Big Data Assignment – Customer Behavior Analysis ASSIGNMENT

By JALPA CHAUHAN
Mark Fernandes
Eshwar Ramesh

Assignment-based Content

1. **Data Preparation**
2. **Data Cleaning**
3. **Exploratory Data Analysis**
4. **Customer Segmentation (RFM Analysis) and Business Insights**
5. **Evaluation and Conclusion**

Problem Statement

- Analyze large-scale e-commerce transaction and customer survey data using Apache Spark to uncover insights into customer behavior.
- The goal is to identify high-value customers, segment users based on purchasing patterns, and detect trends in product demand and engagement.
- This analysis supports data-driven decisions to boost customer satisfaction, retention, and revenue in a competitive market.

🎯 Business Objective

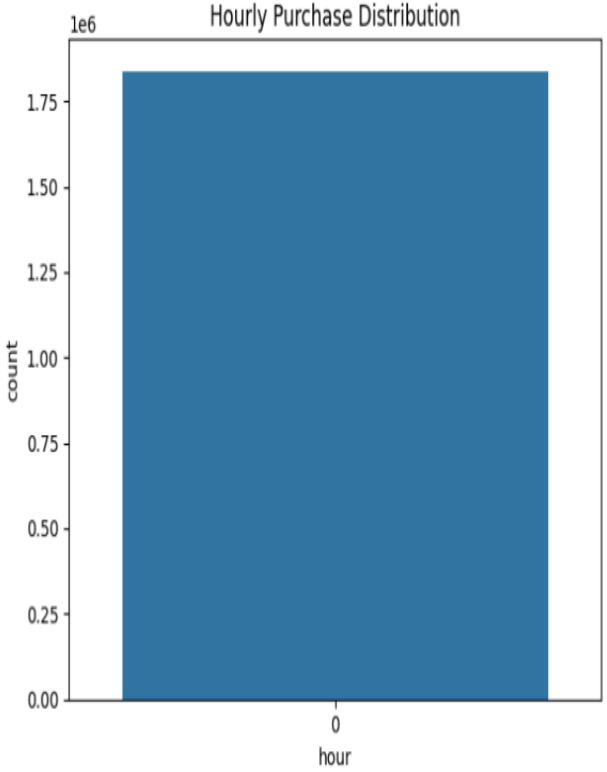
- Understand and analyze e-commerce customer behavior using Apache Spark.
- Leverage big data to:
  - Identify high-value customers
  - Segment users based on purchase patterns
  - Discover trends in product demand and customer engagement
- Drive strategic decisions to improve marketing, personalization, and customer retention.

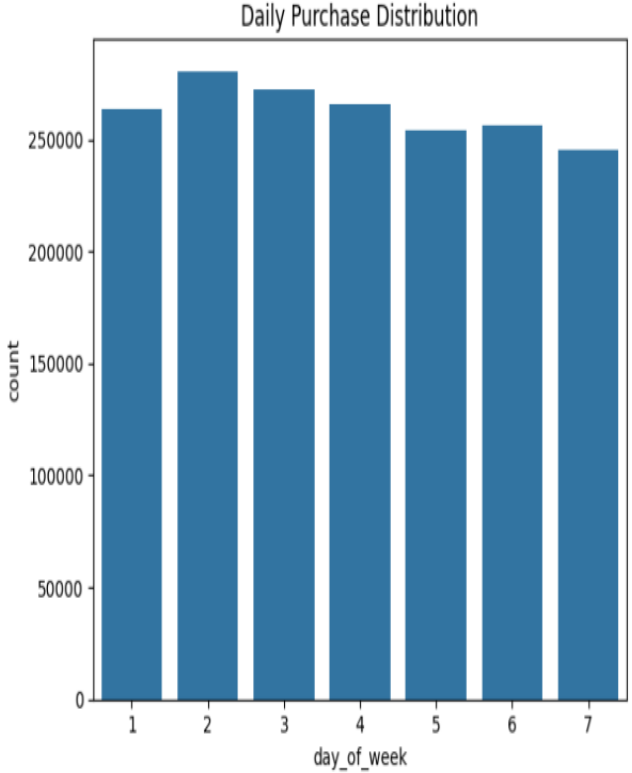
Problem Approach

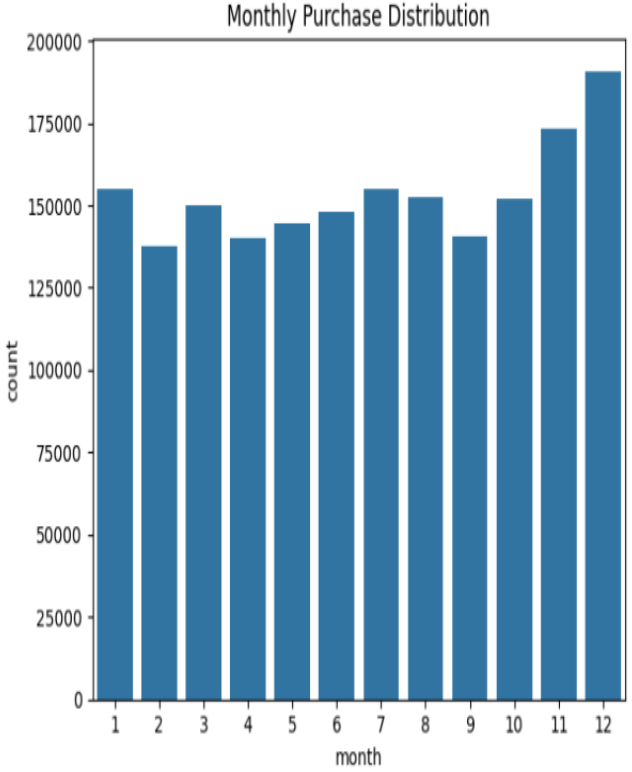
1. Data Preparation
   - Load and integrate transactional, survey, and demographic data.
2. Data Cleaning
   - Handle missing values, remove duplicates, and standardize formats.
3. Exploratory Data Analysis (EDA)
   - Uncover purchase trends, product preferences, and regional patterns.
4. Customer Segmentation (RFM Analysis)
   - Segment customers based on Recency, Frequency, and Monetary value.
   - Identify loyal, high-value, and at-risk customer groups.
5. Business Insights & Evaluation
   - Derive actionable insights to inform marketing and retention strategies.
   - Summarize key findings and propose data-driven recommendations.

