Emory University

Master of Science in Business Analytics Program

Data-Driven Marketing Insights for Pernalonga

Alice Jiang, Cassio Salge, Qiyu Wang

MKT 680 Marketing Analytics

Professor David Sackin and Alvin Lim

2021/2/13

**Executive Summary**

This report was commissioned to develop a marketing analysis that utilizes segmentation to create personalized promotions for Pernalonga’s customers. Pernalonga is an undisputed leader in the retail space of Lunitunia, with over 10,000 products in 400+ categories. In order to drive sales, Pernalonga finds itself in a place where it depends on promotions, as over 30 percent of sales come from promotions. In-store promotions are currently the majority of the promotions at Pernalonga. The issue with in-store promotions is that they offer customers temporary price reductions irrespective of their needs. Therefore, by offering promotions to those who would be willing to buy the items at full price, Pernalonga is losing part of its revenue. Since Pernalonga does not have a strong or thorough marketing campaign, by not effectively targeting stores, customers, or products with their promotions, they are losing potential revenue. Thus, we will help them better understand their customers, products and stores as their consulting team and how to develop a strong marketing campaign that will boost revenue through the use of segmentation.

**Data Understanding & Exploratory Data Analysis**

We were provided two files:

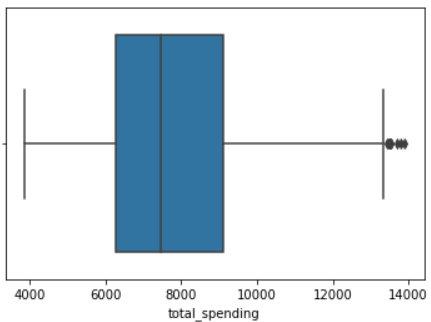
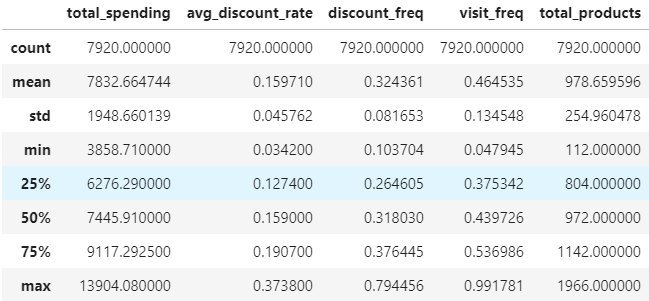
* transaction table which contains 2,961,785 observations and 12 variables covering transaction data from 2016 to 2017.
* product table which contains 10767 observations and 7 variables covering product information.

Some discrepancies were found in the data (i.e. stores with $5 in sales, transactions with random sales volumes and discounts applied incorrectly to products). These observations were removed. Also, we decided not to include bags when clustering. Customers that do not bring bags while shopping might need to buy them to carry out the products when leaving the store since bags are not given away for free. Once we cleaned and then understood the structure of the data, we started looking more in-depth into customers, products, and stores to explore whether there are any trends we could find:

**Customers**

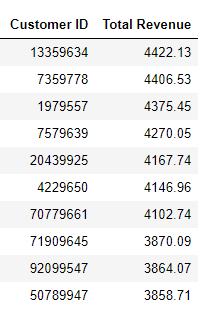
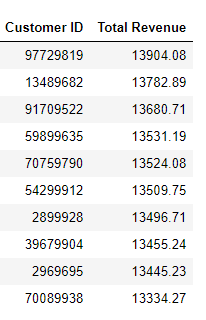
We have aggregated features by customer level in order to understand customer purchasing behavior and explore human factors that can be difficult to express with a single variable:

1. **Revenue analysis:** From the charts, we can see the revenue breakdown by customer. We recognized 7920 unique customers and the majority of them are either in or clustered around the interval between $6276 and $9117, with a mean of $7832. The top 25% can go as high as $13904, and the lower 25% can go down to $3858. The top and bottom consumers have a significant difference in spendings, which is definitely worth noting.



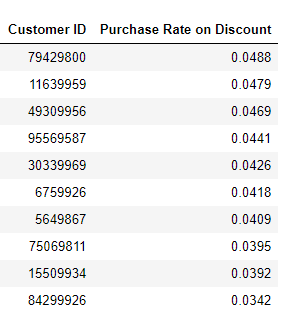
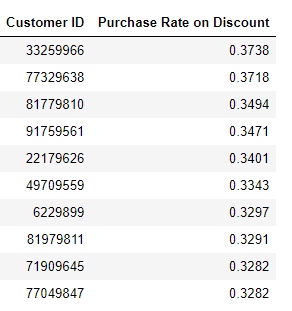
The top/bottom 10 revenue-driving customers are:

**TOP 10 BOTTOM 10**



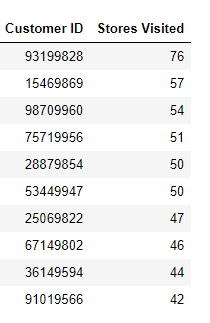
1. **Purchasing behavior on discounted products:**  In order to calculate the percentage of discount money as the total value of a product on an individual customer level, we used the total amount of discounts applied to the product in the transaction divided by sales amount for the product before discounts in the transaction to calculate the discount percentage of money should be spent for each customer. Below we list the top 10 people who saved the most money and the bottom 10 customers who saved the least. Based on our analysis, there is a large gap between the two groups of people.

