E-Commerce Purchase Behavior & Association Patterns Analysis

In this report, we analyze the purchasing behavior of users in an e-commerce store specializing in electronic items. Our goal is to uncover key insights from the purchase history to enhance business decision-making through data-driven strategies. By applying unsupervised learning algorithms to the dataset, we identify purchasing patterns and segment users, ultimately providing actionable recommendations for personalized marketing and product offerings.
As e-commerce continues to grow rapidly, understanding customer behavior, preferences, and feedback is essential for businesses to optimize product recommendations, improve user experiences, and drive sales.

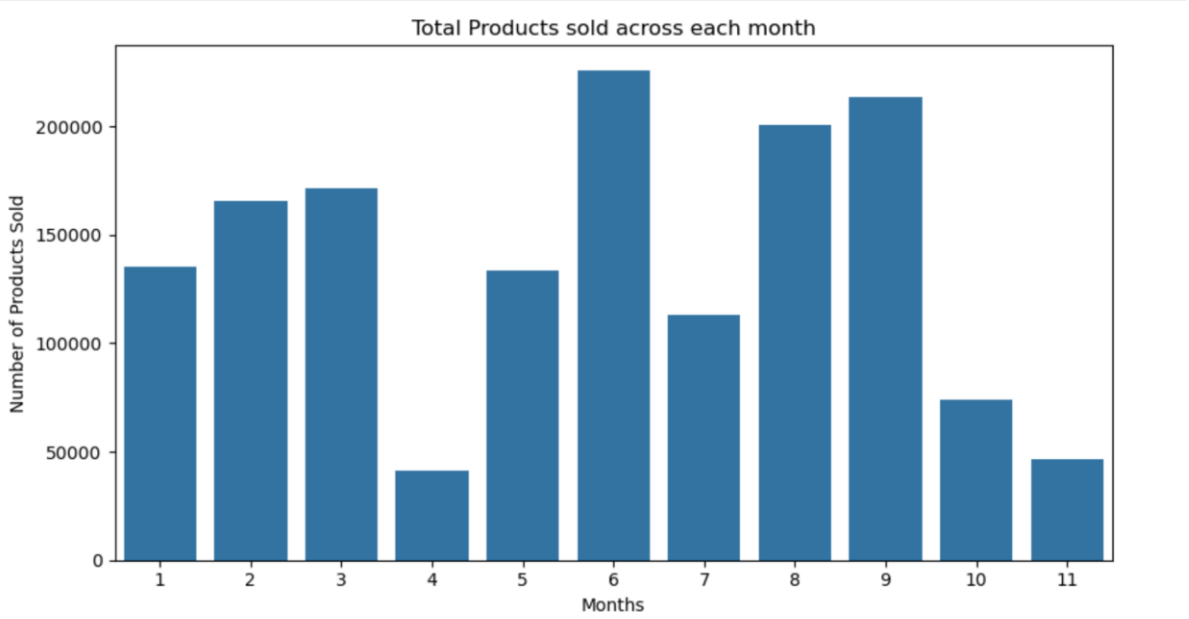
The goal is to segment customers in different groups/clusters and also recommend items based on the clusters as well as based on purchases.

The dataset contains over 2.6 million rows and 8 columns,capturing the transactional data of customers purchasing electronic items. Below are the features in the dataset:

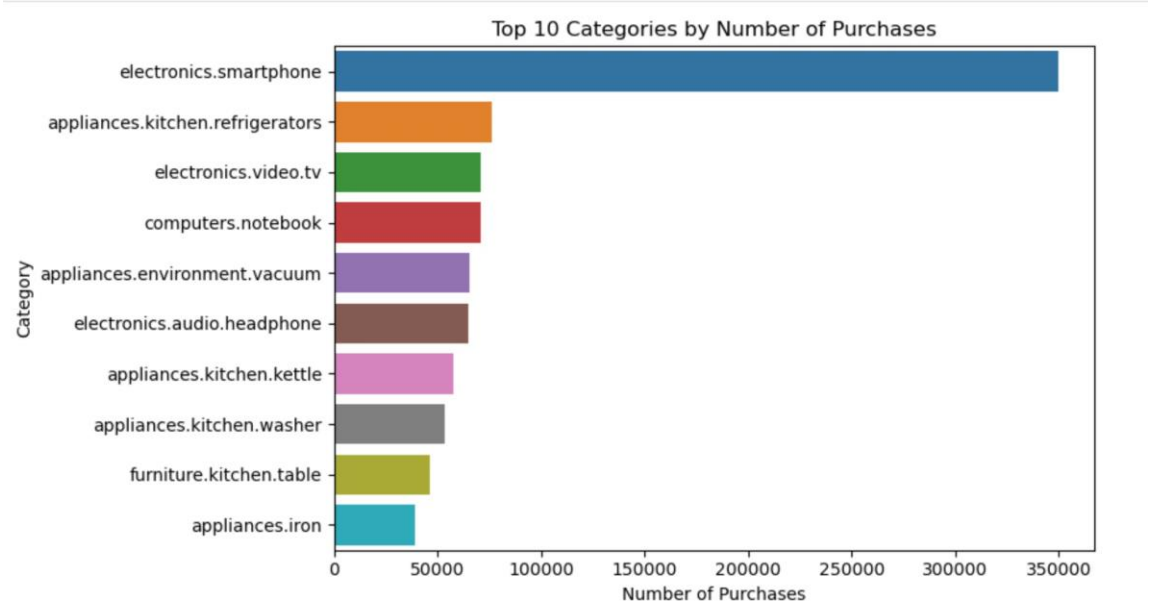
- event_time - Timestamp when the purchase was made
- Order_id - unique ID of the order
- Product_id - unique id of the product
- Category_id - unique ID identifying a specific category
- Category_code - the category in which the electronic item belongs (tablet, headphones, mobile, etc)
- Brand - company/brand name of the product (Samsung, Apple, etc)
- Price - the price of the product
- User_id - unique ID of the user who made the purchase

It has purchase data of different items from an online store for the year 2020.

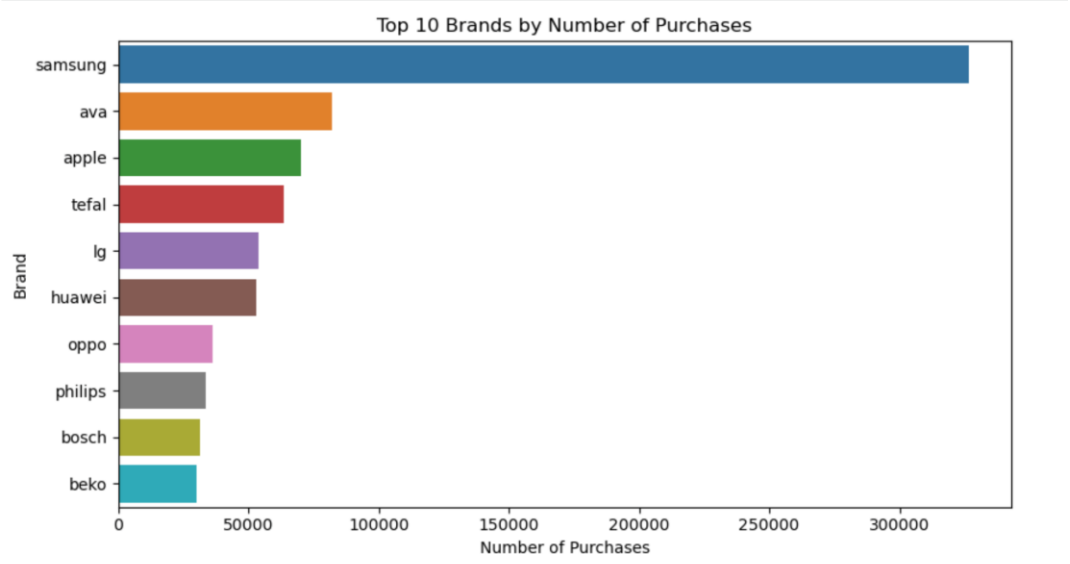
For data cleaning we have dropped the user_id row because it has over 90% missing values and then dropped the missing rows from the other columns.