🔍 Exploratory Data Analysis – 🖌️ Data Cleaning

➢ 3.1 Analyse purchases by hour, day and month

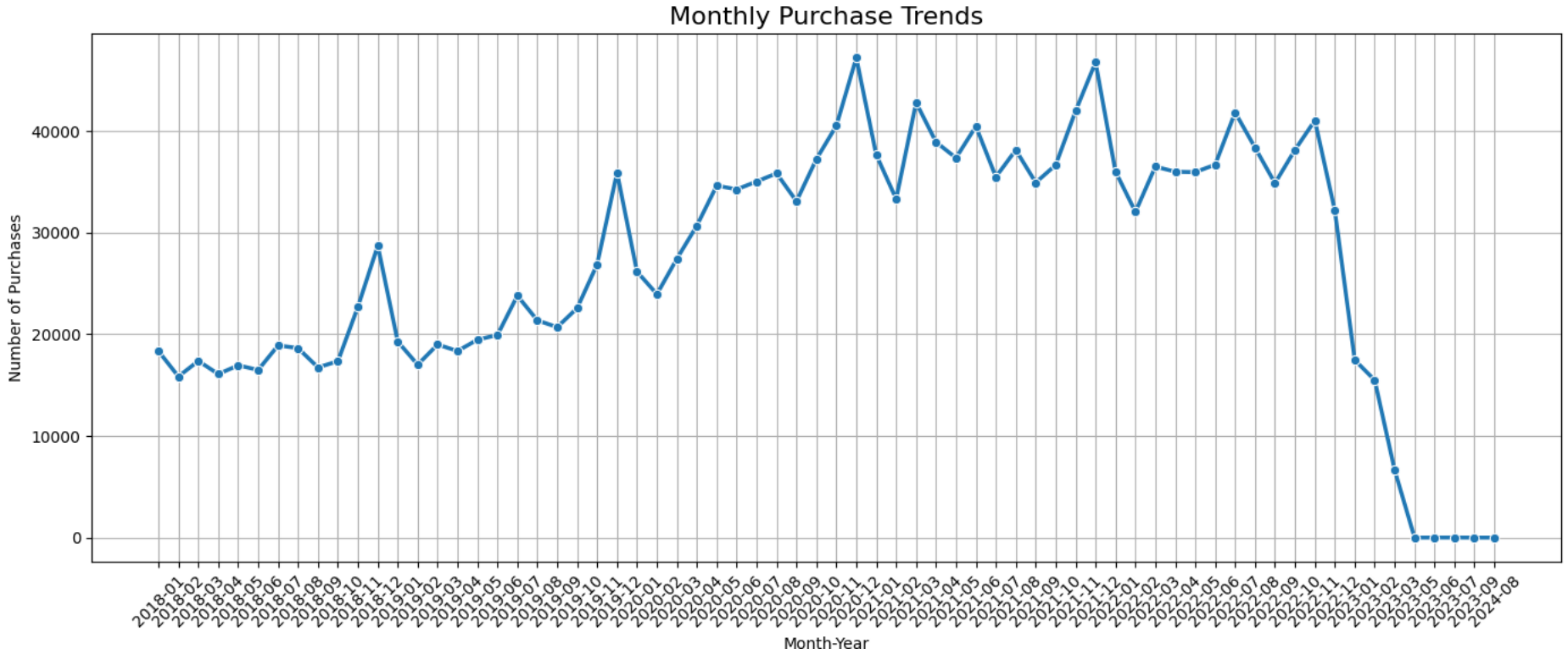






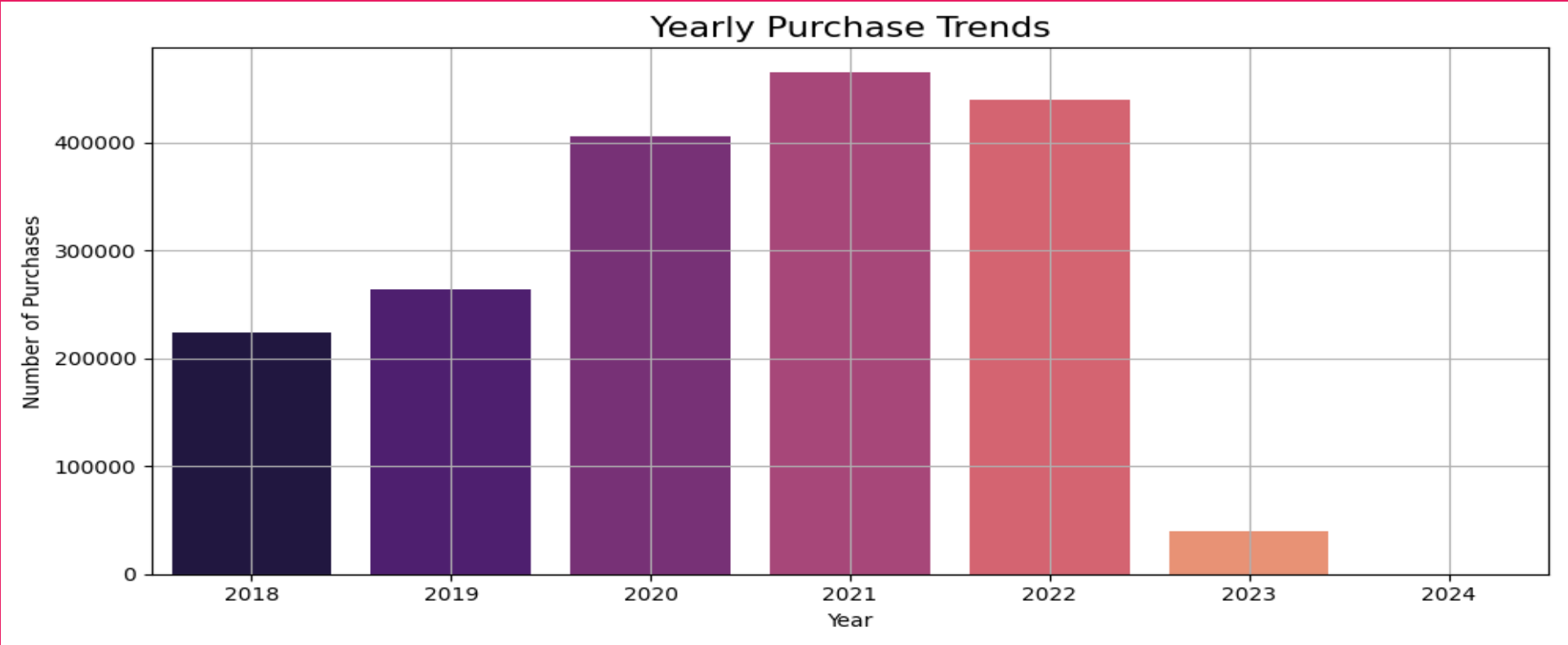
