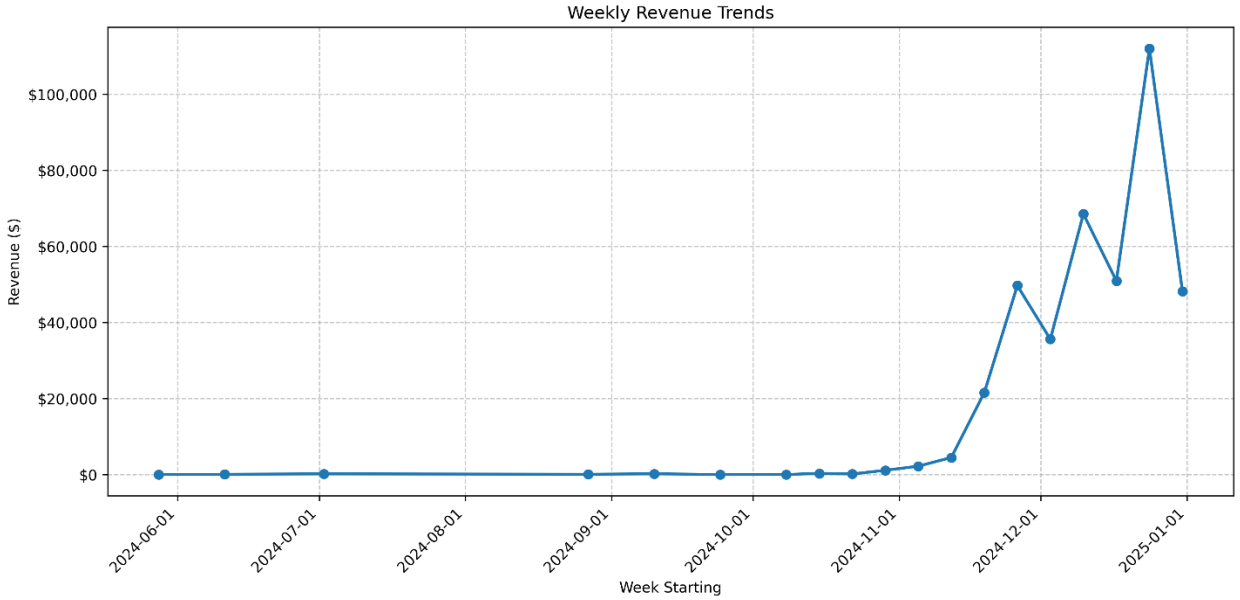
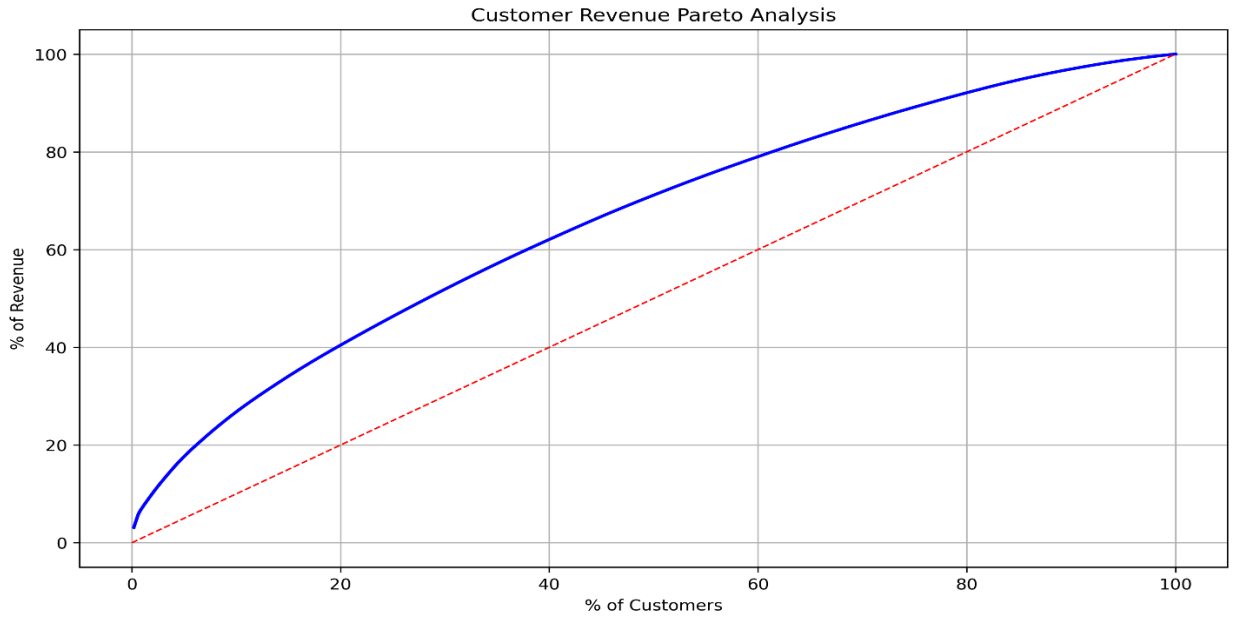
Dataset Summary