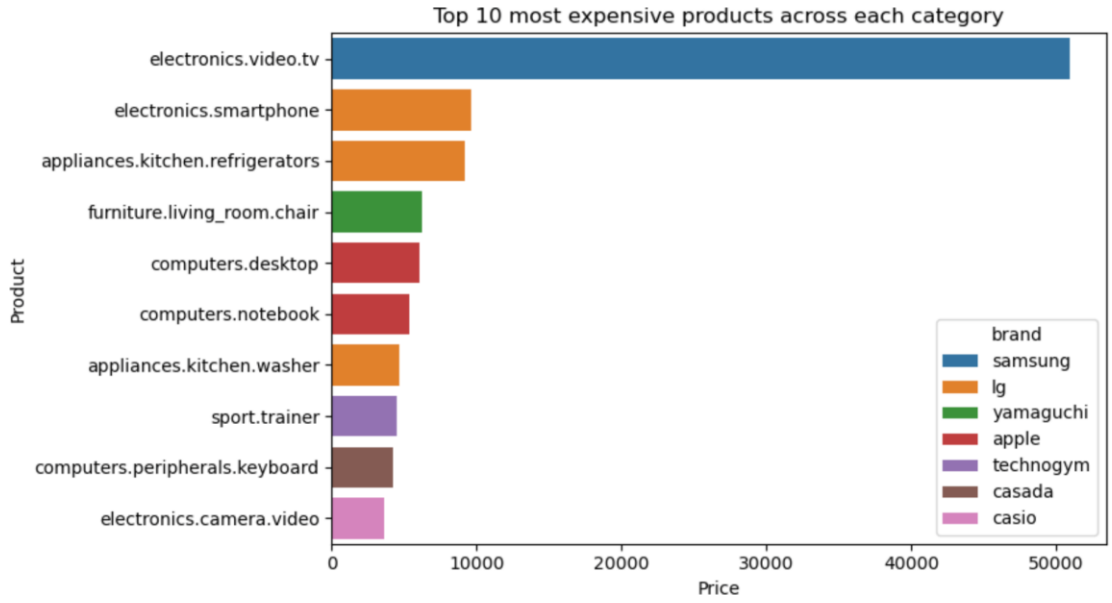
The bar graph illustrates the distribution of product sales across each month, highlighting key trends in customer purchasing behavior. Sales begin at a steady level in January and February, with over 150,000 products sold each month. March sees a slight increase, followed by a noticeable dip in April and May, suggesting a potential seasonal slowdown or lower demand during this period. However, sales peak dramatically in June and remain high through September, with June achieving the highest monthly sales exceeding 200,000 products. This trend suggests heightened shopping activity during the mid-year, possibly due to seasonal promotions or increased demand for specific categories. Sales begin to decline in October and November, indicating a slowdown as the year progresses, which could inform inventory and marketing strategies during these months.



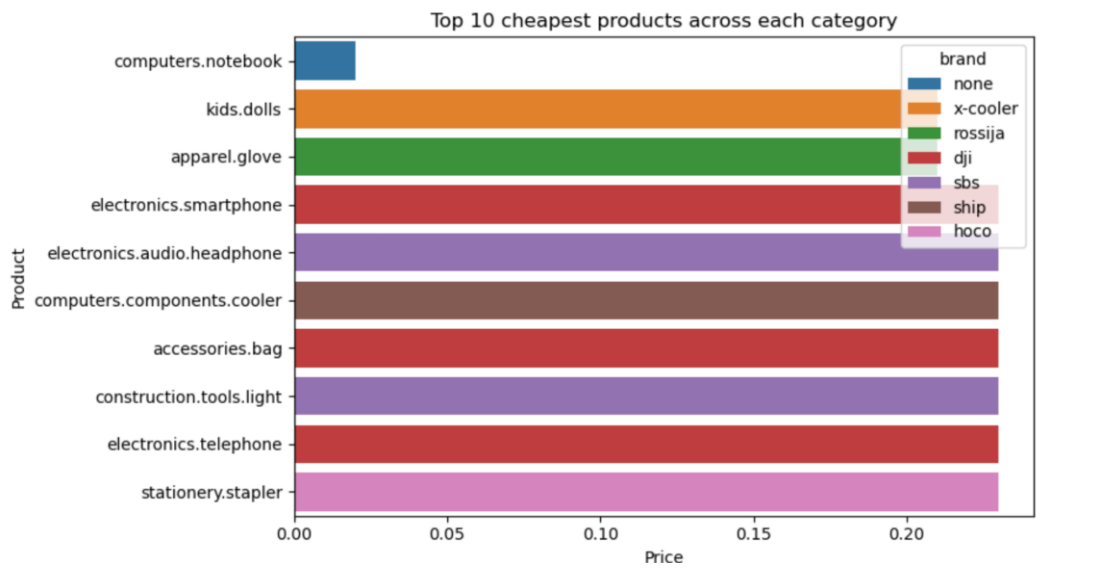
Smartphones were by far the most popular item, with more than 350,000 purchases, showing their importance during that time. Refrigerators and TVs followed, reflecting a focus on home upgrades. Items like laptops, vacuum cleaners, and headphones also saw steady demand, likely driven by remote work, learning, and entertainment needs. Kitchen appliances such as kettles and washers, along with kitchen tables, hint at people creating comfortable and functional home spaces.



