Customer Behavior Analysis Report

1. Executive Summary

This report synthesizes insights from an integrated analysis of Amazon purchase records (2018–2023) and customer survey data. We uncover temporal, demographic, geographic, and product-level trends; segment customers via RFM; and deliver targeted recommendations to enhance inventory planning, customer retention, pricing, and regional marketing.

---

2. Objectives

- Understand seasonal and daily purchase patterns

- Profile customer demographics and geographies

- Identify top products and bundling opportunities

- Segment customers by Recency–Frequency–Monetary value

- Propose actionable strategies to optimize inventory, retention, pricing, and marketing spend

---

3. Data Overview

- **Transactions** (amazon-purchases.csv): Order date, unit price, quantity, product title & ASIN, category, shipping state

- **Surveys** (survey.csv): Demographic attributes (age, gender, income, etc.) and attitudes toward data sharing

- **Field Mapping** (fields.csv): Translates survey codes into full question text

---

4. Methodology

1. **Data Preparation & Cleaning**

   o Standardized column names; parsed dates; cast numeric types

   o Filled nulls (e.g., unknown titles, states, categories); removed duplicates

   o Engineered features: order hour, weekday, month, income/gender encodings

2. **Exploratory Data Analysis**

   o Time-series: hourly, daily, monthly, and yearly purchase trends

   o Demographics: purchase volume by age, gender, income

   o Geography: state-level purchase count and revenue

   o Product performance: top SKUs, categories, and basket co-occurrences

3. **RFM Segmentation**

   o Computed Recency (days since last order), Frequency (total orders), Monetary (lifetime spend) per customer

   o Applied K-Means (k=3) on log-transformed RFM features

4. **Strategic Recommendations**

   o Inventory alignment, targeted retention, dynamic pricing, and geo-targeted marketing

---

5. Key Findings

5.1 Temporal Trends

- **Hourly:** Peaks at midday and early evening

- **Weekday vs. Weekend:** Weekdays drive ~2.5× more purchases than weekends

- **Monthly:** Holiday season (Nov–Dec) surges; Father's Day month (June) sees a bump in personal-care

