

ISSS603-G1-CustX

Aveda Project Final Report

* group 1

ZHU FANGYUAN

CHEN YIMAN

WIDYA TANTIYA YUTIKA

LAM YU HAY GLADWIN

KYAW ZAW LWIN HTUT OO LWIN

Contents

1. Introduction 1

2. Problem Statement 1

3. Approach 1

3.1 Increasing Recency 1

3.2 Increasing Frequency 1

3.3 Increasing Average Basket Size 1

3.4 Alternative Methods to Increase Sales 2

4. Data Dictionary 2

5. Data Preparation Procedures 2

6. Exploratory Data Analysis 2

6.1 Overall Order Performance 2

6.2 Overall Sales Performance 3

6.3 Overall Product Performance 4

7. Customer Segmentation 4

7.1 Segmenting Procedure 4

7.2 Results 5

7.2.1 Retail channel 5

7.2.2 3rd Party Marketplace Channel 5

7.2.3 Ecommerce Channel 6

7.2.4 Corporate Channel 6

7.3 Prioritizing Segments to Focus 6

7.4 Personas of Retail Potential Loyalist 6

7.5 Personas of Retail Need Attention 7

8. Market Basket Analysis 7

8.1 Apriori algorithm 7

8.2 Market Basket Analysis Results: Association Rules 8

8.3 Bundle Recommendations 8

9. Recommendation System 10

9.1 First Item Recommendation 10

9.1.1 For New Customer 10

9.1.2 For Existing Customer 11

9.1.2.1 User – Item Collaborative Filtering 11

9.1.2.2 Replenishment on latest purchase 12

9.1.2.3 Complementary Item in the same collection 12

9.2 Next Item Recommendation 12

10. Discount Analysis 12

10.1 Overview of Discount: How Much Discount Aveda Gave Out? 13

10.2 Analyzing the impact of Discounts on Customer Purchase Frequency 14

10.2.1 Low Spenders 14

10.2.2 Moderate Spenders 15

10.2.3 High Spenders 15

10.3 Analyzing the relationship between discounts and Average Basket Size (ABS) 15

11. Sample and GWP (Gifts with purchase) Analysis 16

11.1 Overview of Sample and GWP 16

11.2 Analyzing Purchase Frequency in Relation to Sample and GWP Distribution 16

11.2.1 Analyzing Purchase Frequency in Relation to Sample Distribution 17

11.2.2 Analyzing Purchase Frequency in Relation to GWP Distribution 17

11.2.3 Purchase Frequency Before vs. After receiving Samples 18

11.2.4 Customers who are sensitive to Samples 19

11.2.5 Recommended Samples 19

12. Conclusion 20

13. Future Work 20

14. Reference 21

# Introduction

Aveda is a haircare, skincare and wellness brand founded in 1978 by Horst Rechelbacher when he moved from Austria to United States. Aveda differentiates itself by applying more use of high quality organic and plant-based botanical ingredients compared to other brands in the same industry. Aveda advocates for natural beauty while meeting sustainability. This is done so via the sourcing of their ingredients for their products and their packaging option. Aveda is a high-end brand and has good presence in United States.

LUXASIA is a distributor of luxury brands in Asia. LUXASIA is the distributor of Aveda in Asia. Prior to Covid-19, LUXASIA saw high spend per transaction, high purchase frequency and short return purchase internal. However, after Covid-19, the situation has changed. Customers are spending less per transaction, have lower purchase frequency and have higher return purchase interval.

# Problem Statement

LUXASIA is concerned about the sales performance of Aveda and would like to revive the sales performance. LUXASIA would like to apply RFM model to increase customer’s repeat purchases. Next, they would like to reduce the repeat purchase intervals of existing customers. Lastly, they would like to increase the sales amount of each basket using cross-selling tactics.

# Approach

Although RFM segmentation is a customer segmentation technique, the variables that make up this segmentation (Recency, Frequency, Monetary value) can impact the sales. Sales or revenue is defined as quantity x selling price. Higher recency can bring higher sales. This is because higher recency refers to how recently a customer has placed an order. A recent purchase brings revenue, hence leads to higher sales. Higher frequency can bring higher sales. This is because higher purchase frequency impacts the quantity purchased over a period. Hence it leads to higher sales. Lastly higher sales amount per basket can lead to higher sales too.

## Increasing Recency

Promotional price can be used to increase recency. The attractiveness of the promotional price can disrupt buyer’s typical purchase timeline and buying behavior by making them make purchases otherwise they would not have. Another way to would be recommending complementary products to the customer. By recommending complementary products, it provides additional customer value proposition hence customer may buy additional products when compared to no product recommendation. This can be done via a product recommender.

## Increasing Frequency

Product recommendation can be used to increase customer purchase frequency as it offers customer new value propositions and use cases of other complementary products. Another way to increase customer frequency is to offer samples to customers. Samples allow customers to try out complementary products at a fraction of the cost of giving the standard size product away. Hopefully with the sample, customers will return thus increasing frequency.

## Increasing Average Basket Size

One way to increase average basket size. By increasing the basket size, it encourages users to buy more quantity or buy higher priced items. Another way is to upsell customers with bundle sets.

## Alternative Methods to Increase Sales

Other methods Aveda can consider increasing sales is via a subscription programme. The subscription programme can guarantee a cap on customer’s recency. It can guarantee minimum frequency and basket size thus providing a more certain revenue cashflow to Aveda. Aveda can also consider increasing their sales distribution channel via affiliate marketing. Aveda can tap on influencer’s connection and pay commissions to these influencers for successful transaction.