The "purchases.csv" dataset contains 500 unique customers making a total of 5443 transactions across 60 distinct products and 12 product categories from June 1st, 2024, through December 31st, 2024, for a total revenue of $395,510.99. The data shows an abrupt pickup in weekly sales starting during the middle of November and continuing on a strong growth trajectory through the end of the year.



The average spending per customer was $791.02 with the standard deviation of spending per customer equal to $728.59. This indicates that some larger spenders may disproportionately contribute to revenue. This is confirmed by a Pareto chart of customer revenue which indicates the top 20% of customers account for approximately 40% of the revenue for the store.



Top Products

The top 10 products by revenue showcase diverse price points and purchasing patterns. The Ergonomic Chair leads revenue generation at $42,448.63 with 137 sales, indicating strong demand for home office equipment at a mid-high price point. The 4K Smart TV, while having fewer sales (48), generates the second-highest revenue ($35,699.52) due to its premium pricing. Wireless Earbuds Pro demonstrates broad market appeal with the highest number of sales (177) among top products, generating $26,918.23 in revenue.

High-ticket items like the Gaming Console X1 and Patio Furniture Set contribute significantly to revenue despite lower sales volumes (36 and 27 sales respectively). Conversely, products like the Skincare Bundle, Fitness Tracker, and Dress Shirt maintain strong revenue through higher sales volumes (143, 140, and 131 sales respectively) at more moderate price points. This diverse mix of high-volume moderate-priced items and lower-volume premium products suggests a well-balanced product portfolio catering to different customer segments and price sensitivities.

Product Name	Number of Sales	Total Revenue
Ergonomic Chair	137	$42,448.63
4K Smart TV 55-inch	48	$35,699.52
Wireless Earbuds Pro	177	$26,918.23
Gaming Console X1	36	$18,699.64
Language Learning Set	131	$17,298.69
Patio Furniture Set	27	$15,899.73
Robot Vacuum	44	$14,049.56
Skincare Bundle	143	$12,438.57
Fitness Tracker	140	$11,198.60
Dress Shirt	131	$9,605.69

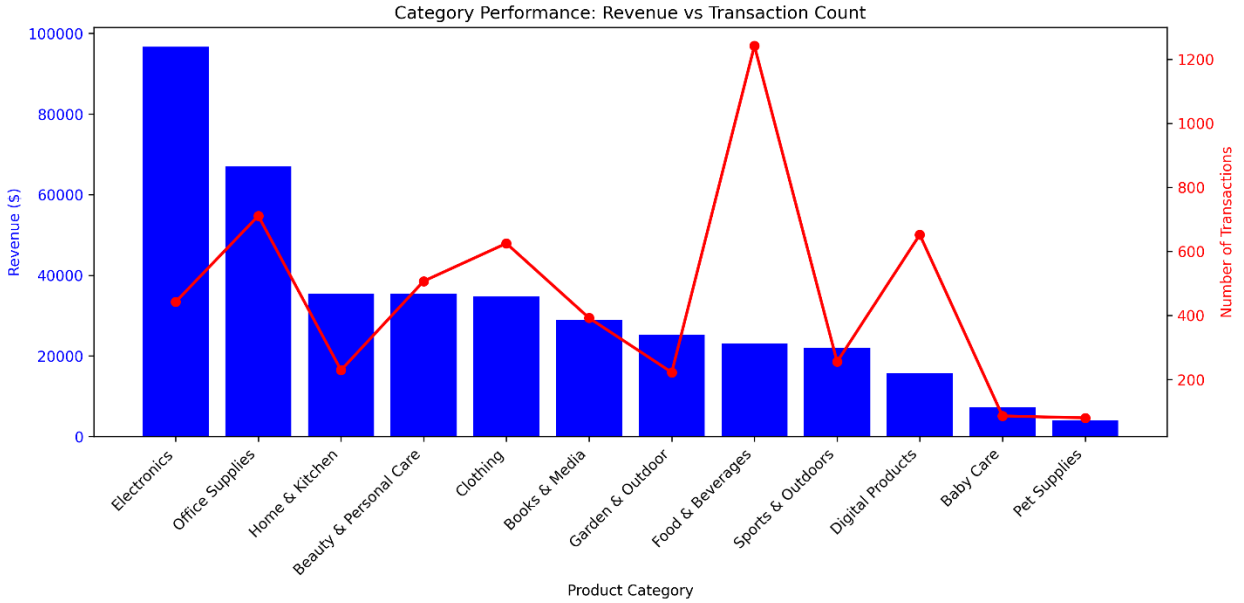
Top Categories

Analysis of top product categories reveals distinct patterns in revenue generation and transaction volume. Electronics leads in revenue at $96,705.58, driven by high-value items like smart TVs and gaming consoles, despite having moderate transaction volume (442 sales). Office Supplies ranks second in revenue ($67,012.89) with the second-highest transaction count (711), suggesting consistent demand for workplace essentials.