- **Yearly:** Growth from 2018 to 2021, followed by a sharp decline in 2023
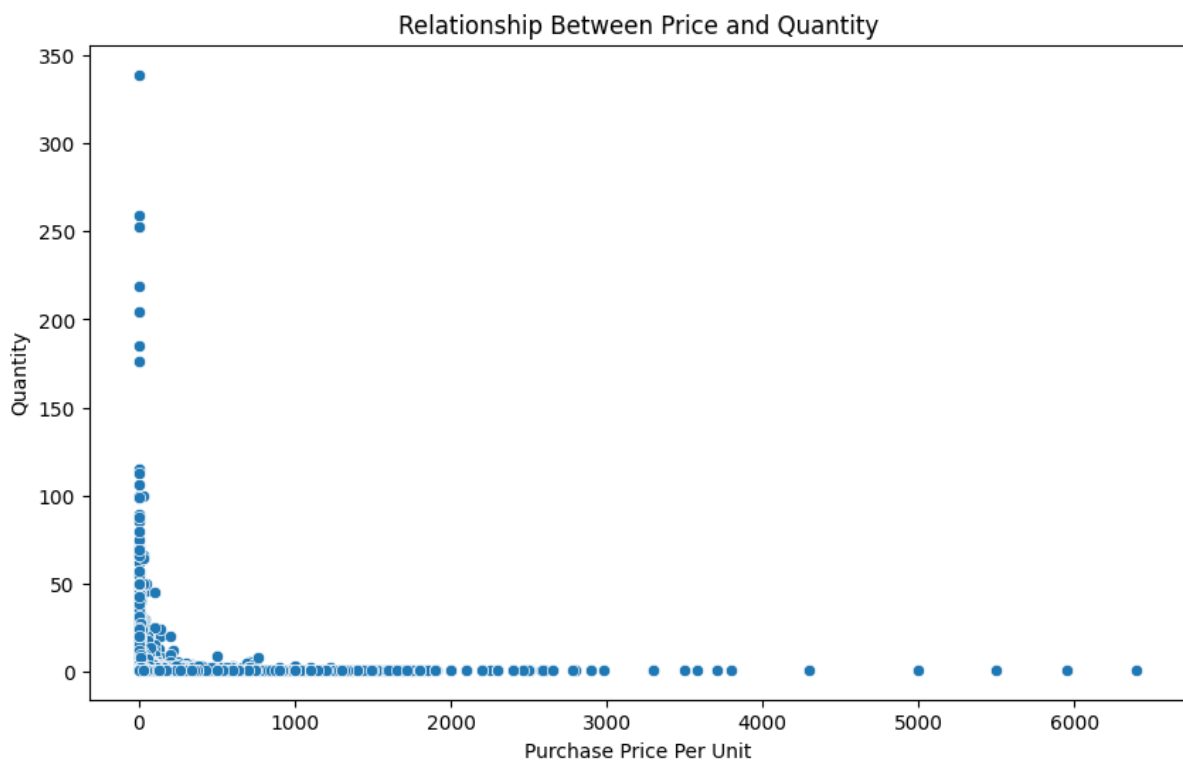