

MARKET BASKET ANALYSIS REPORT

UNSUPERVISED

2024

MARKET BASKET ANALYSIS AND PRODUCT BUNDLING REPORT

1. INTRODUCTION

This report details the market basket analysis and product bundling process, aiming to identify frequently co-purchased items and suggest bundles that could drive sales growth.

2. DATA LOADING AND PREPARATION

- Data Collection:
 - Datasets `order_products.csv`, `products.csv`, and `orders.csv` are utilized to analyze customer purchasing behavior.
- Data Cleaning:
 - The `eval_set` was removed because it contains prior data, including both training and testing sets, with most of the data being from the prior period, which is not relevant for this analysis.
- Product Frequency Calculation:
 - Each product's order frequency is calculated, and the top 100 most frequently ordered products are identified.
- Creation of a Frequent Products List:
 - A list of the top 100 product IDs (`freq_products`) is generated to focus further analysis on popular items.

3. DATA FILTERING AND TRANSFORMATION

Selecting Frequent Products:

- The dataset is filtered to include only orders with the top 100 frequent products, narrowing the focus to high-interest items.

- **Merging Product Names:**

- Product names are merged with the dataset for clarity in identifying items.

- **Data Reshaping for Basket Analysis:**

- The data is restructured into a matrix format where each row represents an order and each column represents a product, suitable for basket analysis.

- **Binary Encoding:**

- Binary encoding is applied to mark the presence of each product in an order, preparing the data for analysis by indicating whether a product was purchased (1) or not (0).

4. MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM FREQUENT ITEM SET

- **Identification:**

The Apriori algorithm is applied with a minimum support threshold of 0.005, capturing item combinations that appear in at least 0.5% of orders, ensuring focus on significant item sets.

- **Association Rule Generation:**

Association rules are generated using lift as the primary metric, with a minimum threshold of 1, to highlight strong associations between products

- **Apriori instead of FP-Growth:**

Apriori provides control over minimum support at each stage, allowing analysts to adjust thresholds easily to find frequent itemsets at specific support levels.

FP-Growth often requires a single, pre-determined support threshold so that's why even if FP-Growth is convenient for larger dataset due to less tuning but due to adjustment of support thresholds it can be suggested to use Apriori Algorithm.

SUMMARY OF ITEMSETS AND THEIR SUPPORT VALUES

	support	itemsets	
0	0.015990	(100% Raw Coconut Water)	
1	0.026006	(100% Whole Wheat Bread)	
2	0.015813	(2% Reduced Fat Milk)	
3	0.036208	(Apple Honeycrisp Organic)	
4	0.028253	(Asparagus)	

- **100% Raw Coconut Water:** Appears in 1.60% of transactions, indicating moderate popularity.
- **100% Whole Wheat Bread:** Has a 2.60% support, reflecting a reasonably high frequency of purchase.
- **2% Reduced Fat Milk:** Purchased in 1.58% of transactions, showing a moderate presence in the dataset.
- **Apple Honeycrisp Organic:** The most popular among the listed items, with a 3.62% support, indicating it is frequently bought.
- **Asparagus:** Appears in 2.83% of transactions, showing a solid level of customer interest.

These itemsets represent a range of product types with varying frequencies of purchase, with organic and fresh produce (like apples and asparagus) showing higher support compared to packaged items like coconut water and bread.

5. PRODUCT BUNDLING STRATEGY

- **Filtering Rules for Strong Bundling Potential:**
 - Rules are filtered to identify potential product bundles with:
 - Lift > 2: Demonstrating a meaningful relationship between products.
 - Confidence > 0.2: Indicating a reliable pattern in purchasing behavior.
- **Top Product Combination Selection:**
 - The filtered rules are sorted by lift and confidence, with the top 10 high-lift and high-confidence product combinations identified as the most promising bundles.
- **Targeted Customer Segments:**
 - The identified product bundles are analyzed to cater to specific customer segments (e.g., organic shoppers, health-conscious buyers), enhancing the effectiveness of the bundles for targeted marketing campaigns.

SUMMARY OF KEY PRODUCT BUNDLING INSIGHTS

	antecedents	consequents	support	confidence	lift
170	(Lime Sparkling Water)	(Sparkling Water Grapefruit)	0.005634	0.286382	8.947487
147	(Bunched Cilantro)	(Limes)	0.005178	0.269832	4.527121
176	(Organic Cilantro)	(Limes)	0.007344	0.248943	4.176653
303	(Raspberries)	(Strawberries)	0.005193	0.210169	3.470076
263	(Organic Lemon)	(Organic Hass Avocado)	0.009059	0.243077	2.680551
310	(Organic Strawberries, Bag of Organic Bananas)	(Organic Hass Avocado)	0.006297	0.239217	2.637984
306	(Organic Baby Spinach, Bag of Organic Bananas)	(Organic Hass Avocado)	0.005085	0.238644	2.631672
242	(Organic Cucumber)	(Organic Hass Avocado)	0.007409	0.218137	2.405520
311	(Organic Strawberries, Organic Hass Avocado)	(Bag of Organic Bananas)	0.006297	0.364745	2.258701
271	(Organic Tomato Cluster)	(Organic Hass Avocado)	0.005575	0.204465	2.254750
278	(Organic Kiwi)	(Organic Strawberries)	0.005295	0.252130	2.232225
284	(Organic Raspberries)	(Organic Strawberries)	0.014324	0.247448	2.190769
305	(Organic Hass Avocado, Organic Baby Spinach)	(Bag of Organic Bananas)	0.005085	0.346620	2.146462
312	(Organic Hass Avocado, Bag of Organic Bananas)	(Organic Strawberries)	0.006297	0.239091	2.116783
239	(Organic Blueberries)	(Organic Strawberries)	0.010112	0.237277	2.100725
290	(Organic Whole String Cheese)	(Organic Strawberries)	0.005843	0.229919	2.035577
50	(Organic Navel Orange)	(Bag of Organic Bananas)	0.005923	0.324205	2.007657

1. SPARKLING WATER PAIRING:

- Lime Sparkling Water often pairs with Grapefruit Sparkling Water (Lift: 8.95).

2. CILANTRO AND LIMES:

- Both regular and organic cilantro frequently co-occur with limes, suggesting a fresh produce bundle.

3. BERRY COMBOS:

- Raspberries and strawberries, as well as organic blueberries and raspberries, show strong associations, ideal for berry packs.

4. ORGANIC PAIRINGS:

- Organic lemons and avocados, as well as combinations like strawberries, bananas, and avocados, appeal to organic shoppers.

5. BANANAS AS A COMPLEMENT:

- Organic bananas commonly pair with items like organic oranges and avocados, ideal for a health-focused fruit bundle.

These bundles target popular pairings and healthy choices, enhancing product appeal in fresh and organic categories.

6. INSIGHTS AND APPLICATIONS FOR PRODUCT BUNDLING

- **Suggested Bundles:**

Based on strong association rules, bundles are recommended that align with common customer buying patterns, enhancing cross-selling opportunities.

- **Cross-Promotion and Marketing Opportunities:**

High-lift product pairs provide ideal candidates for cross-promotions, encouraging sales of both primary and complementary items.

- **Enhanced Customer Experience:**

Bundling frequently purchased items has the potential to simplify shopping, improve customer satisfaction, and encourage repeat purchases.

7. CONCLUSION

- The market basket analysis successfully identifies patterns in customer purchasing behavior, enabling data-driven bundling strategies.
- Suggested product bundles can be effectively used in marketing, promotions, and e-commerce applications to boost average order value and overall sales.