An interesting contrast appears in Food & Beverages, which leads in transaction volume (1,242 sales) but ranks eighth in revenue ($23,059.06), indicating frequent, lower-value purchases. Similarly, Digital Products shows high transaction volume (652) but lower revenue ($15,743.48), characteristic of subscription-based and lower-priced digital goods.

Mid-tier categories like Home & Kitchen, Beauty & Personal Care, and Clothing each generate between $34,000-$35,500 in revenue but with varying transaction volumes, suggesting different price point strategies across these categories. Books & Media, Garden & Outdoor, and Sports & Outdoors maintain moderate performance in both metrics, while Baby Care and Pet Supplies represent smaller but potentially growing segments.

This category analysis suggests opportunities for revenue optimization, particularly in high-transaction categories where average transaction value could potentially be increased through strategic pricing or product mix adjustments.



Customer Segmentation

Our customer segmentation implementation employs a sophisticated hybrid approach combining traditional machine learning with modern LLM capabilities. The process consists of two main components: statistical clustering and dynamic segment naming.

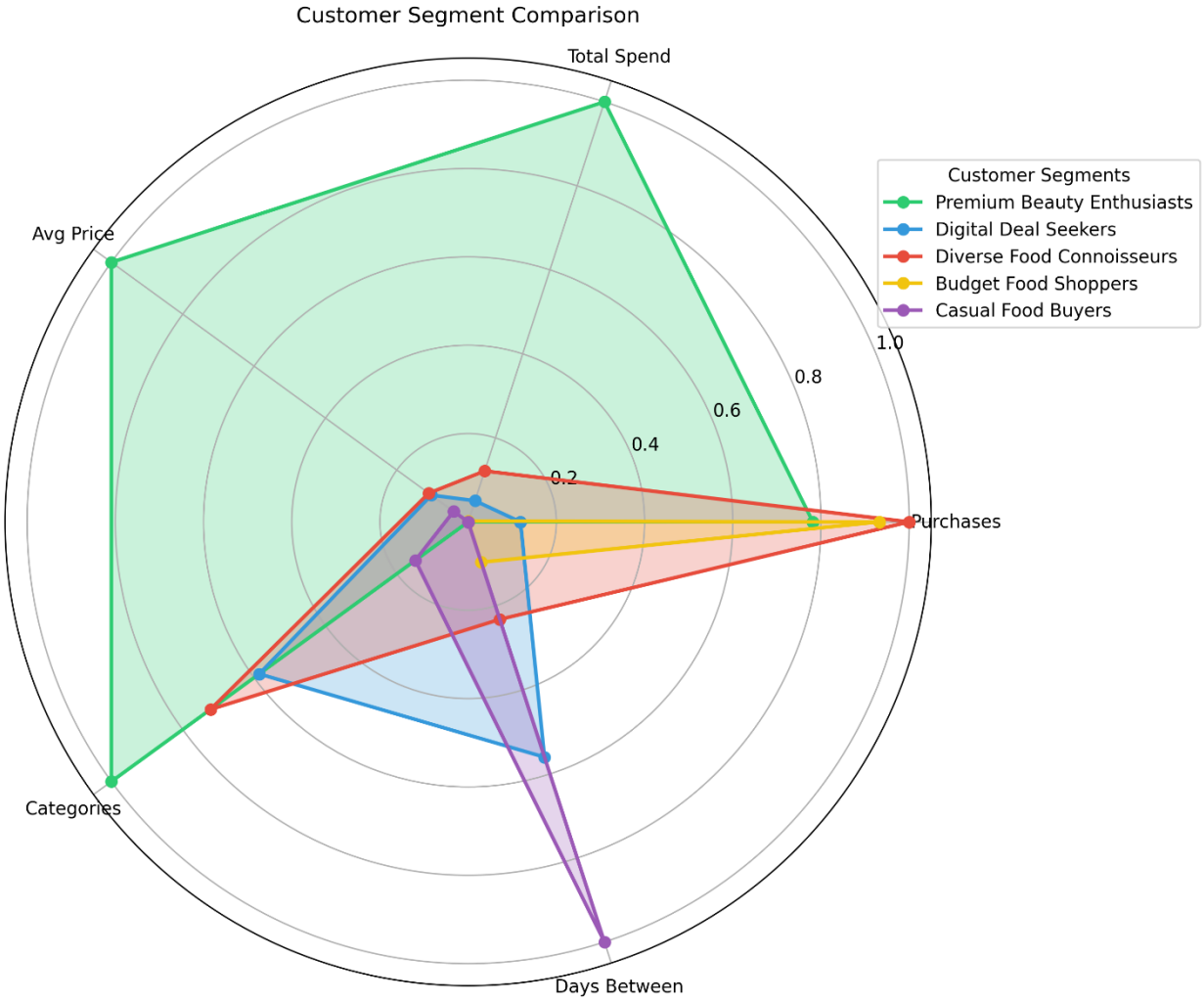
First, we utilize K-means clustering with the following customer features:

- Number of purchases
- Total spending
- Average price per purchase
- Number of unique categories shopped
- Average days between purchases

The data is preprocessed using StandardScaler to normalize the features before clustering. Our implementation is flexible, allowing for anywhere between 2-10 segments to be created at runtime, though we selected 5 segments for this analysis based on the natural groupings in our data.

What makes our approach particularly innovative is the integration of the Anthropic Claude API for dynamic segment naming. Rather than using static labels or manual naming, we make an API call to Claude 3.5 Sonnet with a structured prompt that includes the statistical characteristics of each segment. The prompt engineering is designed to present Claude with the numerical cluster data, request JSON-formatted responses for consistency, and guide Claude to create intuitive, business-relevant names based on the segment characteristics

While the segmentation generally works the same, different run-throughs will lead to different names. For example, in one instance of our 5-segment analysis, Claude identified a group called "Luxury Beauty Enthusiasts" whereas in a separate run it categorized that same segment as "Premium Beauty Enthusiasts". This lack of determinism could be solved by setting the temperature to zero in the API to guarantee consistency of responses.



The radar chart visualizes the distinct characteristics of our five customer segments across key metrics. Premium Beauty Enthusiasts (green) stand out dramatically with the highest total spend and the highest average price point, while also showing strong diversity in category purchases. Digital Deal Seekers (blue) and Diverse Food Connoisseurs (red) show moderate values across most metrics, with the latter maintaining slightly higher purchase frequency. Budget Food Shoppers (yellow) and Casual Food Buyers (purple) both demonstrate lower spending patterns but differ in their purchase frequency - Budget Shoppers make more frequent purchases with lower average prices, while Casual Buyers have more days between purchases. All segments show unique "fingerprints" in their metric combinations, validating the effectiveness of our clustering approach in identifying distinct customer behavior patterns.

Customer Recommendation

The product recommendation system employs a collaborative filtering approach combined with price sensitivity analysis to generate personalized recommendations. The system first creates a user-product matrix from purchasing history, then uses cosine similarity to identify customers with similar purchasing patterns. From there it identifies any products that have been purchased by the cohort of similar customers that have not already been purchased by the customer we are making recommendations for.

The recommendation algorithm balances two key factors when ordering its recommendations out of the candidate list of products: purchase frequency among similar customers and price similarity to the target customer's average purchase amount. The system uses a learned weight parameter (optimized through testing on historical data) to balance these factors effectively, with accuracy shown to peak when purchase frequency is weighted at 0.84 and price similarity at 0.16.

Looking at Customer 314's purchase history over the past month, we see a diverse but focused pattern of spending across several categories. Their total spending of $275.90 across 10 purchases averages to $27.59 per transaction. While Beauty & Personal Care emerges as their favorite category by total revenue (driven largely by a significant $79.99 Skincare Bundle purchase), they show consistent purchasing in Food & Beverages with repeat purchases of Sparkling Water Pack and Organic Coffee Beans. Their purchasing history also includes clothing items (Casual T-Shirt and Classic Denim Jeans), office supplies, and a digital subscription, indicating a customer who shops across multiple categories but maintains consistent price points within each.

The system's top 5 recommendations strongly align with this purchasing pattern. The Trail Mix Bundle ($19.61) received the highest recommendation score (0.964) due to its price similarity to their frequent food and beverage purchases and its high purchase frequency among similar customers. The Electric Toothbrush ($69.99) represents an opportunity to expand their beauty care purchases in a price range they've shown comfort with through their Skincare Bundle purchase. The Protein Bar Pack ($24.99) and Gourmet Chocolate Box ($29.99) align with both their demonstrated interest in Food & Beverages and their typical spending range. The Makeup Set ($99.99), while at a higher price point, builds on their established interest in beauty products, suggesting potential for category expansion. These recommendations successfully balance maintaining category affinity while offering opportunities for careful price-point expansion within their demonstrated interests.