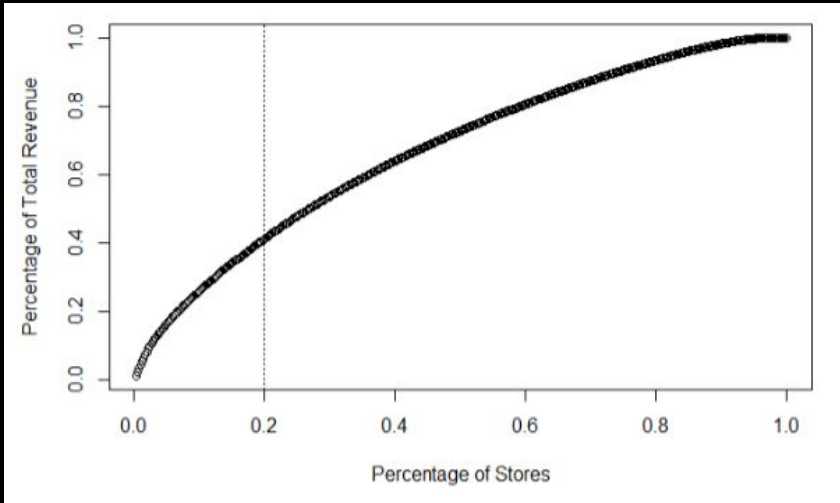
Business Understanding for Each Cluster

Cluster	Total Revenue	Total Transactions	Total categories	Total Stores	Average Discount Rate	
A	Low	Middel	Middle	Low	Middle	cherry-picker
B	Low	Low	Low	Low	Low	passers
C	Middle	High	Middle	Low	Low	nearby
D	Low	Middle	High	Low	Middle	wholesale cherry-pickers
E	High	Middle	High	Low	Low	high-valued customers

4. Store Analysis

Store-Level Exploratory Analysis and Segmentation

Store-Level Sales Doesn't Follow The 80/20 Rule



- ◎ Currently 421 stores in operation
- ◎ Huge gap between best performing stores and worst performing stores
 - Store 342 - \$786,521 in revenue from 30,352 transactions
 - Store 302 - \$4.15 in revenue from 1 transaction
- ◎ Top 20% of the stores generate about 40% of the total revenue

Best Performing Stores

Store ID	Total Volume (count)
324	592883
349	485421
345	441240
344	429165
346	429066

Store ID	Total Volume (kg)
345	77515.39
342	77355.42
349	73962.81
996	64125.80
588	64105.24

Store ID	Total Revenue (\$)
342	786521.10
345	718779.80
349	680640.40
344	624991.70
343	591942.20

Best Performing Stores

Store ID	Total Transaction
342	30352
349	25635
994	24516
347	22628
345	22607

Store ID	Total Customer
344	836
341	781
345	655
342	566
157	421

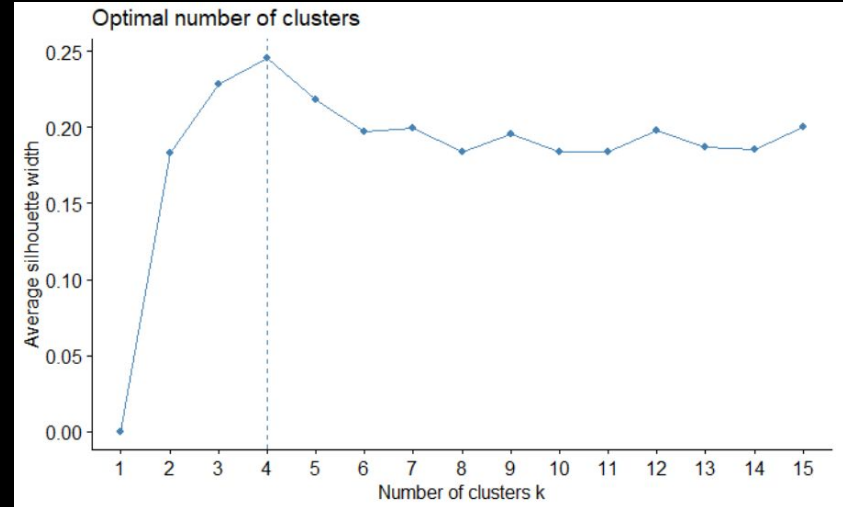
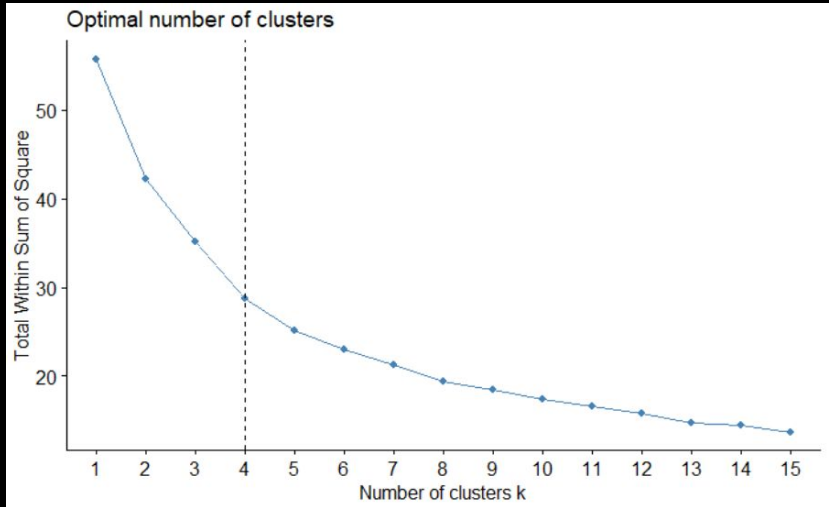