In summary, there are various tools to help Aveda increase sales. The tools are a product recommender, promotional price, sample offerings and creation of more and new bundling sets. For alternative tools, they can consider Aveda subscription and Aveda affiliate programme.

# Data Dictionary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Data Type** | **Description** | **Variable** | **Data Type** | **Description** |
| Order no | Number | Transaction order number | Description | Character | Description of the item |
| Customer | Number | Customer-id | Category Description | Character | Category of the item (Haircare, Skincare, etc.) |
| Country | Character | Country of purchase | Sales Channel | Character | Retail, 3rd Party Marketplaces, Corporate Sales, or eCommerce |
| Brand Description | Character | Brand of purchase | Discount | Number | Discount applied to the transaction |
| Sales Transaction Type | Character | Type of transaction. S refers to sale and R refers to refund | GST | Number | GST amount of the item |
| Document Date | Date | Date of the transaction | Item Discount | Number | Discount applied to the item in the transaction |
| Store ID | Number | Store where the transaction took place | Net Price | Number | Price of the product before GST and discounts |
| Sales Employee ID | Number | The employee who processed the transaction | Retail Price | Number | Price of the products after GST and discounts |
| Article Code | Character | SKU of the item | Qty | Number | Number of items |

# Data Preparation Procedures

The first step of data preparation is to transform the columns “Measure Names” and “Measure Values” from current format into wide format. This transformation consolidates the 6 rows of each transaction into a single row where “Measure Names” serve as column names. In total there are 170,126 rows. Next step of data preparation is to clean the data including removing duplicated rows if exists, finding missing value and replace accordingly and reformatting of the data. For our dataset, there is no duplicated row and some missing value for “Sales Employee Id”, which will not be used for further analysis, so no imputation is required. The cleaned dataset consists of 170,126 rows, with data from August 2021 until July 2023.

# Exploratory Data Analysis

Through this exploratory data analysis, we will delve into various aspects of the dataset, starting with the summaries and descriptive statistics to gain the overall view of the dataset provided. We will also include data visualization to uncover current trends and find valuable insights to determine our analysis focuses.

## Overall Order Performance

A blue and white sign with white text

Description automatically generated A blue and white rectangular sign with black text

Description automatically generated A blue and white sign with black text

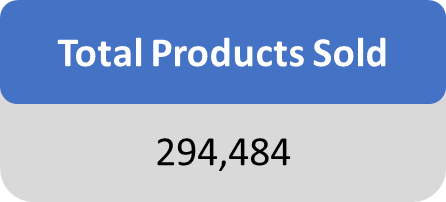
Description automatically generated

In the provided dataset, there are 170,126 number of transactions which include both sales and refund orders. In total, there are around 49,714 sales orders and 737 refund orders.

## Overall Sales Performance

|  |  |
| --- | --- |
|  | In total, the revenue for the 2 years period of data provided is S$12,159,003. In terms of seasonality, November has the highest orders. This could be due to customers doing most of their Christmas or 11.11 sales shopping during that period of time. |
| Figure 1. Monthly Revenue Distribution    Figure 2. Sales Channel-wise Revenue Trends | The revenue made by Retail is the highest compared to the other sales channels. Since AVEDA is a haircare premium brand, people are more comfortable with buying from the actual official physical stores themselves or their trustworthy eCommerce counterparts as opposed to 3rd party sellers (i.e. Shopee, Lazada, etc.) or corporate sales. |
| Figure 3. Store-wise Revenue Trends | Store ID “11” made the most revenue with $6,706,971 which is significantly much more than the other stores ID.  This is due to store ID 11 comprising of all sales channel besides retail. Thus, the actual store with the highest revenue is store “5002”, followed by “6001” and “5009”. |

## Overall Product Performance

  A blue and white sign with black text

Description automatically generated

|  |  |  |
| --- | --- | --- |
| **Sales Channel** | **Average Basket Size (by Quantity)** | **Average Basket Size (by Dollar)** |
| **3rd Party Marketplaces** | 4 | 164 |
| **Corp Sales** | 3 | 121 |
| **Retail** | 4 | 165 |
| **eCommerce Stores** | 197 | 8794 |

Over the course of two years, a total of 294,484 products were sold. The average basket size, when measured by quantity, stands at 6 items per order. In terms of dollars, the average basket size is $247.78 per order.

It is worth noting that the average basket size varies significantly between different sales channels. E-commerce exhibits substantially larger average basket sizes, both in quantity and dollar value per order. Despite the impressive average basket size in e-commerce, retail channels still contribute more to the overall revenue. This is because the number of orders placed through e-commerce is comparatively lower.

# Customer Segmentation

A close-up of a blue and white card

Description automatically generated

We used the RFM customer segmentation to segment the customers. The RFM model examines the recency, frequency and monetary value at customer level and assigns a score to customer for each of the 3 variables.

## Segmenting Procedure

A blue squares with white text

Description automatically generated

There are a total of 4 sales channels – retail, ecommerce, 3rd party marketplace and corporate sales. As the nature of the 4 sales channels each have a very different customer profile, we further segmented the customer into their sales channel to analyze them individually. Next, we computed the recency, frequency, monetary value metrics at customer level to prepare it for RFM segmentation scoring.

The data provided is a snapshot of a customer’s past transaction data between the time windows. We could not tell whether the customer had a prior transaction outside of the time window. We assumed that the first transaction we saw in the data was the customer’s first transaction. We also realized the customers started their relationship with Aveda at a different timing. Users who joined earlier are likely to have higher frequency and monetary value than users who started their relationship later. To ensure fairness between the customers, we used relative recency, relative frequency and relative monetary value instead. The relative metric is computed as the metric over the customer’s lifetime relationship. The customer lifetime relationship is computed as the time difference between the date of data pull and customer’s first order date in the data. See formula below:

A blue squares with white text

Description automatically generated

The RFM scoring was then done using the relative metrics.