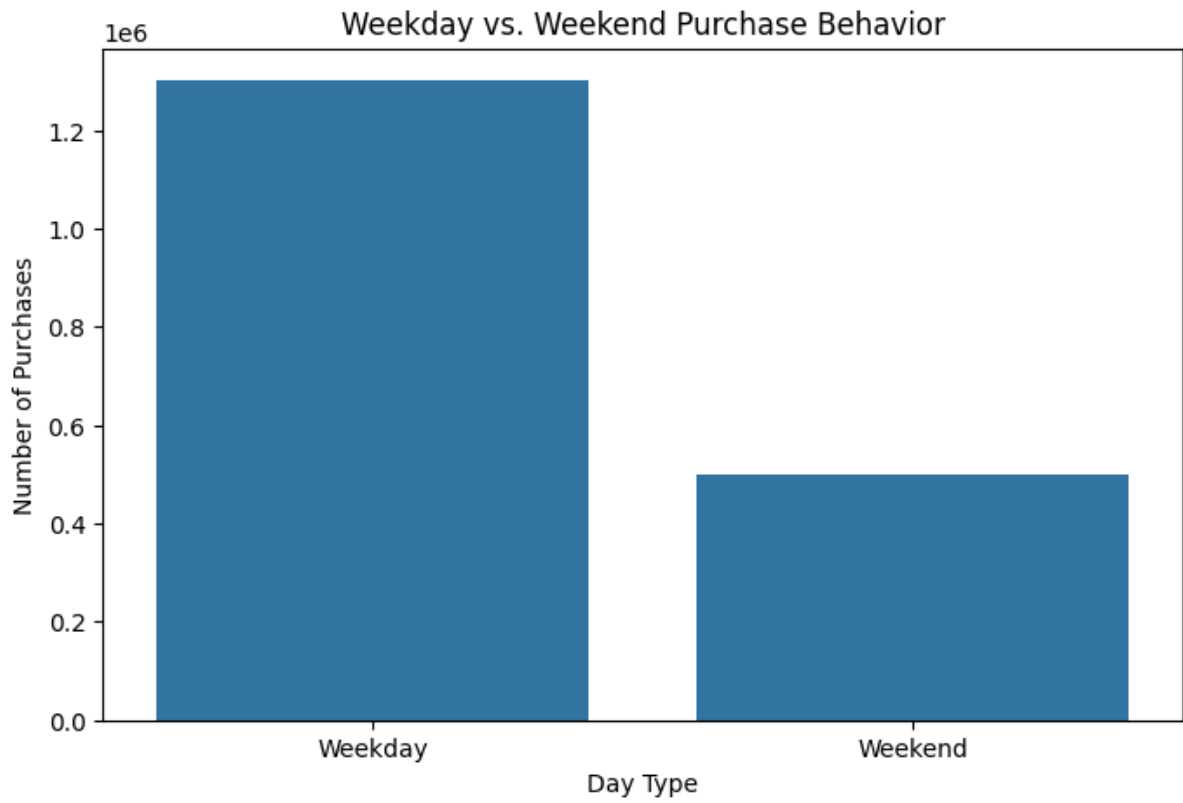
Yearly Purchase Trends
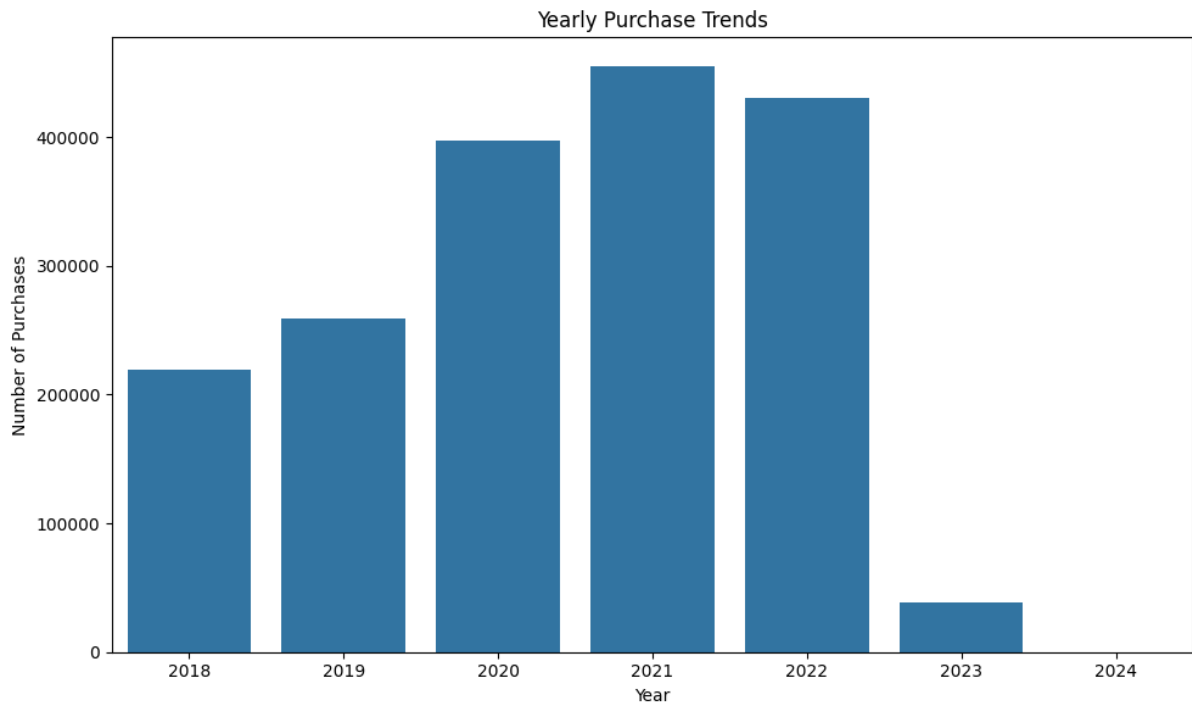
5.2 Demographic Insights