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

- # Monthly Purchase Trends



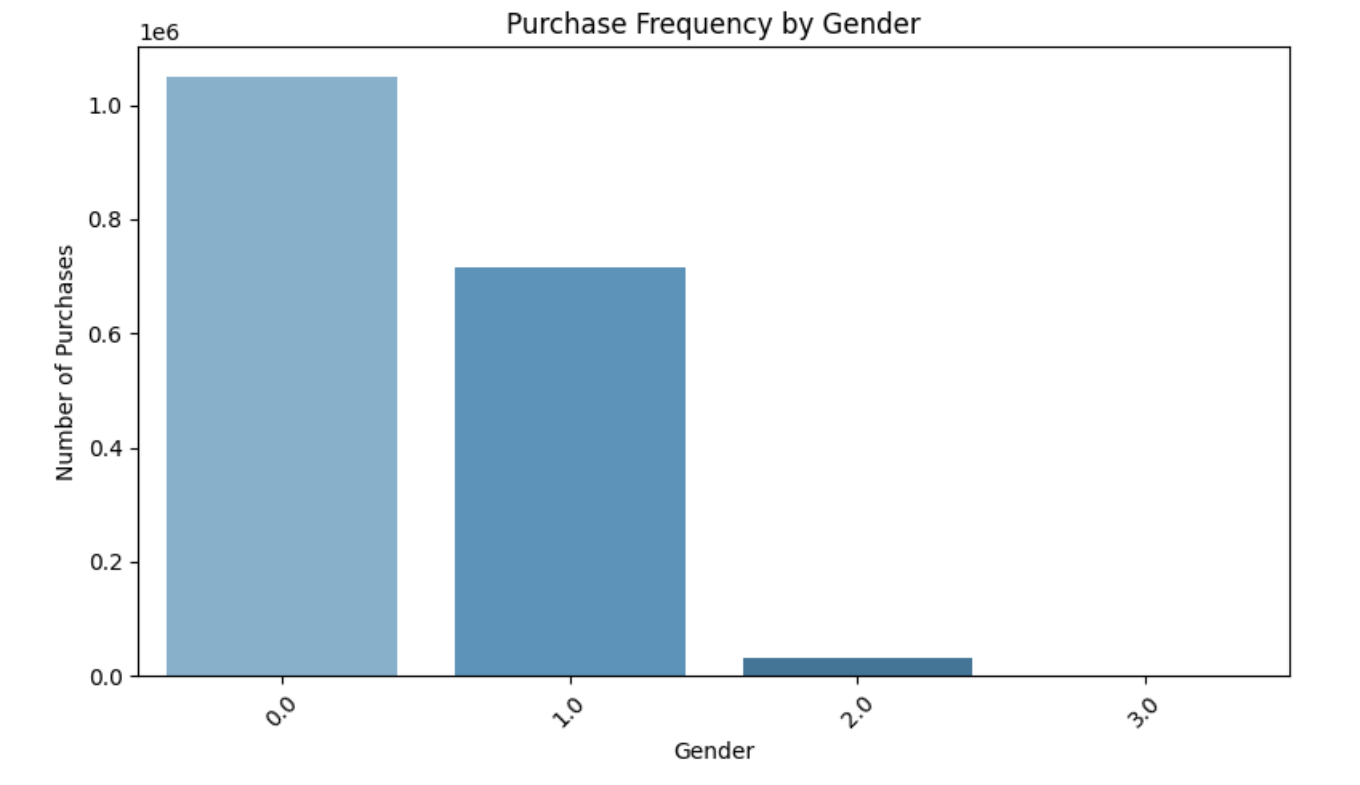
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