## Results

We will dive into the RFM segments based on different sales channel as follows.

### Retail channel

A table with numbers and letters

Description automatically generated

The biggest group of customers belongs to the Churned segment. Churning and churned customers make up about 50% of retail customers. The median spend of churning and churned buyers is lowest. The champion segment is the most valuable as it has the lowest recency, highest frequency, highest GMV spent despite their young lifetime customer relationship.

### 3rd Party Marketplace Channel

A table with numbers and text

Description automatically generated

The biggest group of customers belongs to Churned. They make up 50% of all 3rd party marketplace customers. They have a median frequency of 1 purchase with a low median spent of $139.1. The second biggest group here belongs to Churning. They also have a median frequency of 1 purchase with low median spent of $197. However, they have shorter recency compared to Churned group. Together, the 2 groups occupy 78.7% of all the customers in this sales channel. On the contrary, the Champions segment is the most valuable segment. They have lowest recency of 53 days in average. They have high median frequency of 4 and very high median spent at $800.90.

### Ecommerce Channel

A screenshot of a graph

Description automatically generated

The biggest group of customers belongs to Churned. They make up 76% of all ecommerce customers. They have a median frequency of 1 purchase with a low median spent of $139.10. The second biggest group here belongs to Churning. They also have a median frequency of 1 purchase with higher median spent of $222.80. They have much shorter recency compared to Churned group at 241 days. Together, the 2 groups occupy 92.8% of all the customers in this sales channel. Champions segment is the most valuable segment. They have lowest recency of 227.5 days in average. They have high median frequency of 3 and very high median spent at $857.70.

### Corporate Channel

A screenshot of a graph

Description automatically generated

The biggest group of customers belong to early life high potential. They make up 26.6% of all corporate customers. They have a median frequency of 1 purchase with a relatively decent median spent of $170.10. The second biggest group here belongs to Churned. They also have a median frequency of 1 purchase, but much lower median spent of $46.7. Mature loyal most profitable (ie Champions) segment is the most valuable segment with 5 purchases in average and high median spent of $629.60.

## Prioritizing Segments to Focus

Due to limited time and manpower, Aveda can focus on increasing sales from Retail channels. They can focus targeting the Potential Loyalist and Needs Attention segments. This is because these segments are a good size, bring in a big portion of Aveda sales and will be the fastest way to see any GMV uplift, if any.

## Personas of Retail Potential Loyalist

We mapped in the needs of the users in this segment via the product collections they have previously bought. We then performed K means clustering, to further split them into smaller segments. We used the elbow plot in deciding the optimal number of K clusters. We end up with below personas:

|  |  |
| --- | --- |
| A graph with a line  Description automatically generated | A screenshot of a computer  Description automatically generated |

|  |  |  |
| --- | --- | --- |
| **The desperate “Hair Thinning” middle aged man** | **The very concerned “Damaged Hair” lady** | **The mildly “Damaged Hair” first time buyer** |
| * 1 past purchase * Last purchase about 3 months * Spent $288 * Hair Care Needs: Experiencing hair loss | * 3 past purchases * Last purchase about 6 months * Spent $517 * Hair Care Needs: Damaged hair and dry hair | * 1 past purchases * Last purchase about 2 months * Spent $197 * Hair Care Needs: Damaged hair and wants a healthy scalp |

## Personas of Retail Need Attention

Similarly, we mapped in needs of users in the segment based on the product collections they bought. We then split them using K means and using the elbow plot to determine the optimal number of clusters. We end up with below personas:

|  |  |
| --- | --- |
| A graph with a line  Description automatically generated | A screenshot of a computer  Description automatically generated |

|  |  |  |
| --- | --- | --- |
| **The impatient skeptical “Hair Thinning” man** | **The not too concerned “Damaged Hair” lady** | **The “Dry hair” but not too concerned, price sensitive bulk buyer** |
| * 1 past purchase * Last purchase about 6.5 months * Spent $159 * Hair Care Needs: Experiencing hair loss | * 2 past purchases * Last purchase about 7.5 months * Spent $202 * Hair Care Needs: Damaged hair | * 2 past purchases * Last purchase about 11 months * Spent $352 * Hair Care Needs: Dry hair |

# Market Basket Analysis

Market Basket Analysis (MBA) is a data mining and statistical technique used in retail and e-commerce to discover associations between products that are often purchased together by customers. The primary goal of MBA is to identify patterns and relationships in transaction data, such as customer purchase history and to determine which items tend to be bought in conjunction with others. This analysis can help retailers understand customer preferences, optimize pricing and promotion strategies, and increase revenue through cross-selling and upselling.

## Apriori algorithm

The Apriori algorithm is a widely used method for performing Market Basket Analysis and it works by generating association rules that identify frequent item sets in transaction data. The parameters used to generate frequent item sets are:

|  |  |
| --- | --- |
| **Support** | This is the minimum threshold for an itemset to be considered frequent. It is typically defined as the percentage of transactions that contain a particular itemset. |
| **Confidence** | This parameter determines the strength of the association between items in an association rule. It's calculated as the percentage of transactions containing the antecedent (left-hand side) item where the consequent (right-hand side) item is also present. |
| **Lift** | Measures how much more likely the consequent is bought when the antecedent is bought compared to when it's bought without considering the antecedent. Lift values greater than 1 indicate a positive association. |
| **Minimum Items per Rule** | This parameter sets the minimum number of items in an association rule. |

The minimum support used was 0.01, meaning that an itemset must appear in at least 1% of the transactions to be included in the analysis. The minimum threshold for lift was set to 1.0 as we are only interested in items that are positively associated (frequently bought together in an order).