**TOP 10**  **BOTTOM 10**



1. **Loyalty to store locations:**  We found that the majority of customers (6644) only went to 10 stores or less, taking up 83.89% of the total customer population. The number of customers visiting 30 stores or more is small (26), taking up less than 0.3% of the whole customer population. Customers who only visit a single store (381), take up 4.81% of the total population.

**TOP 10 BOTTOM 10**

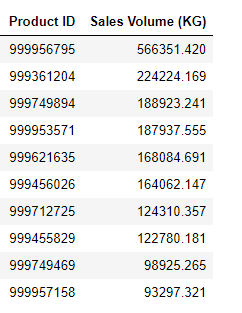
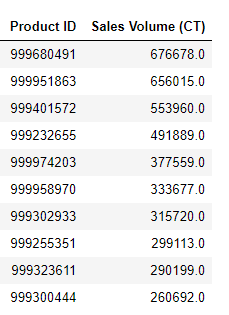


**Products**

There are 10755 distinct products. The top 20% of products have brought in $45,863,210.66 which represents 74% of total revenue of Pernalonga’s 421 stores. We recommend Pernalonga to spend the majority of resources on the products that bring the most revenue.

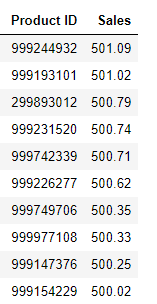
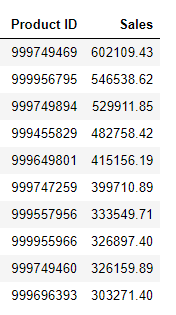
1. **Highest sales volume:**  We analyzed the sales quantity to find out most sold items in terms of counts and kilograms. We found that Mineral Waters had the highest number of sales in counts, followed by Pao Manufacture, Fresh UHT Milk, Yogurt Health and Special UHT Milk. For items sold in terms of kilograms, Fresh Poultry Meat has the highest sales followed by Fresh Pork, Banana, Fresh Beef and Apple. We also noticed that these are high-demand life necessities products which customers buy frequently.

**TOP 10 TOP 10**



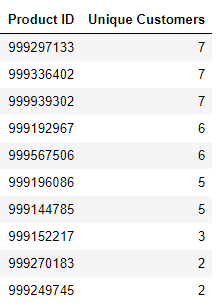
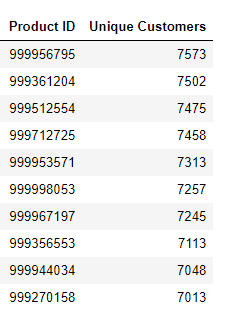
1. **Highest revenue:** Similarly, we see overlap in products that bring the highest revenue. Fresh Pork brings in the highest revenue, followed by Fresh Beef, Fresh Poultry Meat, Dry Salt Cod and Fine Wines.

**TOP 10 BOTTOM 10**



1. **Sold to the highest number of unique customers**: Additionally, we believe it is meaningful to look up products with the highest number of unique customers, meaning these products are more likely to be essential for a wide group of customers. These products are Rice, Fine Wafers, Mineral Waters, Scanned Vegetables and Fresh Poultry Meat.