- # Yearly Purchase Trends



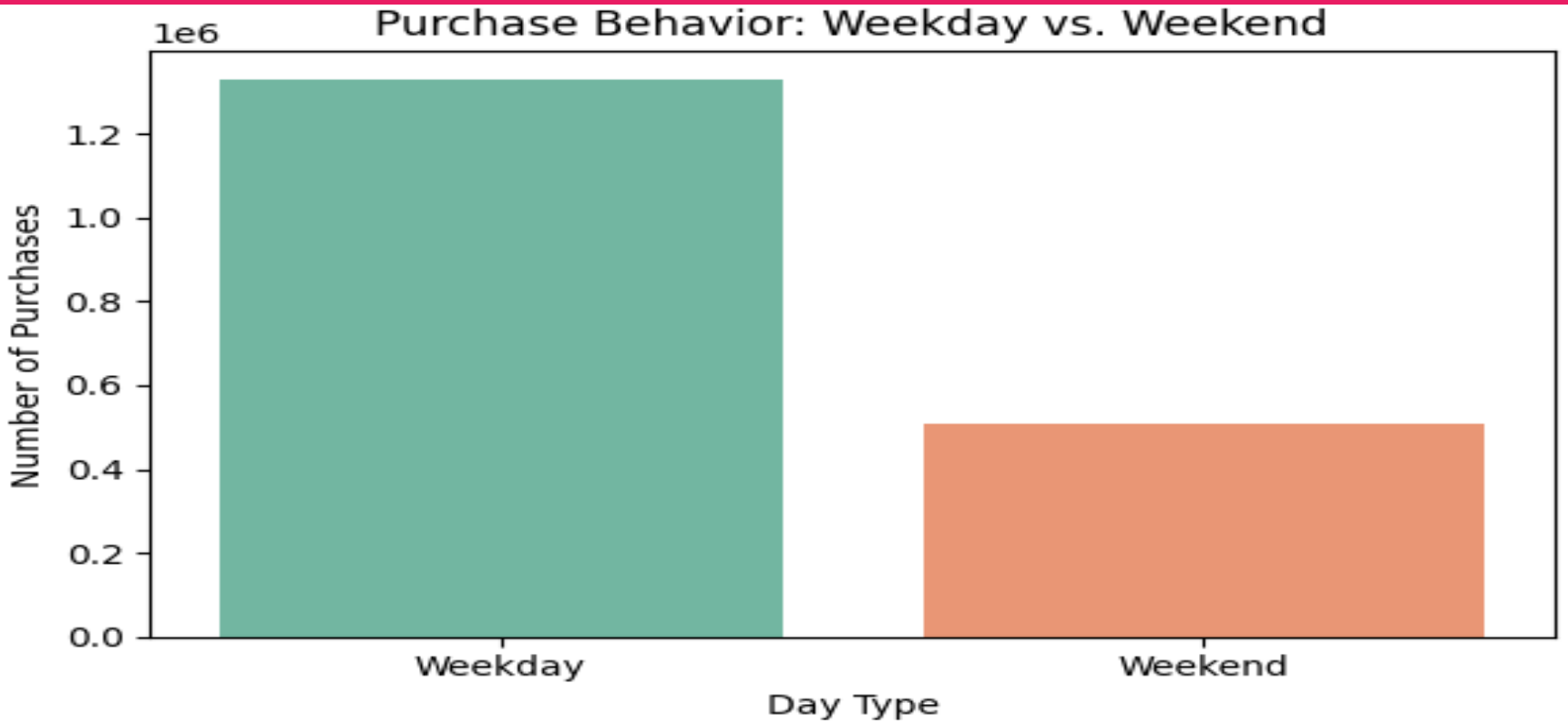
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.2 Customer Demographics vs Purchase Frequency