## Market Basket Analysis Results: Association Rules

Based on frequent item sets, we generate association rules and determine the optimal bundling of items. Order numbers & Article codes are parameters used for the Market Basket Analysis.

**Association Rules: Antecedents -> Consequents**

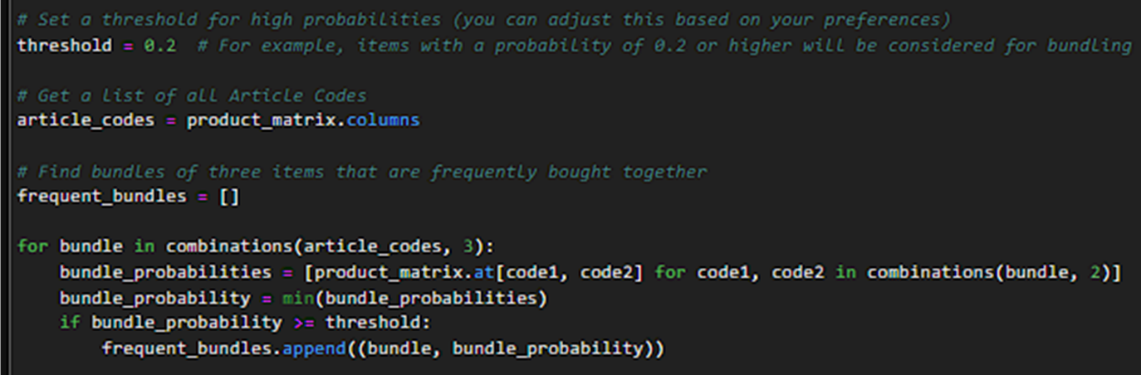
Below are the snippets of our top association rules:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **antecedents** | **consequents** | **antecedent support** | **consequent support** | **support** | **confidence** | **lift** |
| **frozenset({'A09A700000-1'})** | **frozenset({'A1XG010000'})** | **0.0604** | **0.0739** | **0.0125** | **0.206** | **2.791** |
| **frozenset({'A1XG010000'})** | **frozenset({'A09A700000-1'})** | **0.0739** | **0.0604** | **0.0125** | **0.169** | **2.791** |
| **frozenset({'AMFW010000'})** | **frozenset({'A09A700000-1'})** | **0.1000** | **0.0604** | **0.0115** | **0.115** | **1.902** |
| **frozenset({'A09A700000-1'})** | **frozenset({'AMFW010000'})** | **0.0604** | **0.1001** | **0.0115** | **0.190** | **1.902** |
| **frozenset({'A1XG010000'})** | **frozenset({'AEY9010003'})** | **0.0739** | **0.06696** | **0.0102** | **0.138** | **2.056** |

The antecedents (on the left-hand side of the rule) represent the items that are already in the customer's basket or transaction. While consequents (on the right-hand side of the rule) represent the items that are likely to be added to the customer's basket in addition to the antecedents. In other words, consequents are the items that the analysis suggests customers might buy in conjunction with the antecedents. The goal is to find associations or patterns between antecedent and consequents with high associations and so optimal bundles can be determined.

## Bundle Recommendations

Higher confidence and lift values are used to determine the significance of association rules. The minimum thresholds set can filter out rules that are strong and relevant. The rules can be used to bundle, and targeted marketing to encourage customers to purchase both the antecedents and consequents together, thereby increasing sales and customer satisfaction. We study which items go well together in a bundle. The probability threshold is set to 0.2 while the number of items per bundle was set to 3 as shown in the code below. The probability matrix shows the probability of the article code of a product being bought together with the article code of another product.



Product Probability Matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Article Code** | **A09A700000-1** | **A0K6010000** | **A0RF010000** | **A0TK010000** | **A139010000** |
| **A09A700000-1** | **1** | **0.0157** | **0.0372** | **0.0286** | **0** |
| **A0K6010000** | **0.00160** | **1** | **0.00429** | **0** | **0** |
| **A0RF010000** | **0.0139** | **0.0157** | **1** | **0** | **0** |
| **A0TK010000** | **0.000533** | **0** | **0** | **1** | **0** |
| **A139010000** | **0** | **0** | **0** | **0** | **1** |
| **A13F010000** | **0** | **0** | **0** | **0** | **0.0357** |
| **A14G010000** | **0.00320** | **0.00524** | **0.00429** | **0** | **0** |
| **A14H01J000** | **0.00533** | **0** | **0.00572** | **0** | **0** |
| **A14Y010000** | **0.0107** | **0** | **0.00715** | **0** | **0.0357** |

From the table above, A14Y010000 (AVH COLOR CONSERVE SHMP 1000ML) has the highest probability of being bought together with A09A700000-1(AVH PADDLE BRUSH) with a probability of 0.0107. Below are the recommended bundles (bundles of 2 and 3 products) and their probabilities of being bought together based on different sales channels:

* Retail Channel

|  |  |  |  |
| --- | --- | --- | --- |
| **Product 1** | **Product 2** | **Product 3** | **Probability** |
| AVH SHAMPURE HAND & BODY WASH 250ML | AVH BEAUTIFYING HAND RELIEF 40ML TUBE | AVH TLS FACIAL DRY BRUSH | 0.519 |
| AVH RM H/B WASH 250ML | AVH BEAUTIFYING HAND RELIEF 40ML TUBE | AVH TLS FACIAL DRY BRUSH | 0.429 |
| AVH SS EXFOLIATING SCALP TREATMENT 150ML | AVH SS OVERNIGHT SCALP RENEW SERUM 50ML | AVH SCALP SLTN STIMULATG SCALP MASSAGER | 0.394 |
| AVH SS EXFOLIATING SCALP TREATMENT 150ML | AVH SS BALANCING SHAMPOO 200ML | AVH SS REPLENISHING CONDITIONER 200ML | 0.393 |
| AVH LITER PUMP | AVH SS BALANCING SHAMPOO 1000ML | AVH SS REPLENISHING COND 1000ML | 0.333 |

|  |  |  |
| --- | --- | --- |
| **Product 1** | **Product 2** | **Probability** |
| AVH SS BALANCING SHAMPOO 200ML | AVH SS REPLENISHING CONDITIONER 200ML | 0.833 |
| AVH COLOR CTRL SHAMP 10ML/.34FLOZ | AVH COLOR CNTL COND 10ML/.34FLOZ | 0.832 |
| AVH 19 ROSEMARY M PURIFY SHAMPOO 250ML | AVH SS BALANCING SHAMP 1000ML | 0.789 |
| AVH 19 ROSEMARY M PURIFY SHAMPOO 250ML | AVH SS REPLENISHING COND 1000ML | 0.778 |
| AVH RM H/B WASH 250ML | AVH INVATI ADV LITRES | 0.681 |

* E-commerce Channel

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Product 1** | **Product 2** | | **Product 3** | | **Probability** |
| AVH INVATI ADV EXF SHAMP LIGHT 1000ML | AVH INVATI ADV EXF SHAMP RICH 200ML | | AVH BLONDE REVIVAL SHAMP 200ML | | 0.933 |
| **Product 1** | | **Product 2** | | **Probability** | |
| AVH INVATI ADV EXF SHAMP LIGHT 1000ML | | AVH INVATI ADV EXF SHAMP RICH 200ML | | 0.933 | |

It is noticed that the recommended bundles of products for e-commerce are slightly larger in volume compared to those of retail.

* Corporate Sales Channel

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Product 1** | **Product 2** | | **Product 3** | | **Probability** |
| AVH SMOOTH INFUSION SHMP 1000ML | AVH NUTRIPLENISH COND LT 1000ML | | AVH NUTRIPLENISH MASK LGT 150ML | | 0.333 |
| Product 1 | | Product 2 | | Probability | |
| AVH SMOOTH INFUSION SHMP 1000ML | | AVH NUTRIPLENISH COND LT 1000ML | | 0.333 | |

The products in the recommended bundles for Corp Sales are even bulkier than in e-commerce. The top recommended bundle for Corp Sale is usually 1000 ml.

* 3rd Party marketplace Channel

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Product 1** | **Product 2** | | **Product 3** | | **Probability** |
| AVH ORGANIC TOTE BAG | AVH INV REVITALIZER DUO (OL) | | AVH INV COND 1L + LE MINI PADDLE | | 0.704 |
| **Product 1** | | **Product 2** | | **Probability** | |
| AVH ORGANIC TOTE BAG | | AVH INV SP LT 1L + COND + REVIT 10ML X 2 | | 0.969 | |

For 3rd party marketplaces recommendation, the MBA recommended bundles to be bought together.

# Recommendation System

In this section, a recommender system was created to improve customer experience by catering to customers preferences. We designed 2 types of recommendation system. One is for the first item recommendation and another one for following item recommendation.

## First Item Recommendation

For the first item recommendation, there are 2 separate recommendation systems for new customers and existing customers.

### For New Customer

Depending on the sales channel, we will recommend best sellers for new customers with some rules such as only recommending products from haircare category, single item products and smaller size of these best sellers. Below are the top 3 best sellers for each sales channel.

A screenshot of a product list

Description automatically generated

### For Existing Customer

There are 3 distinct functions that are developed to tailor for existing customer preference based on the products they have previously purchased.

#### User – Item Collaborative Filtering

In this approach, recommendations are made by finding similar users to the target user and recommending items which those similar users have liked. In this dataset, there is no data about customer rating on the items whether they like the item or not, so we define new metric to measure the how much users like certain items which is “Repurchase Count”. Based on repurchase count, we use rating scale of 1 to 5, the higher the repurchase count, the higher the rating scale is which also means that the more they like the products. Since there are some items without being repurchased by certain customers, instead of assuming that they do not like the products by giving low rating scale, we leave the rating scale to be empty (sparse matrix). The metric to measure the similarity between users that we tested are:

* + **Cosine Similarity**. It is scale-invariant, meaning that it focused on the direction between vectors but not the by the magnitude of the rating scale. This can be advantageous when we want to emphasize the similarity in the directions of user-item interaction vectors rather than the absolute ratings. However, in this similarity measure, magnitude of ratings is ignored, and missing ratings are treated as negative.
  + **Pearson Correlation Coefficients.** It takes into account the centering of rating scales, which means that it considers the variation in the mean value of vectors. This can be advantageous when we want to capture both the direction and strength of the linear relationships between vectors. However, Pearson correlation may not work well with the sparse matrix as it relies on having significant number of common ratings between users/items.

