Customer Segmentation, Sales Analysis, and Lookalike Modeling

1. Introduction

This report presents a comprehensive analysis of customer behavior, purchasing patterns, and product performance based on recent transactional and profile data.

The analysis integrates:

- Customer Segmentation: Grouping customers based on spend, quantity purchased, product preferences, and demographics.
- Sales Insights: Trends, regional performance, product-level analysis, and correlation metrics.
- Lookalike Modeling: Identification of highly similar customers using feature-based similarity scoring to support targeted marketing.

The objective is to provide actionable insights that can guide personalized marketing strategies, inventory optimization, and revenue growth initiatives.

---

2. Methodology

2.1 Customer Segmentation

- Feature Engineering: Total spend, average order value, total quantity purchased, favorite category, geographic region, and signup date.
- Clustering Algorithm: K-Means clustering with k = 4.
- Evaluation Metric: Davies-Bouldin Index (DB Index) to measure cluster separation.

2.2 Sales & Trend Analysis

- Aggregated transactional data was analyzed to determine top-performing regions, products, and categories.
- Monthly sales trends and Pareto analysis (80/20 rule) were used to identify revenue drivers.
- Correlation heatmaps measured relationships between price, quantity, and revenue.

2.3 Lookalike Modeling

- Process: Customer data was merged, features were engineered, and categorical fields were one-hot encoded.
- Normalization: Feature scaling was applied before computing cosine similarity.
- Filtering: Matches with a similarity score ≥ 0.90 were retained.
- Output: A ranked list of lookalikes for each customer, with none having more than 20 lookalikes (most had fewer).

---

3. Results

3.1 Clustering Summary

- Number of Clusters: 4
- DB Index: 0.6017 (good separation)
- Cluster Characteristics:
    1. Cluster 0 – High-Value Bulk Buyers:
        - High spenders (all in top total spend quartile: $4,764–$10,674).
        - Largest quantities purchased (majority in top quartile: 6–11 units).
        - Balanced category preferences; slightly higher interest in Electronics and Books.
        - Concentrated in South America and Asia.
    2. Cluster 1 – Core Average Buyers:
        - Moderate total spending ($2,137–$3,137).
        - Mid-range quantities purchased (mostly 3–5 units).
        - Broad product interest with Books and Electronics slightly ahead.
        - Evenly distributed across all regions.
    3. Cluster 2 – Low-Engagement Customers:
        - Low total spend (<$2,137) and lowest quantities purchased (<3 units).
        - Slight preference for Clothing and Books.
        - Distributed evenly across all regions.
        - Likely mix of new and inactive customers.
    4. Cluster 3 – High-Value Niche Buyers:
        - High total spending (mostly $3,137–$4,764).
        - Strong preference for Books, followed by Home Decor.
        - Higher presence in Europe and South America.

3.2 Sales Insights

Regional Performance:

- Top Region: South America (highest total sales value, >210k).
- Middle Tier: Europe (>160k).
- Lowest: Asia and North America (~150k).

Product Performance:

- Top Product by Revenue: ActiveWear Smartwatch.
- Top Product by Quantity Sold: ActiveWear Smartwatch (~100 units), followed closely by SoundWave Headphones.
- Across categories, Books and Electronics dominate the top sellers.

Category-Level Highlights:

- Books: Strong recurring demand (SoundWave Cookbook, BookWorld Biography, TechPro Textbook).
- Clothing: Consistent mid-tier sales volumes (ActiveWear Jacket, TechPro T-Shirt).
- Electronics: High revenue impact (Smartwatch, Headphones, Smartphone).
- Home Decor: Steady niche performance (HomeSense Desk Lamp, ActiveWear Rug).

Pareto Analysis (80/20 Rule):

- A small subset of products (top ~20%) drives the majority of sales volume, confirming concentration of demand.

Monthly Trends:

- Peaks observed in January, May, July, and September.
- Notable dips in March, June, and November, indicating potential seasonal effects or promotion gaps.

Correlation Analysis:

- Price and revenue show a strong positive correlation (0.72).
- Quantity sold has a moderate correlation with revenue (0.61), suggesting that both price strategy and volume contribute significantly to sales performance.

3.3 Lookalike Model Insights

- High Similarity Pairs: Many customers achieved similarity scores >0.99, indicating strong behavioral matches.
- Marketing Opportunity: These lookalike pairs can be used to replicate successful campaigns—e.g., promotions that worked for one customer can be extended to their top lookalikes.

- Engagement Potential: Lower-engagement customers who are lookalikes of high-value buyers can be prioritized for targeted upsell campaigns.
- Scalability: Customers with the most lookalikes had, with the majority of customers having fewer, suggesting clear but selective matching rather than overly broad similarity.

---

4. Strategic Recommendations

Cluster-Based Actions:

- Cluster 0: Maintain loyalty with VIP programs, early product access, and cross-category recommendations.
- Cluster 1: Introduce tiered incentives to push into higher spend quartiles.
- Cluster 2: Deploy re-engagement campaigns and welcome offers to spark activity.
- Cluster 3: Deepen niche engagement with targeted campaigns in Books and Home Decor.

Sales-Driven Actions:

- Focus on maintaining stock and marketing for the top 20% products driving most revenue.
- Time promotions around seasonal peaks (January, May, July, September).
- Introduce targeted offers during historically low months (March, June, November).

Lookalike Model Applications:

- Personalization: Extend high-performing campaigns from active customers to their top lookalikes.
- Acquisition: Use behavioral profiles to identify and target prospects similar to best customers.
- Retention: Monitor and re-engage low-activity lookalikes of high-value segments.

---

5. Conclusion

This integrated approach—combining segmentation, sales trend analysis, and lookalike modeling—provides a multi-dimensional view of customer behavior.

It enables:

- Granular targeting based on cluster traits.
- Product prioritization using revenue and quantity concentration insights.
- Scalable personalization through lookalike matching.

Together, these insights support data-driven decisions for marketing, merchandising, and customer relationship management, with the potential to increase both short-term revenue and long-term customer value.