Selected Attributes for K-Means Store Segmentation

	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amount	Avg. Visit
Min	3274	25	686	0.082	9.809	2.457
1st Quartile	89573	76	4998	0.17	19.032	38.429
Median	124163	106	5815	0.187	22.068	55.773
Mean	149894	126	5733	0.187	22.086	61.795
3rd Quartile	183328	150	6612	0.205	24.861	80.464
Max	786521	836	9520	0.277	38.712	187.296



Selected K Value for K-Means Store Segmentation



Optimal K = 4



Numeric Result of K-Means Store Segmentation

Cluster	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amt	Avg. Visit
1	87375.070	98.991	4771.557	0.185	21.851	51.106
2	292545.100	172.739	7464.065	0.193	22.365	84.926
3	173114.910	131.595	6531.718	0.186	21.975	72.413
4	614919.610	475.500	9211.700	0.226	27.613	60.019



A Closer Look: Store Segmentation

Cluster	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amt	Avg. Visit
1	Low	Low	Low	Low	Low	Low
2	Mid	Mid	Mid	Low	Low	High
3	Lower Mid	Lower Mid	Mid	Low	Low	Upper Mid
4	High	High	High	High	High	Mid

7-11

Small-town Kroger

Large-town Kroger

Walmart Supercenter

© Cluster 1 - "7-11"

- Lowest in total revenue, total customer, total product, and average visit per customer
- Small in size but large in quantity, making them easy to find
- Small orders covering limited kinds of products
- 228 stores



A Closer Look: Store Segmentation

Cluster	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amt	Avg. Visit
1	Low	Low	Low	Low	Low	Low
2	Mid	Mid	Mid	Low	Low	High
3	Lower Mid	Lower Mid	Mid	Low	Low	Upper Mid
4	High	High	High	High	High	Mid

7-11

Small-town Kroger

Large-town Kroger

Walmart Supercenter

- ◎ Cluster 2 - "Small-town Kroger"
 - Mid level of total revenue and high average visit per customer
 - Probably have no or few competitors
 - People go there frequently but don't buy a lot
 - 46 stores



A Closer Look: Store Segmentation

Cluster	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amt	Avg. Visit
1	Low	Low	Low	Low	Low	Low
2	Mid	Mid	Mid	Low	Low	High
3	Lower Mid	Lower Mid	Mid	Low	Low	Upper Mid
4	High	High	High	High	High	Mid

7-11

Small-town Kroger

Large-town Kroger

Walmart Supercenter

- ◎ Cluster 3 - "Large-town Kroger"
 - Cover similar amounts of products as Cluster 2
 - Lower total revenue and total customer
 - Probably more competitors
 - 131 stores



A Closer Look: Store Segmentation

Cluster	Total Revenue	Total Customer	Total Product	Avg. Discount Rate	Avg. Transaction Amt	Avg. Visit
1	Low	Low	Low	Low	Low	Low
2	Mid	Mid	Mid	Low	Low	High
3	Lower Mid	Lower Mid	Mid	Low	Low	Upper Mid
4	High	High	High	High	High	Mid

7-11

Small-town Kroger

Large-town Kroger

Walmart Supercenter

- ◎ Cluster 4 - "Walmart Supercenter"
 - Attract the most customers
 - Generate the highest revenue
 - Offer the highest discount rate
 - Only 10 Stores



5. Result and Conclusion

“The Numbers Mason, What Do They Mean???”



Customer Segmentation (K-Means Model)

- ① With the products, customers and stores clustered and analyzed, we lastly take a look at the implications of these clusters and how they interact with other clusters.
- ② The correlation is calculated by looking at the percentage of each green category that the blue categories have bought. Thus, each vertical column should sum up to 100% the green category.

Customer – Product Correlation

Customer\Product	Cash-Drivers	Promotional Purchases	Constantly Needs Help	Necessities	High-Shelf
Cherry Picker		✓		✗	
Passers		✗	✓		
Lives Nearby				✓	✗
Wholesale Cherry Pickers			✗		✓
High-Value	✓	✓	✓	✓	✓

Customer – Store Correlation

Customer\Store	7-11	ST Kroger	LT Kroger	Walmart SC
Cherry Picker	✗	✓	✓	
Passers	✓			✗
Lives Nearby	✓	✗	✗	
Wholesale Cherry Pickers	✗	✗	✓	?
High-Value				✓



Product – Store Correlation

Product\Store	7-11	ST Kroger	LT Kroger	Walmart SC
Cash-Drivers	✓	✗		
Promotional Purchases		✗	✓	
Constantly Needs Help				✓
Necessities			✗	✓
High-Shelf	✗			✓



Next Steps

- ◎ Based on our analysis, here is what we suggest Pernalonga should do next:
 - Implement targeted marketing tactics that would've previously been ineffective if deployed to the entire product, customer or store groups.
 - Develop targeted marketing campaigns for each customer segment based on their purchase pattern and affinity to certain clusters of products.
 - Rank customers by their total value and maximize the revenue from the top customers with specialized discounts.
 - Rearrange store layouts to promote products with lower sales volume (e.g. constantly needs help) and to highlight cash-drivers.
 - Create loyalty programs to reward high-value customers, lock in cherry pickers and encourage all categories to purchase more.
 - Create collaborations between certain products and store locations (e.g. product demonstration stands, sampling, etc.) based on store category to further drive sales of the popular product segments or promote circulation of the lesser-popular product segments.



IN CONCLUSION

We are confident that our exploratory data analysis and thorough segmentation analysis has laid the perfect foundation for future successful marketing strategies.



Thank you!

ANY
QUESTIONS?

We are Team 4(04)

404