Instead of using a pure cosine similarity metric, we use neighborhood-based approach (K-Nearest Neighbor Baseline) to help alleviate the problem of missing ratings affecting similarity calculations. Using the surprise library in Python, several models are tested using GridSearchCV to obtain the best parameters as follows:

* + K= 10
  + Similarity Matrix = Cosine
  + Min-Support = 10

For the above best model, the RMSE is 0.536, meaning that on average, the predicted ratings deviate from the true ratings by approximately 0.536 on a scale of 1-5. With this User-Item Collaborative Filtering function, customers without repeat purchase items will not get any recommendation, so to incorporate that, some additional functions are created.

#### Replenishment on latest purchase

In this approach, our aim is to suggest items that are likely to be running low based on the latest purchase items. Our research indicates that, assuming an average hair wash consumes 10ml of the product, and individuals are generally recommended to wash their hair every two days, we estimate that individuals use around 150ml of hair products in a month. To determine when it's time for replenishment, we compare this estimated monthly usage to the time elapsed since their latest purchase.

In addition, our analysis of the quantity sold distribution has led us to consider that the quantity of items sold can influence the timing of replenishment. Specifically, when the quantity sold is 10 or less, we classify it as a B2C (Business-to-Consumer) situation, and the time for replenishment will be affected by the quantity sold as well. While, for quantity sold surpassing 10, we categorize it as a B2B (Business-to-Business) scenario, and time for replenishment is assumed to only depend on the volume of the products.

With this Replenishment function, customers who recently shopped may not get any recommendation as it is not yet time for replenishment, so an additional function are created.

#### Complementary Item in the same collection

This is a higher-level recommendation in which based on the latest purchase items, we will suggest items in the same collection which are complementary with previously purchase items. For this method, we conduct market basket analysis based on the function of the haircare products (i.e. shampoo, conditioner, and etc.).

A close up of a text

Description automatically generatedA close up of a bottle

Description automatically generated

As shown by the figure above, since customers bought Aveda colour control conditioner, we will recommend complementary items in the same collection which are Aveda colour control leave-in treatment and shampoo.

The implementation of these three distinct functions will guarantee that our existing customers receive recommendations.

## Next Item Recommendation

The next item recommendation function recommends the following item a customer should consider based on the products they have already added to their cart. It checks the antecedents of the association rules generated through market basket analysis and if there is a match, it retrieves the consequents as the recommendation.

# Discount Analysis

Discount analysis is an analysis on how discounts impact demand for a good or service. It can be used to understand customer behavior and to optimize sales and product offerings. When a company offers a discount, it is typically trying to stimulate demand for its product. If the discount is large enough, it may cause consumers to purchase more than when there is no discount offered.

A graph showing the value of a curve

Description automatically generated with medium confidence

Figure 4. Demand Curve

From the above demand curve, the quantity demanded is lower when the price of product is higher. From here, we can infer that boosting demand may be possible if we reduce the price by offering discount. We also study if one discount has similar impact on different customer segmentations’ demand.

In our analysis, we demonstrate ‘demand’ from 2 angles: frequency of every customer and Average Basket Size per person. We will explore how these 2 features are affected by different discounts in the section 10.2 and 10.3.

***Total demand = total number of customers \* Average Basket Size per Customer***

***Total demand = total number of customers \* Average spending per order \* Frequency***

## Overview of Discount: How Much Discount Aveda Gave Out?

A graph with a number of numbers

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated

From the distribution of discounts, we notice that most customers benefit from a 10% discount, followed by 40% discount. To assess the effectiveness of two discount levels at boosting customer visit frequency,10% and 40% off, we applied ANOVA (Analysis of Variance) test.

A number and text on a white background

Description automatically generated A black and white text

Description automatically generated with medium confidence

From the ANOVA test result, the difference in visit frequency between the 40% off and 10% off groups is 9.747 with lower bound of 8.57 and 10.9. This indicates that on average, customers who received a 40% discount visited more frequently than those with a 10% discount. From this result, we can infer that offering a 40% discount is more effective at boosting customer visit frequency as compared to a 10% discount and so Aveda may want to explore this and assess the implications of adjusting discount strategies to leverage the higher discount level for increased customer engagement and visit frequency for certain items, rather than giving a uniform 10% discount to many products.

## Analyzing the impact of Discounts on Customer Purchase Frequency

A graph of a number of objects

Description automatically generated with medium confidenceA screen shot of a graph

Description automatically generated

Figure 4. Demand Curve

On the scatter plot, every dot represents a distinct customer, with a unique combination of a consuming frequency and average price per product. The average price of each product purchased by a customer is a key metric in customer analysis that can provide valuable insights into their preferences and consuming power.The price on X-axis is used to categorize customers into three spending categories based on the value of their spending per product.

### Low Spenders

Customers with an average price per product below 100 are categorized as "Low Spenders". For low spenders, we segmented again by purchase frequency, a noteworthy trend emerges. Specifically, for customers with a frequency of below 20, we notice that as the price per product increases, their purchase frequency also rises, indicating a willingness to invest in higher-priced items. However, this increase in frequency reaches a peak at a certain price point, after which their purchase frequency begins to decrease. This suggests that there is an optimal price range that maximizes their buying frequency.

From the result, we determine the specific price range at which low spenders reach their peak purchase frequency which ranged from $20 to $40. Aveda may consider giving discounts on popular items to fall within range or to make a smaller size item to cater for these customers.

As for low spenders with frequency of above 20, we notice that the purchase frequency decreases as the price per product increases. This segment exhibits random purchases, and the number of customers in this category is very limited. Thus, it is worth delving deeper into their behaviors to understand their unique characteristics and preferences. Giving product diversification within the optimal price range may benefit this segment. In addition, Aveda can offer cross-selling products or bundling to encourage complementary products, but to keep in mind the optimal price range.