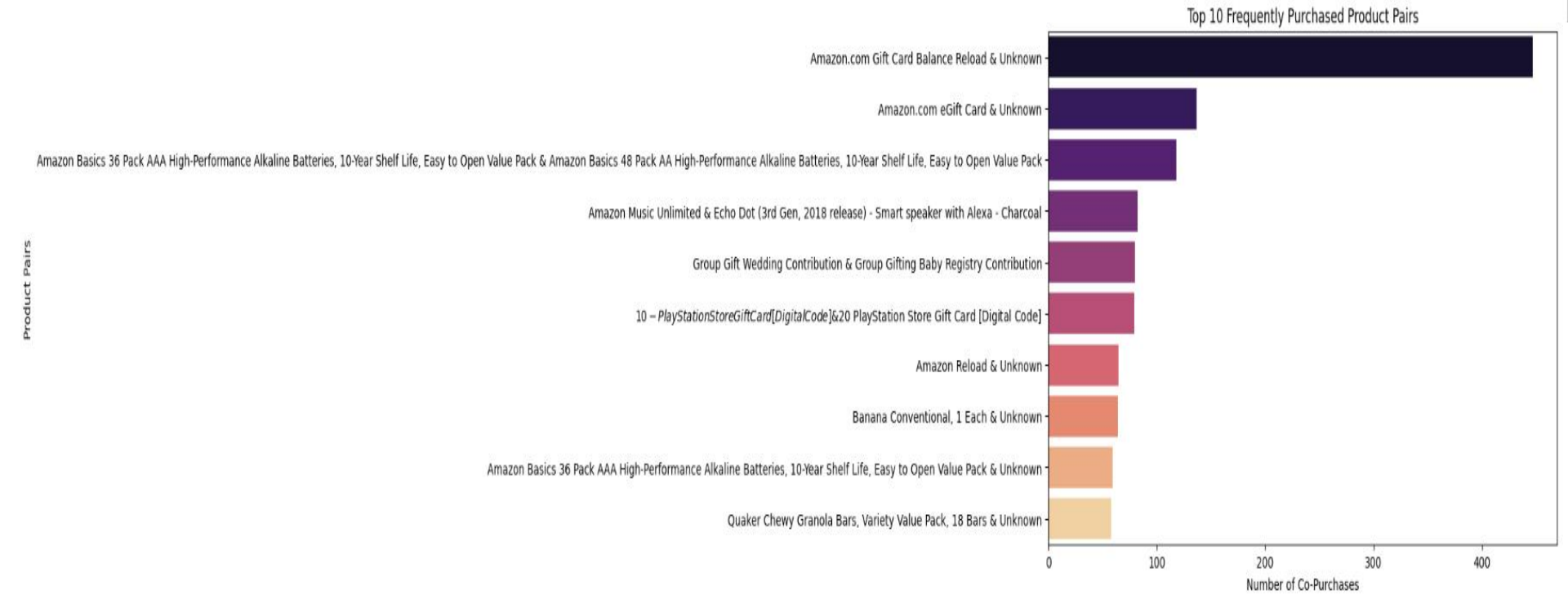
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.3 Purchase behavior weekend vs weekday



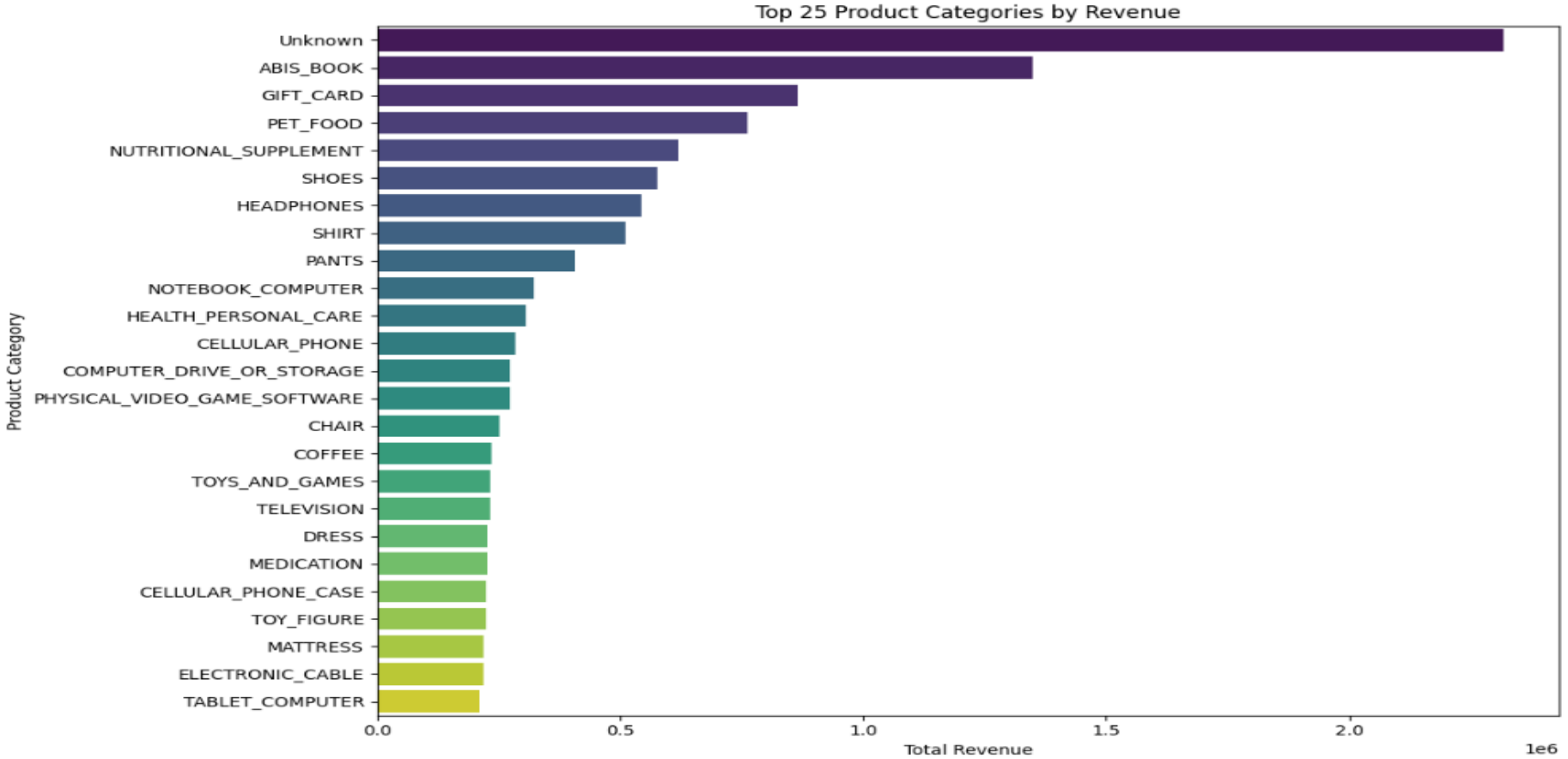
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.4 Frequently purchased product pairs