Samsung dominates the scene, with over 300,000 purchases, far outpacing other brands. This highlights its strong presence across various product categories, especially smartphones. Following Samsung, brands like Ava, Apple, and Tefal make a significant impact, showing steady demand for their products. LG, Huawei, and Oppo also hold Strong presence, reflecting consumer interest in their offerings. Meanwhile, Philips, Bosch, and Beko rounded out the top 10, showing their relevance in specific product categories.



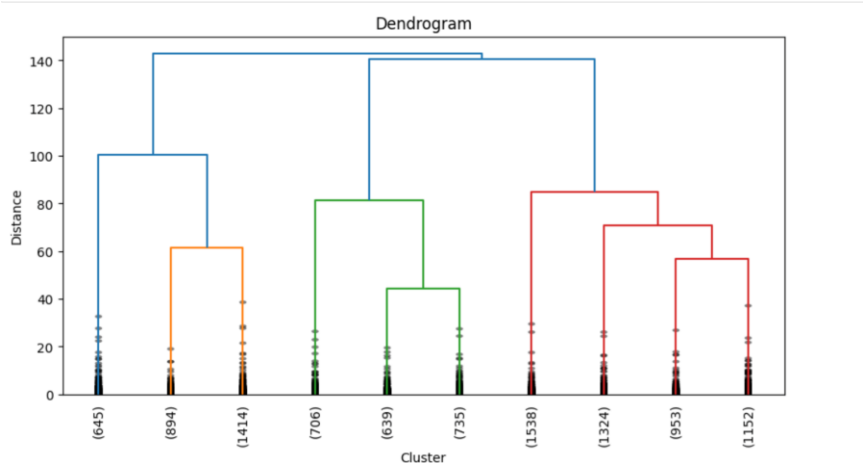
Televisions dominate the list with an exceptionally high price, reflecting the premium nature of high-end models, particularly from Samsung which might be a custom made TV or an outlier because of its exorbitant price (50000). Other high priced products like smartphones, refrigerators, and furniture items like living room chairs, emphasizing the value of technology and home essentials. Desktop computers, notebooks, and washers also make the list, pointing to investments in productivity and home functionality. Additionally, sport trainers, keyboards, and cameras appear as high-value items. The LG smartphone which is around $10000 could potentially be an error or a specialized custom product.



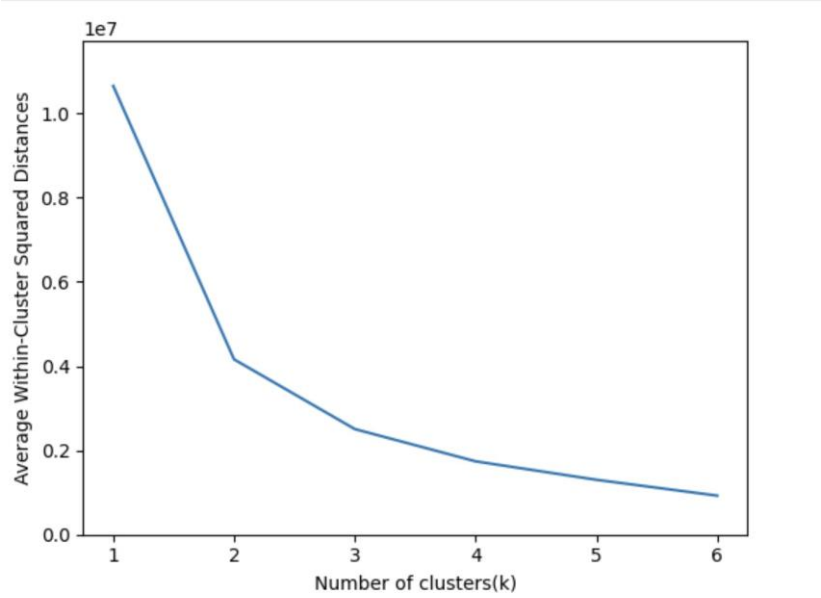
Notebooks, smartphones, coolers, and telephones are listed with incredibly low prices, which seems highly improbable for such items, suggesting potential data inaccuracies or outliers. Other affordable items, such as kids' dolls, gloves, and accessories like bags and staplers, align more with expectations for low-cost products.