### Moderate Spenders

Customers with an average price per product between 100 and 200 are categorized as "Moderate Spenders.

For moderate spenders, a noteworthy trend is that as the price per product increases, their purchase frequency tends to decrease. This indicates that providing more substantial discounts may actually lead to reduced purchase frequency among this customer segment.

### High Spenders

Customers with an average price per product larger than 200 are categorized as "Moderate Spenders.  While high spenders possess considerable purchasing power and exhibit a preference for expensive products, they are small and infrequent customer segments. For these customers, we can recommend bundles of higher priced items.

In conclusion for Discount analysis, discount offers should be tailored by different customers segment especially for customers who responds well with discounts, but to keep in mind the discount range (sweet spot for customers). Aveda can also consider product assortment to ensure that it complements discount strategy. High-quality and popular products can be paired with discounts to maximize the impact on spending. Last but not least, maintaining a balance between discounting and profit margins is essential.

## Analyzing the relationship between discounts and Average Basket Size (ABS)

A screen shot of a graph

Description automatically generated

Figure 5. Scatterplot for Total Expenditure and Discount

The scatter plot indicates that the total net expenditure by customers reaches its peak at a **50% discount**, suggesting that, on average 50% discount is the most effective way to maximize customer spending.

* **Excessive Discounts:** Customers on the right side of the peak point, who received excessively high discounts, did not result in substantial growth in sales or total net price. This indicates that offering extremely deep discounts may not always yield a proportional increase in spending, and businesses should be cautious about giving away too much margin.
* **Positive Correlation**: On the left side of the peak point, there is a noticeable positive correlation between the total net price and the discount. This suggests that customers are more likely to spend more when they receive a discount, even though this correlation may diminish as the discount level increases beyond the optimal point.

Based on the scatter plot data, AVEDA may consider adjusting discount strategy to focus on giving discount below than 50% as excessive discounts may not yield a significant increase in spending. While discounts are essential for attracting customers, they should not erode profitability.

# Sample and GWP (Gifts with purchase) Analysis

Samples and GWP (Gifts with purchase) play a crucial role in advertising new products, introducing fresh products to the clients, improving customer loyalty, and offering an opportunity to exchange essential information during the process of giving out samples and GWP.

## Overview of Sample and GWP

*A graph of orange lines

Description automatically generated*

Figure 6. Timeline of Sample and GWP Distribution

In total, there are 503 product variations from the dataset provided and among which 40 products are samples and 26 are GWP from Haircare Category. The average number of samples and GWP given out per day has been growing since the beginning of 2023 as shown in the line chart above. Therefore, it is important to get some insights from our analysis so that the company can give out samples and GWP with a precise target and direction. In our analysis subsequently, we have excluded new customers who made their first purchase in the latest month.

## Analyzing Purchase Frequency in Relation to Sample and GWP Distribution

In this section, we want to know whether giving out samples and GWP will improve the purchase frequency. First, we analyzed the relationship between customers’ Purchase Frequency and whether they have ever received sample/GWP. We divided customers who have ever received samples as group “Y”, and those who have not as group “N” for both sample and GWP. We used the ANOVA test to compare differences between the 2 groups upon sample and GWP respectively.

### Analyzing Purchase Frequency in Relation to Sample Distribution

A graph of a number of points

Description automatically generated with medium confidenceNull Hypothesis*: There is no difference between the means of “Y” and “N” groups.*

From the test result, the adjusted p-value is near 0, which means that we can reject the Hypothesis, and the difference in average Purchase Frequency between the group “Y” and “N” is 7 times.

diff lwr upr p adj

Y-N 6.738243 6.456744 7.019742 0

Recieved\_SMP count mean sd

*<chr>* *<int>* *<dbl>* *<dbl>*

1 N 7924 3.73 4.36

2 Y 7732 10.5 12.0

In addition, the average Purchase Frequency of customers who have ever received samples is 11 times, while that of customers who have not is 4 times.

### Analyzing Purchase Frequency in Relation to GWP Distribution

Null Hypothesis*: There is no difference between the means of “Y” and “N” groups.*A graph of a number of lines

Description automatically generated with medium confidence

From the test results, the adjusted p-value is also near 0. The difference in average Purchase Frequency between the group “Y” and “N” is 8 times.

diff lwr upr p adj

Y-N 8.012912 7.721211 8.304613 0

`Received \_GWP` count mean sd

*<chr>* *<int>* *<dbl>* *<dbl>*

1 N 10351 4.35 4.88

2 Y 5305 12.4 13.5

The average Purchase Frequency of customers who have received GWP is 13 times, while that of customers who have not is 5 times.

Based on the test results, we can conclude that there exist relationships between Sample/GWP and Purchase Frequency. Customers who have received Sample/GWP do come back more frequently than others. Furthermore, Spearman Test results show a positive association between Sample/GWP and Purchase Frequency as shown below.

A white text with black text

Description automatically generated

The p-value is near 0, meaning that the association relationship is statistically significant. And from the Coefficient value, we can see that x2 (=GWP) contributes more to the Purchase Frequency. Although we concluded that GWP is more associated with Purchase Frequency than Samples, giving out too many GWP products is unsustainable due to GWP’s higher cost and function of making customers feel special. So, the company should still pay major attention on samples, and thus we further analyze on Samples.

### Purchase Frequency Before vs. After receiving Samples

It is not enough to conclude that receiving samples/GWP results in higher Purchase Frequency. In order to find the causality between Purchase Frequency and Receiving Samples, we need to consider following questions:

* *Are there any changes about Purchase Frequency before and after customers received samples?*
* *How to cut the time point of before and after in concern that the time within the dataset is not a full life cycle?*
* *How to evaluate the effectiveness?*

But in our analysis, we simplify the question as to compare the Purchase Frequency changes before and after the **First time** the customers receiving samples within the time of the dataset.