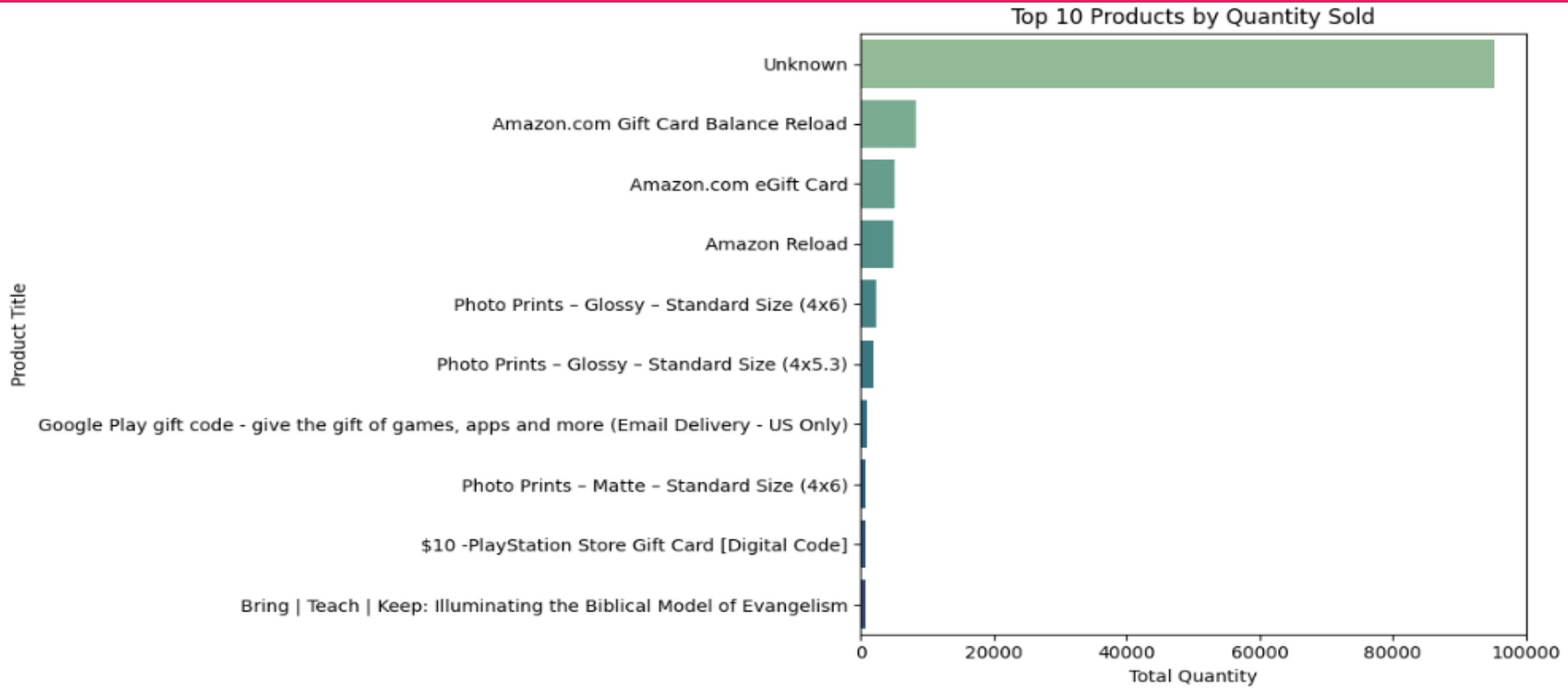
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.5 Examine Product Performance