User Segmentation and Recommendations

User segmentation using clustering



Based on the visual representation of the data through the dendrogram, it is evident that the dataset naturally divides into three distinct clusters. These clusters indicate that the data points within each group share strong similarities, while there are noticeable differences between the groups.



The elbow point at k=3 suggests that 3 clusters is the optimal number for this data. At this point, the clusters are well-formed, and further increasing k leads to diminishing returns, making additional clusters unnecessary.

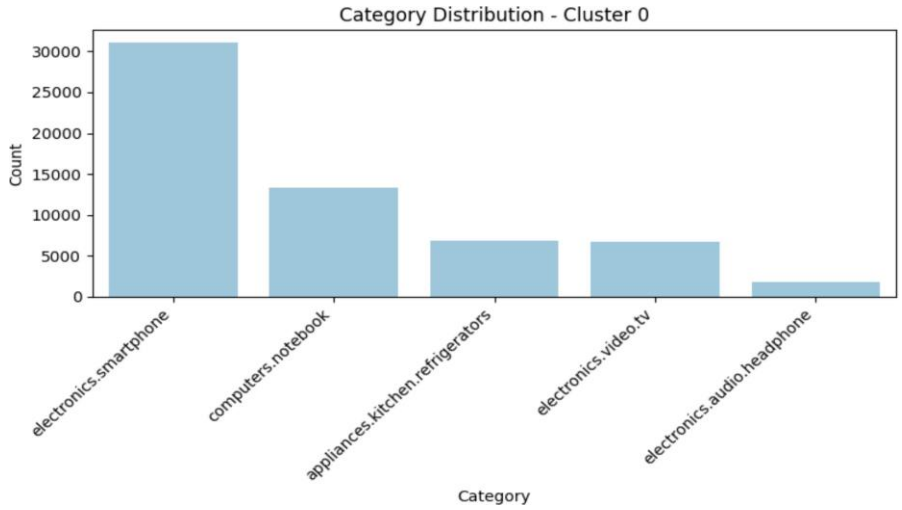
1. Cluster 0: Mid-Volume, High-Value Products

Key Characteristics:

Total Products: 67,424 – medium-sized cluster.
Average Price: $954.09 – higher-value products.
Price Range: $53.22 to $11,574.05 – spans mid to premium pricing tiers (very expensive).

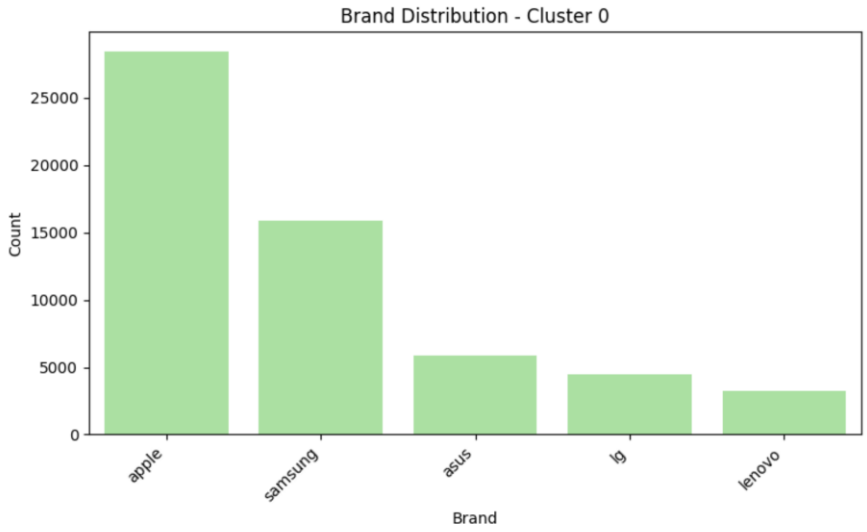
Category Distribution:

Dominated by electronics.smartphone (31,034 products), followed by computers.notebook and other tech categories.



Brand Distribution:

Apple is the leading brand, followed by Samsung, Asus, and others, which are prominent in high-value consumer electronics.