**TOP 10 BOTTOM 10**

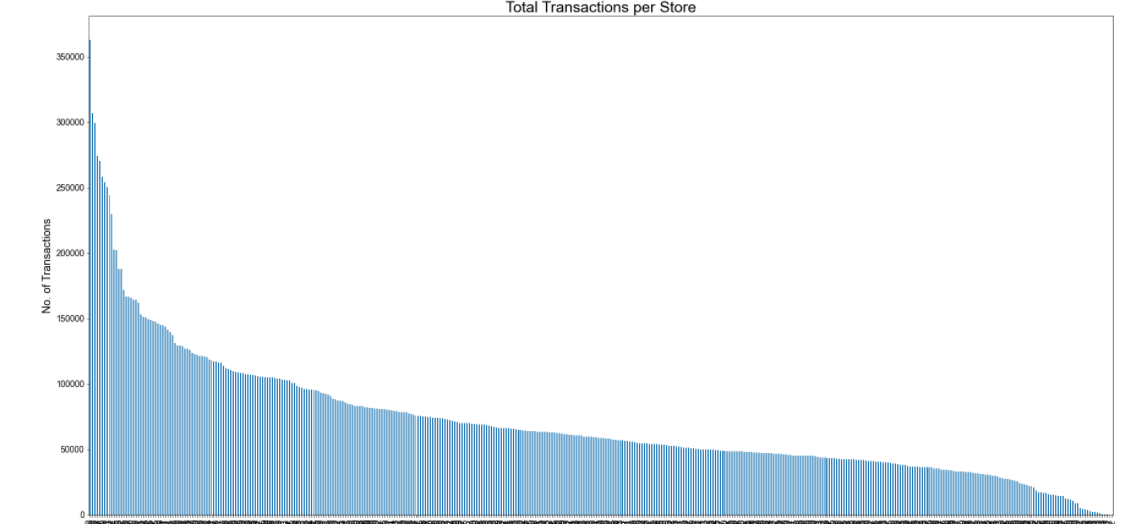


**Stores**

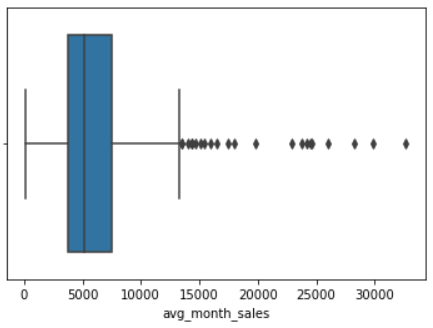
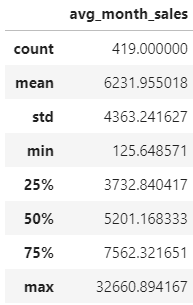
Pernalonga operates 421 stores that generated a total of $62,034,704.77 in revenue. However, the stores have an unequal distribution of total revenue. After observing the performance among stores, we noticed that two locations (store id 302 and 102) have extremely low total revenue and transactions. These observations were removed.

We also noticed that the stores, unlike products, don’t follow the 80/20 rule. The top 20% of the stores generate approximately 40% of the total revenue.

1. **Revenue**: By analyzing data on stores, we discovered that there is a general growth in revenue for 2017 compared to 2016. Moreover, we found the number of transactions or sales amount follow the long-tail phenomenon well, where most sales are done by the top stores.



From the box charts, we recognized that the monthly sales volume varies significantly across stores. Sales can go from $125 all the way to $32660. Almost all the stores’ sales volume are below $14000, but there are few stores having great leads in sales compared to others.



To find top stores within Pernalonga, we examined their sales volume, revenue, transaction, and customers.

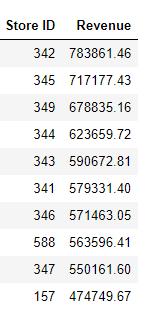
**Stores by highest volumes (CT):**

**TOP 10 TOP 10**



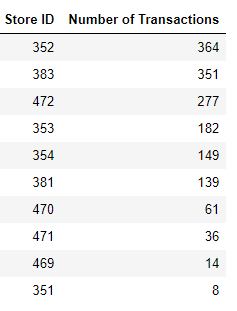
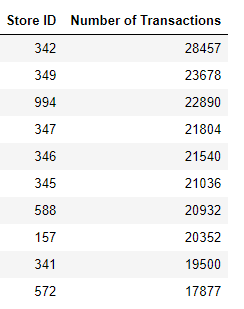
**Stores by total revenue:**

**TOP 10 BOTTOM 10**

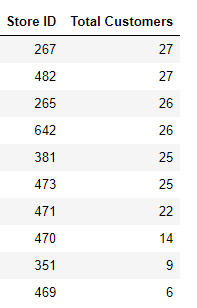
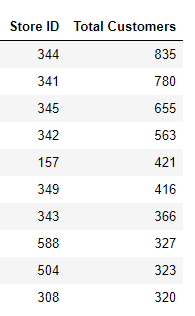


**Stores by transactions:**

**TOP 10 BOTTOM 10**

****

**TOP 10 BOTTOM 10**

****

**Segmentation & Insights**

In order to better segment customers and deliver meaningful insights to provide personalized promotion recommendations, we had to transform and normalize our data into more meaningful features. For example, we created columns for week and month, normalized historical total sales to monthly average sales and historical total visits/transactions to monthly average visits/transactions. Moreover, we calculated discounts versus sales ratio, discount frequency ratio, and more. By doing these, we were able to focus more on the current profitable status of Pernalonga’s customers/stores/products and consequently enhance their profitability.