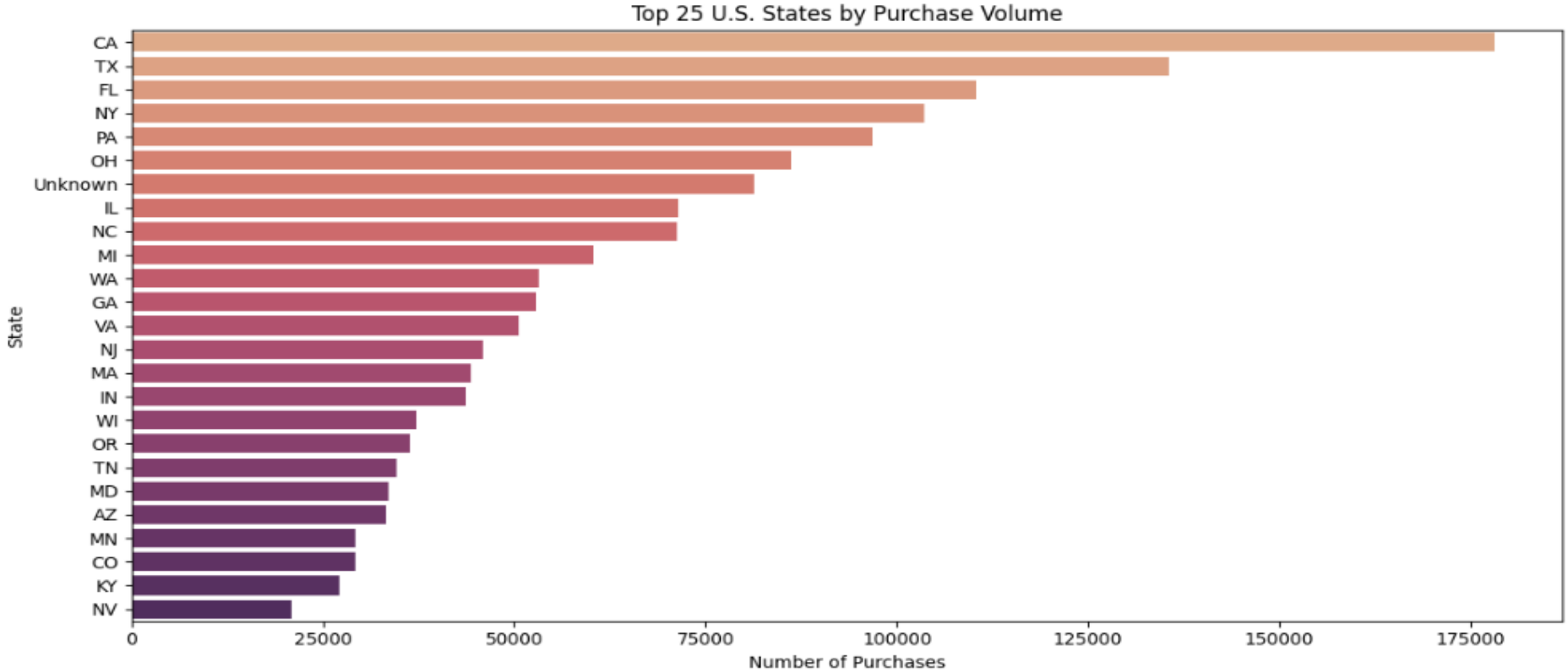
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.6 Top products by quantity



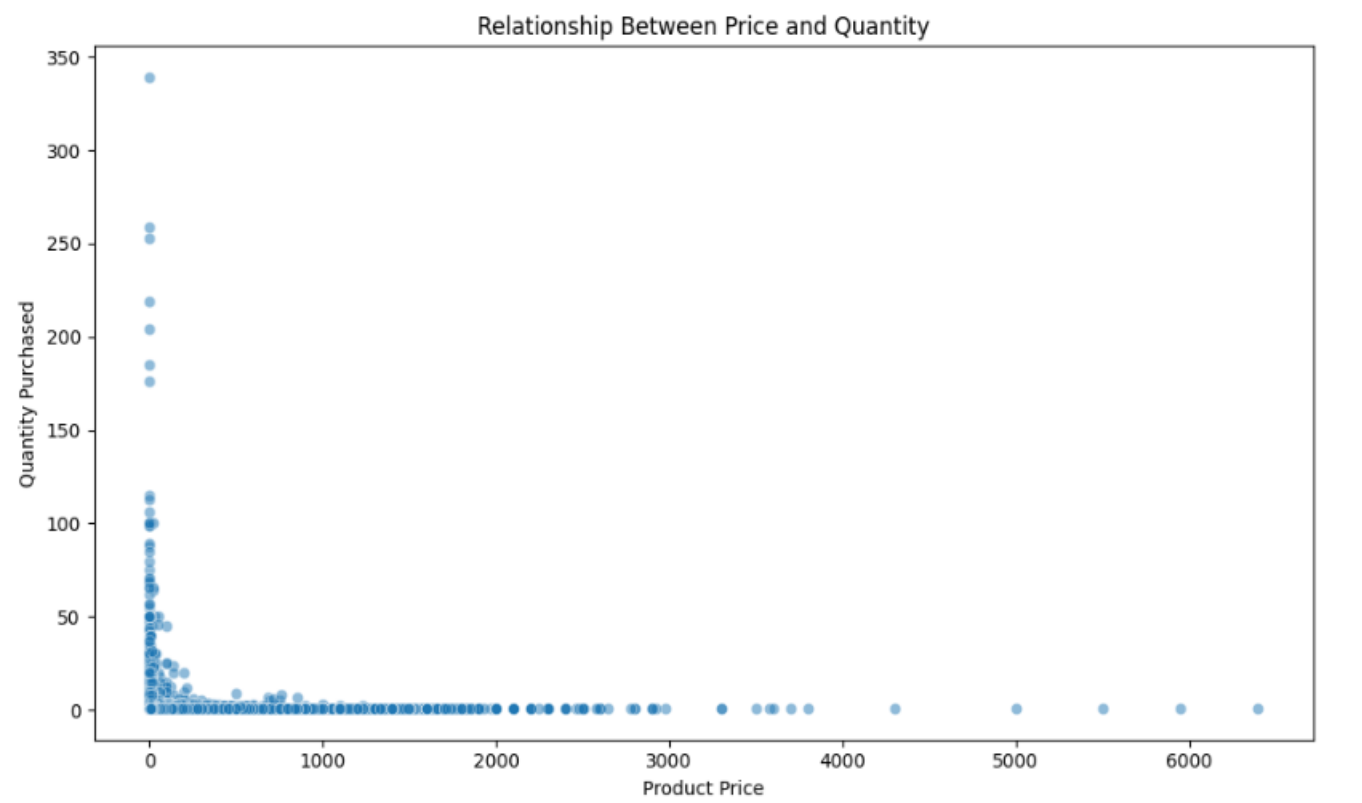
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.7 Distribution of Purchases by State