- **Age:** 16–34 bracket accounts for >60% of purchases

- **Gender:** Females slightly outpace males; non-binary and prefer-not-to-say segments are meaningful

- **Income:** Highest volumes in $50K–$99K range; spending tapers at both extremes
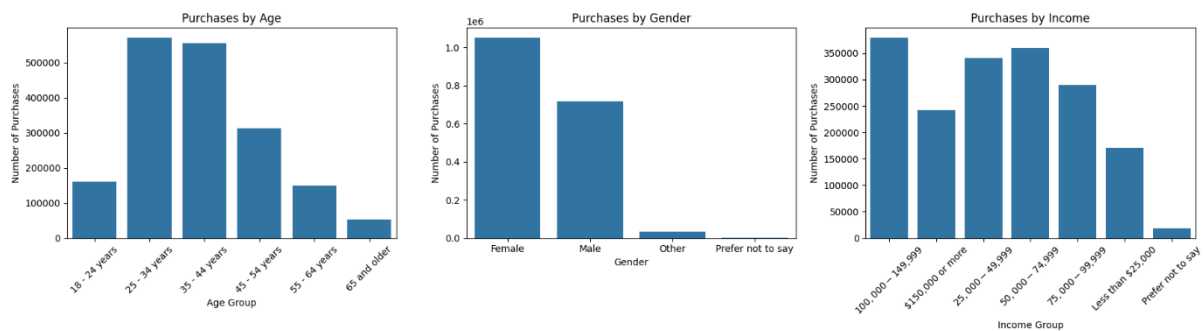
5.3 Geographic Distribution

- **Top States by Revenue:** California, Texas, Florida, New York (>40% combined)

- **High Per-Capita Spend:** Washington and Massachusetts

- **Data Quality Note:** "Unknown" addresses represent a non-trivial share

Weekday vs. Weekend Purchase Behavior

Relationship Between Price and Quantity

Total Revenue by State
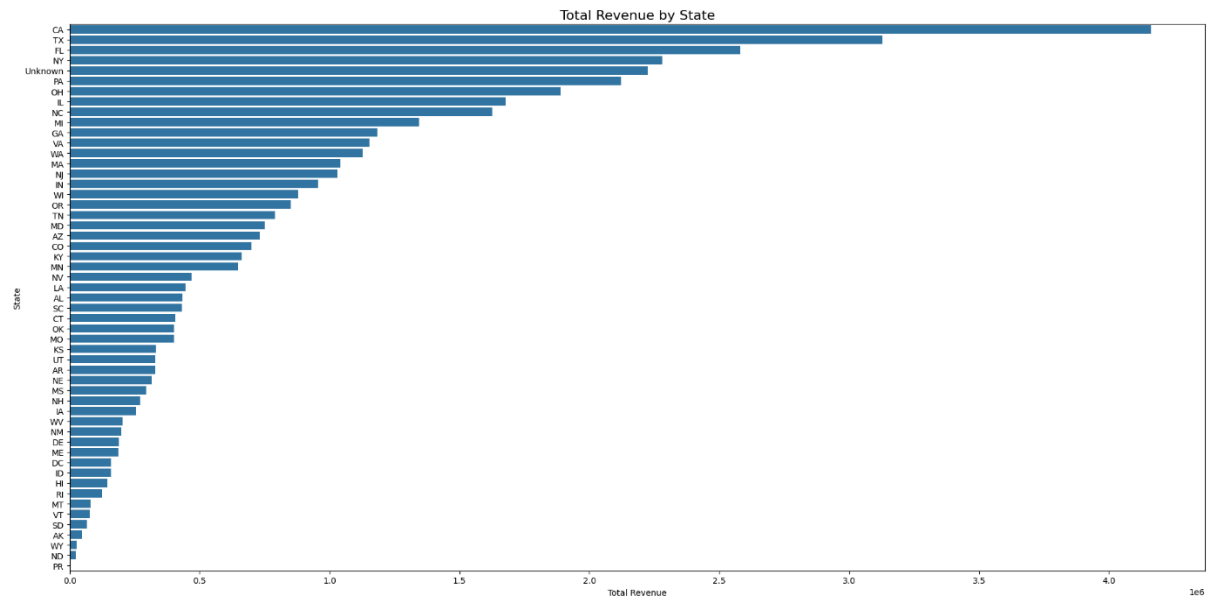
5.4 Product & Basket Analysis