We filtered out the transactions of customers who have received samples before and found out the *Document date* of their purchase when they received samples. Then for each customer, we calculated their Monthly Purchase Frequency before and after that date, where

=

=

**A table with numbers and a few digits

Description automatically generated**

We generated a table showing the Monthly Purchase Frequency for customers as shown above.

**A screenshot of a computer screen

Description automatically generated**

Figure 7. Monthly Purchase Frequency BEFORE and AFTER giving samples.

The box plot above shows the average Monthly Purchase Frequency BEFORE and AFTER giving samples indicating that the Purchase Frequency of customers after the first-time receiving samples sharply dropped down. But we cannot tell that Product Samples do not affect the Purchase Frequency as we do not know if the first date of receiving samples in the dataset is the real first date. But at least we can filter out customers who are sensitive to samples based on current analysis.

### Customers who are sensitive to Samples

Based on the above conclusion, we can safely filter out **295** customers whose purchase behaviors on frequency did alter after receiving samples. Therefore, we can aim to give out more effective samples to increase their purchase frequency.

Apart from changes in customers’ purchase frequency, what products customers bought after receiving samples also demonstrates their sensitivity to samples. We usually hope that samples can bring customers back to buy the same products, if so, it means that the customers care about the samples, and on the other hand, these samples are effective. However, by comparing the transaction date, we found that there was no customer who bought the product for the first time after receiving product samples, but there were some who bought products from the same collections as shown in the table below.

A close up of a document

Description automatically generated

There are **5127** suchcustomers who care about the product samples. We can target them as potential customers of cross-selling strategies as they are willing to accept advertisements for new products.

### Recommended Samples

Extending from the last step, we can also filter out those samples that do bring customers back to buy products from the same collection as the samples. There are **33** out of 40 items in total that can be safely deemed as effective. Grounded on the effective sample list as shown partially below (sorted from largest to smallest in terms of sales generated), the employees can pick ones when giving out samples for different purposes while balancing the costs of samples with the business needs.

A screenshot of a computer

Description automatically generated

When giving away samples, AVEDA can focus more on the sample-sensitive customers listed in the previous section to increase the frequency of their purchases. In addition, customers who have purchased products from the same collection as the samples received are prime candidates for cross-selling. On the other hand, employees can provide more effective samples as partially listed in the above table when promoting new products or new collections to customers.

# Conclusion

Aveda can leverage on several tools to help increase their customer recency, purchase frequency and total to date sales amount. These tools are segmentation, product recommender, market basket analysis, discounting and offering appropriate samples. To further finetune, they should do A/B testing to see how sensitive the sales and customers are to these tools while planning their sales strategy.

# Future Work

There are several avenues for future work that will contribute to the development and improvements of the projects as follows.

* Explore Subscription Model

Aveda Subscription model can be used. The features of the subscription allow users to try out different product line. If it is unsuitable, it can be exchanged free of charge to a different collection. It will also allow users to cancel any time. Users under this subscription can schedule in advance when they need to top up their products and a free delivery to their doorstep will be made to send them the items.

* Explore Additional Distribution Channel

Aveda can consider an affiliate programme. They can tap on influencer’s fame to market Aveda products. For every successful transaction from influencer’s referral, Aveda can do some profit sharing with them to keep them motivated in marketing for Aveda.

* Try More Personalised Marketing

Aveda should explore more personalized marketing strategies by developing personalized marketing campaigns and promotions for each customer segments, focusing on their specific needs and preferences to enhance customer loyalty, encourage repeat purchases, and increase the probability for closing out each sale.

* Refresh RFM Segments with Latest Sales Data

To enhance the average order value in the future, Aveda should dive deeper into customer profiles, preferences, and purchase histories to identify opportunities for cross-selling. They can start with profiling via RFM segmentation first as the first level before going further into segmentation. However, all RFM models will be obsolete with time hence Aveda should continuously update it.

# Reference

Barilliance. (n.d.). RFM Analysis: Understanding Your Customers Better. Retrieved from <https://www.barilliance.com/rfm-analysis/#:~:text=is%20RFM%20Analysis%3F-,A%20definition%20and%20context.,much%20they%27ve%20spent%20overall>.

Kursun, I. (n.d.). Customer Segmentation with RFM Analysis: Learn More About Your Customers. Medium. Retrieved from <https://medium.com/@ilaydakursun/customer-segmentation-with-rfm-analysis-learn-more-about-your-customers-f12b348acbd>.

Kadlaskar, A. (2023) Market basket analysis: A comprehensive guide for businesses, Analytics Vidhya. Retrieved from <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis.>

Baldha, S. (2022) Introduction to collaborative filtering, Analytics Vidhya. Retrieved from <https://www.analyticsvidhya.com/blog/2022/02/introduction-to-collaborative-filtering/>

Schildge, G. (2022, July 9). What is a Product Life Cycle Analysis and Management. Matrix Business AI. <https://matrixmarketinggroup.com/product-life-cycle-analysis/>

Dopson, E. (2021, November 22). Product Sampling: 6 Tips and Best Practices to Delight Customers Today. Shopify. <https://www.shopify.com/sg/retail/product-sampling>

Pinsker, J. (2014, October 1). The Psychology Behind Costco's Free Samples. The Atlantic. <https://www.theatlantic.com/business/archive/2014/10/the-psychology-behind-costcos-free-samples/380969/>