**Customers**

After performing K-Means clustering, we decided to divide our customers into three groups: regular customer, loyal customer and cherry pickers. Each group is described in the table below.

**Regular customers**: those who spend an average amount, do not care particularly about products on sale, visit stores at an average rate and buy a reasonable amount of products.

**Loyal customers:** spend the most in stores, do not care as much about products on sale but visit stores at a high frequency and buy a large number of products.

**Cherry Pickers:** spend the least amount at the stores but have the highest probability to buy and highest frequency to buy during discounts; they buy an average amount of products.

|  |  |  |  |
| --- | --- | --- | --- |
| **Features/Clusters** | **Regular Customer** | **Loyal Customer** | **Cherry Picker** |
| Total Spending($) | Average  (7120.8) | High  (10015.5) | Low  (6893.06) |
| Probability to buy during discounts | Low  (0.12) | Average  (0.15) | High  (0.20) |
| Frequency to buy during discounts | Low  (0.26) | Average  (0.32) | High  (0.40) |
| Frequency visiting stores | Average  (0.46) | High  (0.53) | Low  (0.41) |
| Total number of products bought (Count) | Low  (815) | High  (1203) | Average  (983) |

\* Average = actual average, ex: Average ranking of Total Spending is the actual average of Total Spending

**Products**

Products are divided into three groups after K-Means segmentation: High Selling Essentials, Low Selling and “Luxury” Items.

**Low price:** low-priced products that have the highest number of buyers. These products are likely to be life essentials products such as water, rice and paper.

**Low selling (no discount):** items are not so popular amongst customers, not many people buy them; they are also least likely to be on sale.

**Highly discounted:** high-priced grocery products that have the highest average unit price but also have the highest discount rate.

|  |  |  |  |
| --- | --- | --- | --- |
| **Features/Clusters** | **Low Selling**  **(No Discount)** | **High Selling Essentials** | **Highly Discounted** |
| Average Monthly Customers (Count) | Low  (24.10) | High  (215.56) | Average  (54.85) |
| Average Discount Rate | Low  (0.053) | Average  (0.064) | High  (0.323) |
| Discount Frequency | Low  (0.184) | Average  (0.226) | High  (0.673) |
| Product Unit Price($) | Average  (4.947) | Low  (1.864) | High  (7.29) |
| Average Monthly Transactions  (Count) | Low  (11.93) | High  (27.97) | Average  (13.02) |

**Stores**

By performing K-Means clustering on store data, we were able to divide stores into four groups: flagship stores, high-performance stores,low performance / low discount stores, and low performance / high discount stores.

**Flagship stores:** they have the highest monthly sales, highest number of customers, highest product count and highest discount rate.

**High-Performance performance stores:** high monthly sales, high number of customers, high product count and low discount rate.

**Low performance / low discount stores:** lowest monthly sales, low number of customers, lowest product count and lowest discount rate.

**Low performance / high discount stores:** low monthly sales, lowest number of customers, low product count and high discount rate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features/Clusters** | **Low Performance/Low Discount Store** | **Low Performance/High Discount Store** | **High-Performance Store** | **“Flagship” Store** |
| Average Monthly Revenue ($) | Lowest  (4186.53) | Low  (4450.97) | High  (9847.38) | Highest  (24957.48) |
| Average Monthly Customer  (Count) | Low  (32.19) | Lowest  (28.23) | High  (61.55) | Highest  (172.75) |
| Average Monthly Product  (Count) | Lowest  (902.92) | Low  (965.70) | High  (1689.77) | Highest  (2948.96) |
| Average Discount Rate | Lowest  (0.136) | High  (0.171) | Low  (0.159) | Highest  (0.183) |
| Discount Frequency | Lowest  (0.285) | High  (0.342) | Low  (0.316) | Highest  (0.360) |

**Recommendations:**

Based on our analysis, we can now move on to suggesting and implementing targeted marketing tactics which would previously have been ineffective if deployed without proper analysis. Here is our suggestions to Pernalonga:

* Implement targeted marketing tactics that would’ve been inefficient 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. For example, we can offer promotions to selected groups such as cherry pickers where discounts are highly related to sales.
* Target consumer segments that are high spenders. For example, we can target the loyal group with some loyalty events to enhance the attachment of the group to Pernalonga.
* Rearrange store layouts to promote products with lower sales volume and to highlight products that bring the most revenue. For example, we can allocate high-selling essential goods along with high discounted products so when consumers are looking for high-selling goods they will acknowledge the recent promotions.
* Increase the number of products and promotions at Low-performance low discount stores to match up with low-performance high discount stores, since low-performance high discount stores are doing better in terms of sales.
* 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.