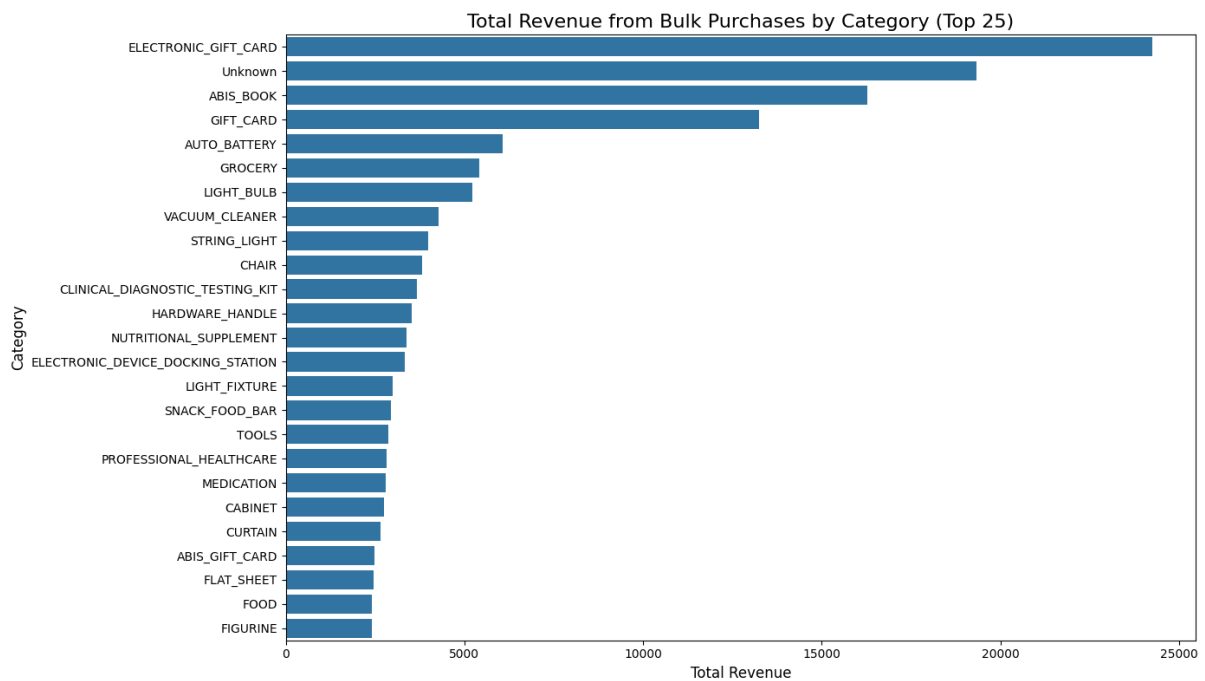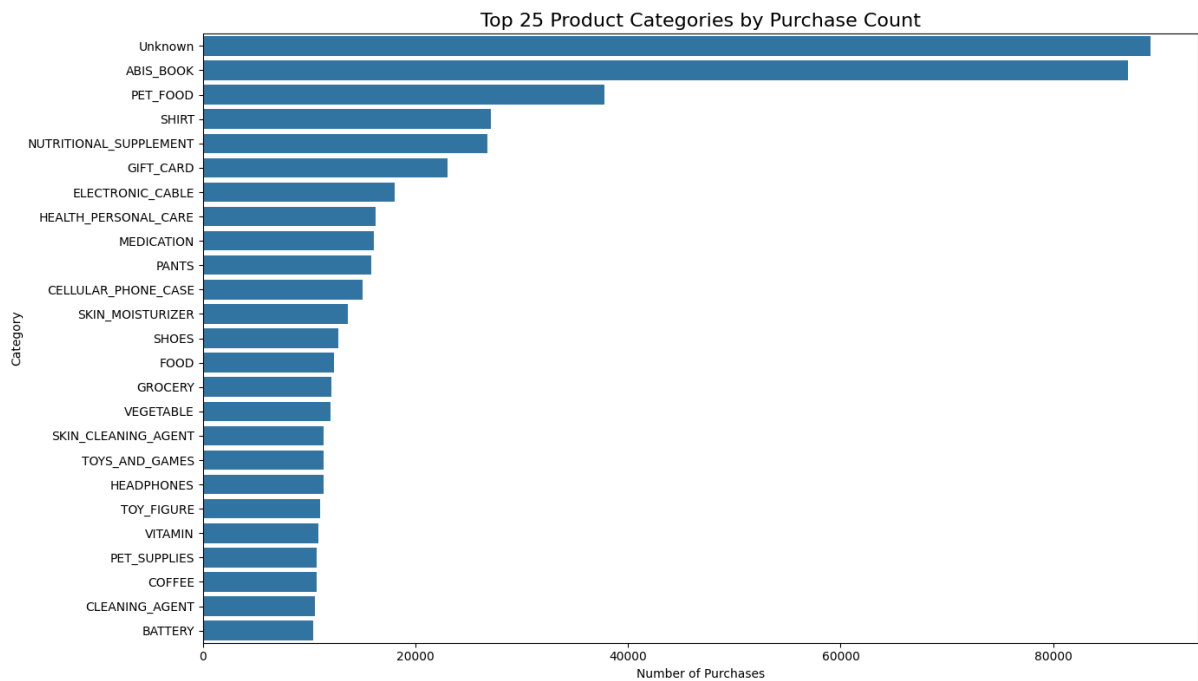- **Top SKUs:** SanDisk Ultra memory cards; Betron earphones; Amazon gift cards

- **Category Leaders:** Books, pet food, gift cards, electronic cables

- **Bulk Purchases:** Consumables and gift cards dominate mid-week bulk orders (>5 units)

- **Market Basket:** Strong co-occurrence among cables, memory cards, earphones—ideal for "Frequently Bought Together"