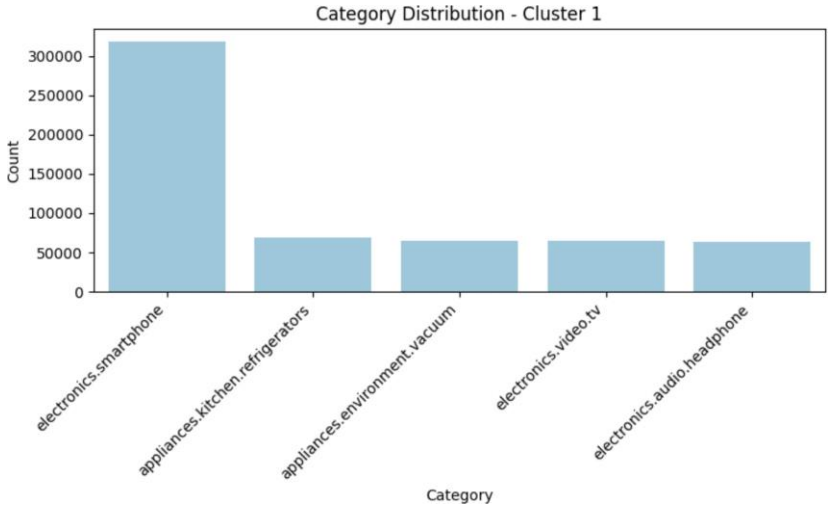


Characterization:
This cluster represents mid-volume, high-value products, appealing to consumers purchasing premium gadgets, smartphones, and laptops.

2. Cluster 1: High Volume, Low-Value Products

Key Characteristics:
Total Products: 1,453,969 – the largest cluster by volume.
Average Price: $161.13 – indicates low-to-mid-range pricing.
Price Range: $0.02 to $1,504.61 – ranges from super cheap to a bit expensive products

Category Distribution:
Dominated by electronics.smartphone (318,565 products), followed by appliances.kitchen and other affordable product categories.



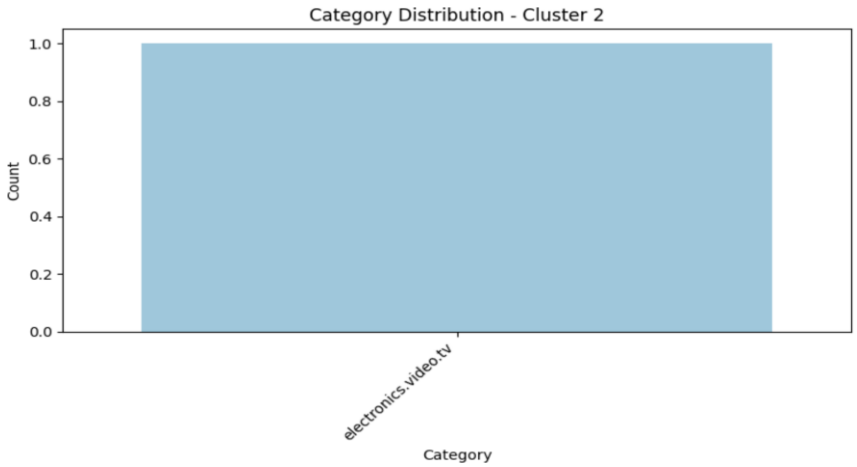
Brand Distribution:
Samsung dominates, followed by Ava and Tefal, which are likely manufacturers of affordable, high-volume consumer goods.

Characterization:
This cluster represents high-volume, low-value products, with affordability as the primary appeal. It reflects the mass-market consumer base.

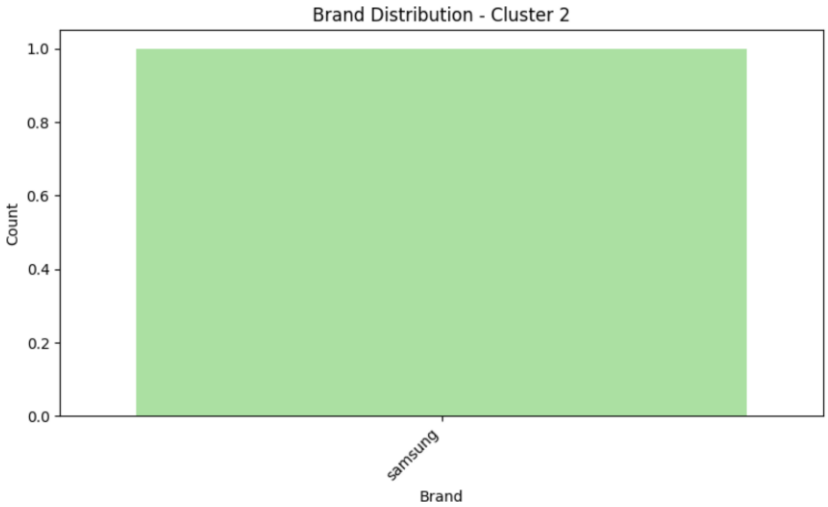
3. Cluster 2: Rare, Ultra-Premium Product