Top 20 Frequently Purchased Product Pairs

Top 25 Product Categories by Purchase Count

Category	Number of Purchases
Unknown	~88000
ABIS_BOOK	~86000
PET_FOOD	~37000
SHIRT	~27000
NUTRITIONAL_SUPPLEMENT	~26000
GIFT_CARD	~24000
ELECTRONIC_CABLE	~18000
HEALTH_PERSONAL_CARE	~16000
MEDICATION	~15000
PANTS	~15000
CELLULAR_PHONE_CASE	~13000
SKIN_MOISTURIZER	~12000
SHOES	~11000
FOOD	~11000
GROCERY	~11000
VEGETABLE	~11000
SKIN_CLEANING_AGENT	~10000
TOYS_AND_GAMES	~10000
HEADPHONES	~10000
TOY_FIGURE	~10000
VITAMIN	~10000
PET_SUPPLIES	~10000
COFFEE	~10000
CLEANING_AGENT	~10000
BATTERY	~9000

Top 25 Product Categories by Purchase Count — bar chart

Total Revenue from Bulk Purchases by Category (Top 25)

Total Revenue from Bulk Purchases by Category (Top 25) — bar chart showing categories including ELECTRONIC_GIFT_CARD, Unknown, ABIS_BOOK, GIFT_CARD, AUTO_BATTERY, GROCERY, LIGHT_BULB, VACUUM_CLEANER, STRING_LIGHT, CHAIR, CLINICAL_DIAGNOSTIC_TESTING_KIT, HARDWARE_HANDLE, NUTRITIONAL_SUPPLEMENT, ELECTRONIC_DEVICE_DOCKING_STATION, LIGHT_FIXTURE, SNACK_FOOD_BAR, TOOLS, PROFESSIONAL_HEALTHCARE, MEDICATION, CABINET, CURTAIN, ABIS_GIFT_CARD, FLAT_SHEET, FOOD, FIGURINE
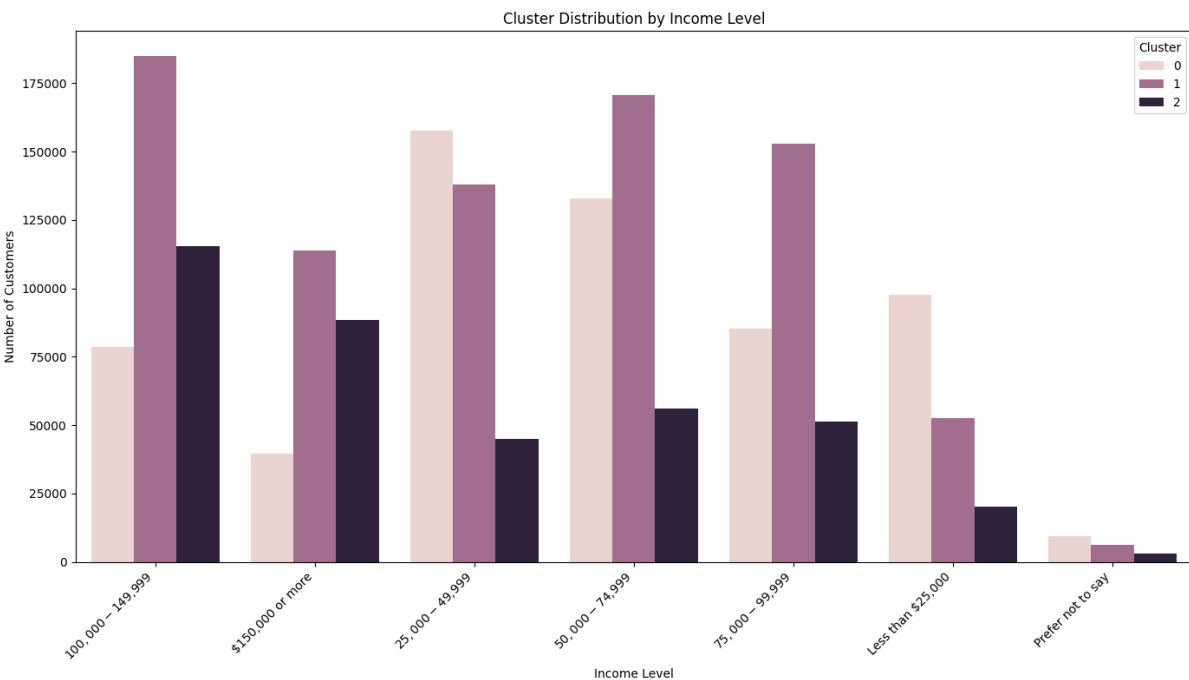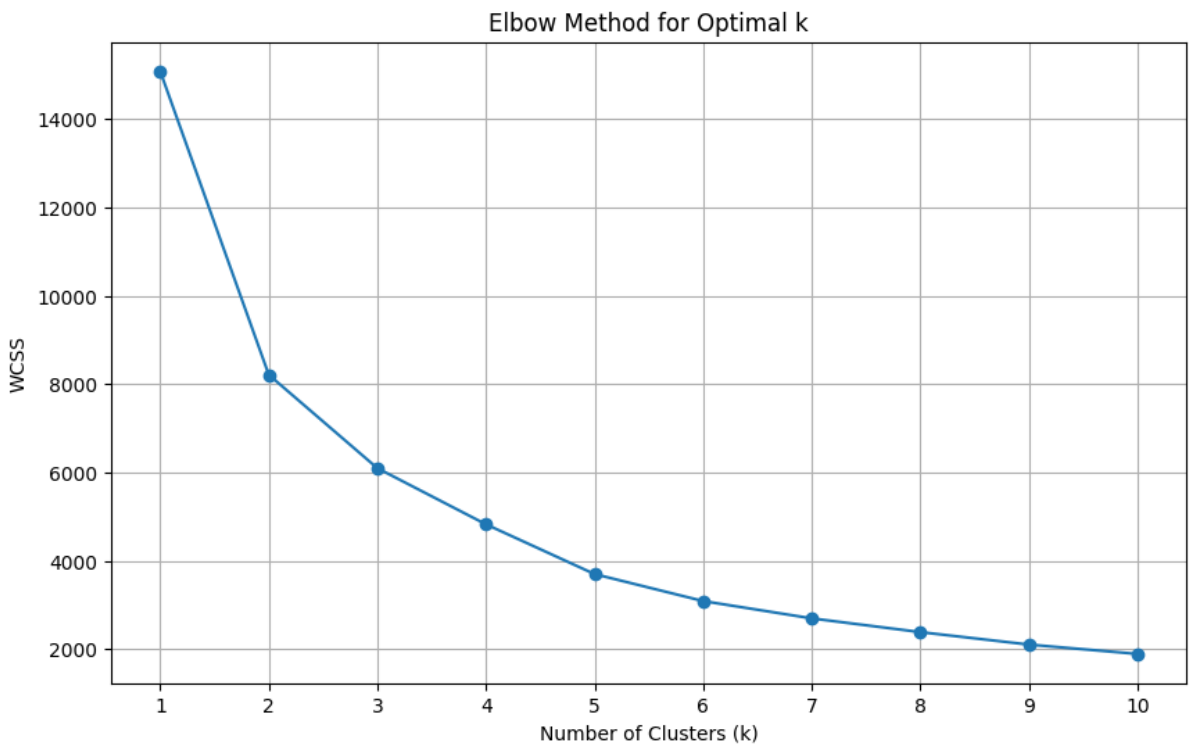
---

6. RFM Segmentation & Behavioral Profiles

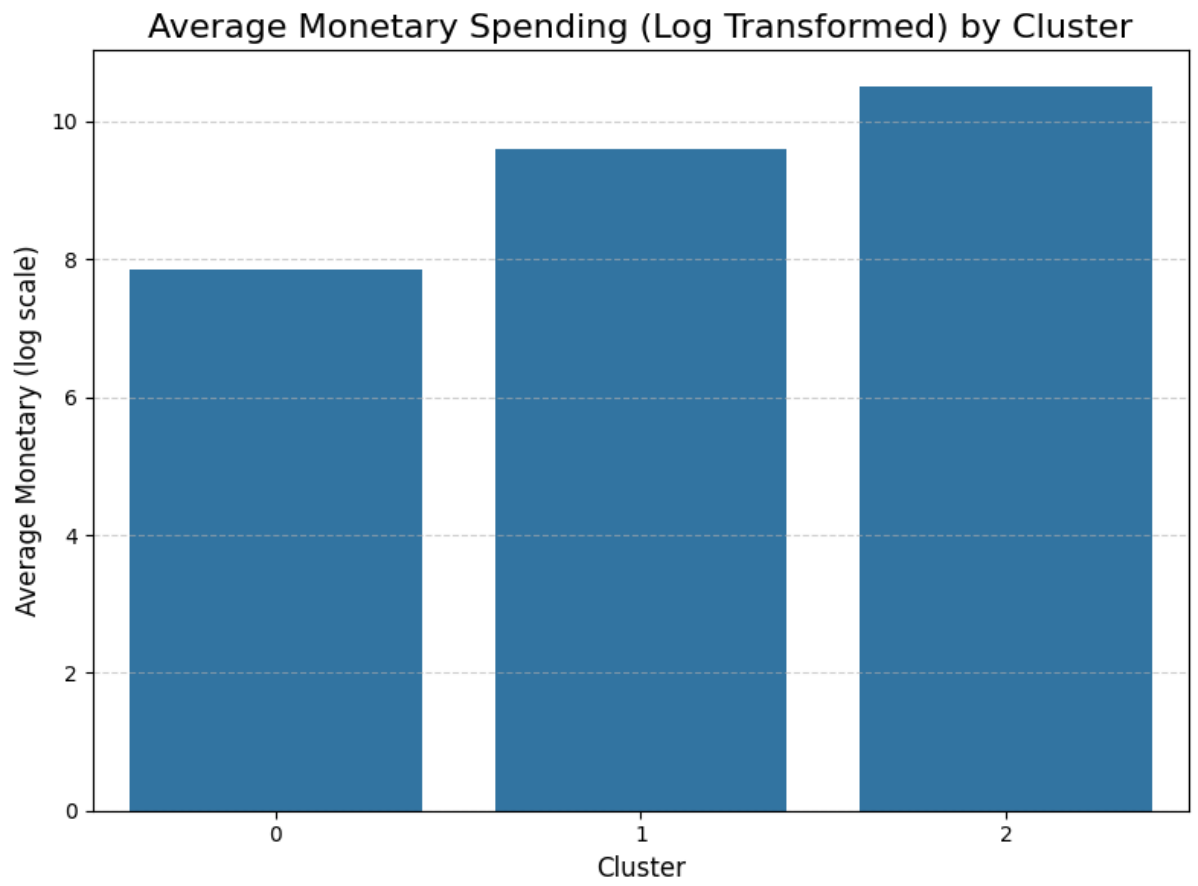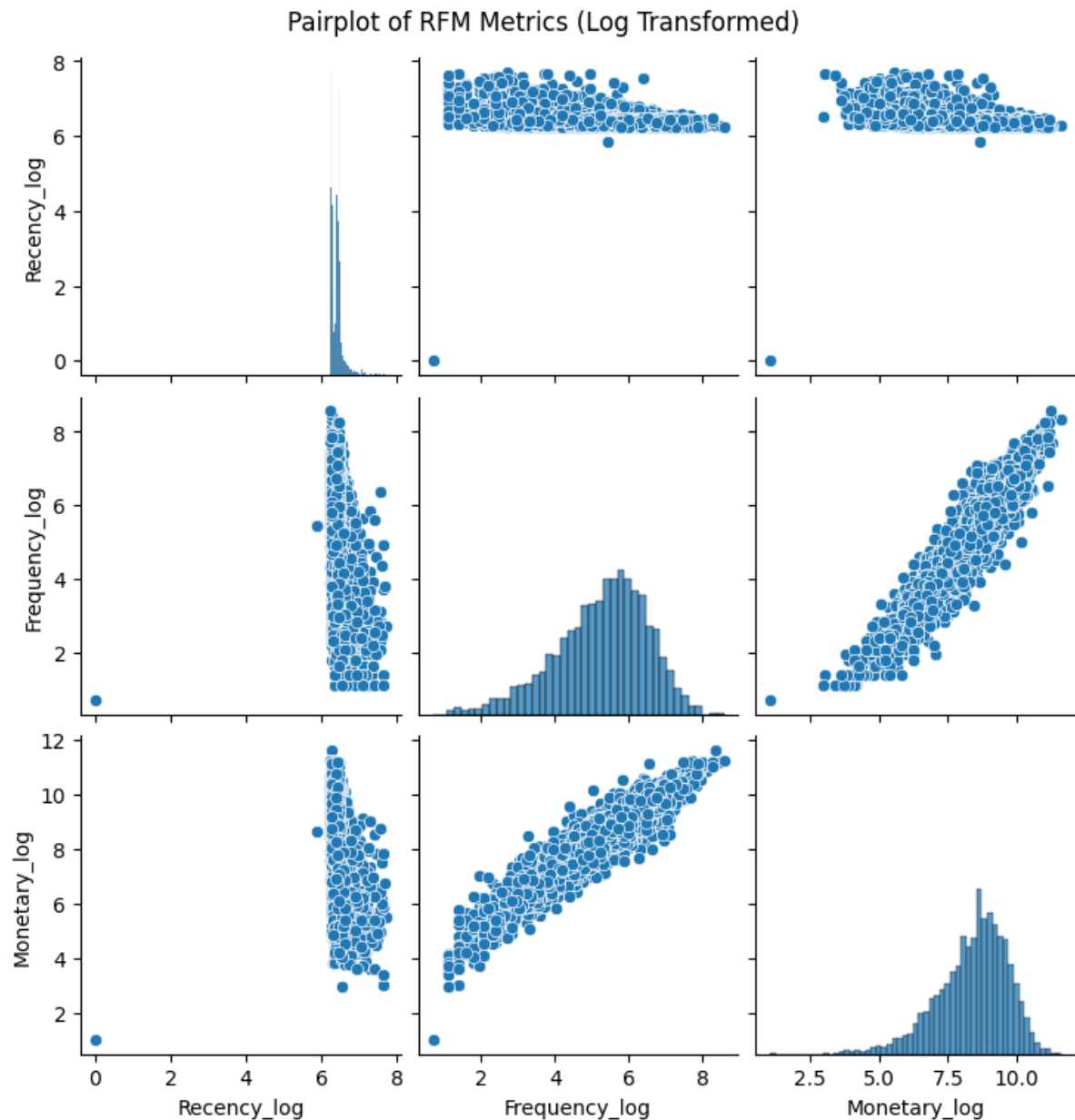| Cluster	Recency	Frequency	Monetary	Business Label
0	High	Low	Low	At-Risk
1	Medium	Medium	Medium	Moderate

Cluster Recency Frequency Monetary Business Label

2   Low   High   High   Champions

- **Champions (Cluster 2):** Recent, frequent, high spenders

- **Moderates (Cluster 1):** Consistent, mid-value buyers

- **At-Risk (Cluster 0):** Dormant or low engagement

Elbow Method for Optimal k

Cluster Distribution by Income Level

Average Monetary Spending (Log Transformed) by Cluster

Pairplot of RFM Metrics (Log Transformed)

---

7. Strategic Recommendations

7.1 Inventory Management

- **Pre-Season Stocking:** Boost electronics accessories inventory by 30% in October

- **Father's Day Bundles:** Assemble grooming kits and grooming-related consumables in May

- **SKU Tiering:**

    o A-SKUs (top 10 by revenue): expedited reorders, premium shelf locations

    o B-SKUs (next 15): weekly forecast reviews

    o C-SKUs: consolidated slow-moving storage

- **Vendor Partnerships:** Seasonal volume agreements and just-in-time restocks for high-turn categories