Key Characteristics:
Total Products: 1 – an outlier.
Average Price: $50,925.90 – exceptionally high price.
Price Range: Fixed at $50,925.90 – single ultra-premium product.

Category Distribution:
Solely electronics.video.tv, suggesting a high-end television.



Brand Distribution:
Only Samsung, potentially representing a flagship or custom-built product.



Characterization:
This cluster represents an outlier or ultra-premium product, appealing to an extremely niche audience.

Recommendation with rules

Our analysis revealed fascinating insights into customer purchasing behavior, uncovering associations that range from well-established brand ecosystems to niche, cross-category patterns. These findings can guide targeted recommendations, bundled offers, and marketing strategies to better serve diverse customer segments.

Customers who purchase or interact with Samsung headphones are strongly inclined to purchase Samsung smartphones as well. This reflects a consistent loyalty to the Samsung brand, where customers prefer staying within the brand ecosystem for complementary products. Retailers can capitalize on this behavior by promoting bundled offers such as headphones alongside smartphones or targeted accessory recommendations to increase basket size.

A particularly strong pattern emerges around AWAX smartphones and niche TV brands like Megogo and Okko.

Customers buying AWAX smartphones frequently purchase Megogo TVs or Okko TVs, and the reverse relationships hold true.

These bi-directional dependencies, reflected in Rules 6–9, have exceptionally high lift values (111.59 and 109.88), suggesting that customers view AWAX smartphones and these smart TVs as complementary products.

This pairing likely caters to a focused demographic—customers seeking affordable, less mainstream brands for their home entertainment needs. This may be influenced by regional pricing strategies or bundled promotions. Retailers can leverage this by creating combo deals, such as AWAX smartphones and Megogo/Okko TVs, to encourage cross-category purchases.

An unusual but intriguing relationship exists between Dogland pillows and Apple smartphones (Rule 2). While seemingly unrelated, the rule shows a confidence of 55.6%, indicating that over half of the customers purchasing Dogland pillows also buy Apple smartphones. However, the low support value (0.33%) highlights this as a rare behavior occurring within a niche group.

This link points to a lifestyle-based customer segment—households that value both premium home comfort and high-end technology. For example:

Pet-centric households or customers focused on home aesthetics might also prioritize owning premium gadgets.

Retailers could craft targeted campaigns, such as luxury living for Tech-Savvy, Pet-Loving Homes, to connect seemingly unrelated product categories and tap into this niche.

Rule 4 highlights a subtle yet interesting transition: customers purchasing Samsung smartphones may experiment with AWAX smartphones. Though confidence is low (2.66%), the lift value (3.58) indicates that AWAX smartphones are over-represented compared to random chance.

This pattern could signal early signs of shifting customer interest toward emerging or cost-effective brands like AWAX. Retailers can monitor these and explore promotional campaigns for AWAX products, targeting cost-conscious customers or tech explorers seeking alternative options.

While mainstream brands like Samsung and Apple dominate purchasing behavior, their lift values (under 4) indicate broader, less-targeted associations. For example:

Samsung smartphones → Samsung headphones (Rule 0) is a reliable link but lacks the focus seen in niche pairings.

In contrast, niche brands like AWAX, Megogo, and Okko exhibit specialized relationships with far higher lift values, emphasizing their importance to smaller, more focused customer groups.

Final Recommendations:

To enhance the recommendation system and drive customer engagement:

- Implement brand-specific recommendations for Samsung, Apple, and Awax products.
- Develop bundle deals targeting entertainment enthusiasts (smartphones and TVs) and lifestyle shoppers (home goods and premium electronics).
- Personalize recommendations through the clusters obtained based on their preferences and spending behavior.
- Use cross-category recommendations to encourage customers to explore complementary products they might not have considered, boosting overall cart value.

Source of dataset - https://www.kaggle.com/datasets/mkechinov/ecommerce-purchase-history-from-electronics-store