Monthly Revenue Trends for Top 25 Product Categories
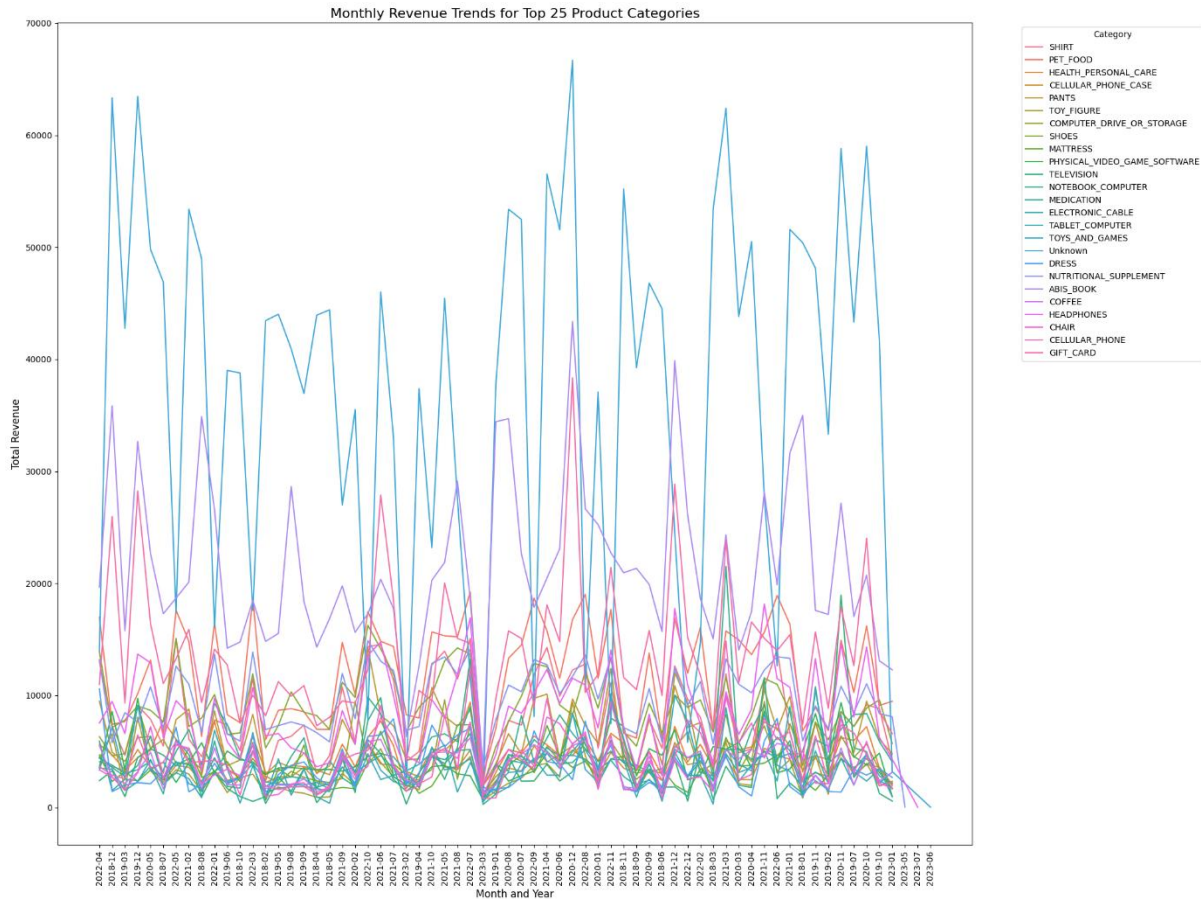
7.2 Customer Retention

- **Champions:** Enroll in VIP loyalty with exclusive offers and early-access sales

- **Moderates:** Send mid-cycle "frequent-buyer" coupons 6 weeks post-purchase

- **At-Risk:** Trigger "We Miss You" reminders 60 days after last order with personalized discounts

- **Subscription Nudges:** Automated refill prompts for consumables (e.g., shaving cream every 30 days)

7.3 Personalized Pricing

- **Bulk Discounts:** 5+ units → 10% off; 10+ units → 15% off to drive institutional orders

- **Surge Pricing Alerts:** Time-limited offers on Fri–Sat for high-demand SKUs

- **Cluster-Specific Flash Sales:**

- o   Champions: fullmargin bundles + value-add (free shipping)

- o   Moderates/At-Risk: 5–10% off in abandoned-cart reminders

- **A/B Testing:** Controlled experiments on staple SKU price points to gauge elasticity

7.4 Regional Marketing Allocation

- **Budget Split:**

  - o   50% digital spend focused on CA, TX, NY

  - o   20% targeting high-spend, low-volume states (WA, MA) with premium product ads

- **Micro-Fulfillment:** Establish same-day hubs in LA, TX metros, and NYC to support rapid delivery promises

- **Localized Campaigns:** Weekend "local flash sales" and geo-targeted creatives featuring top SKUs per state

---

8. Conclusion

By aligning inventory with seasonal peaks, tailoring retention journeys by RFM segment, deploying dynamic pricing tactics, and concentrating marketing in high-value regions, the business can sustainably grow revenue and deepen customer loyalty. This data-driven approach lays the foundation for scalable operations and targeted, high-ROI campaigns.