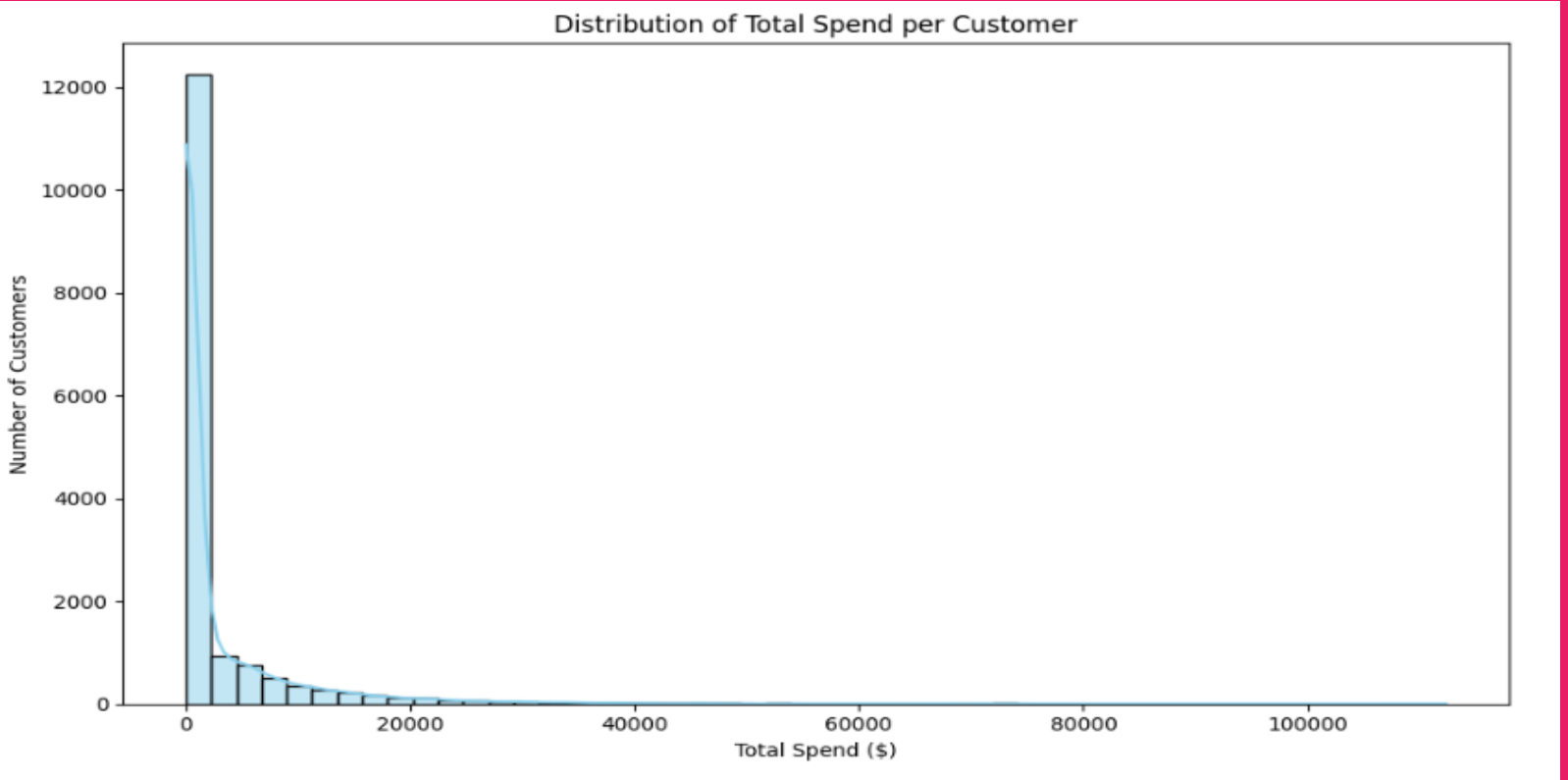
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.8 Price vs Product Quantity



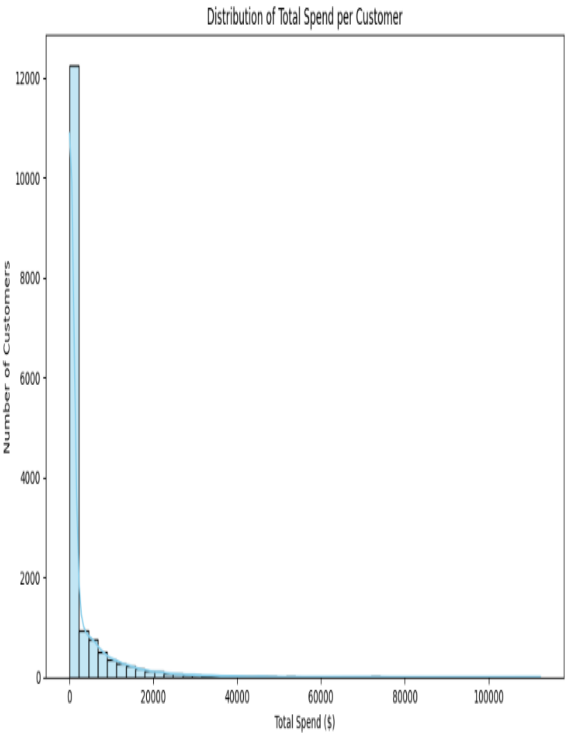
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

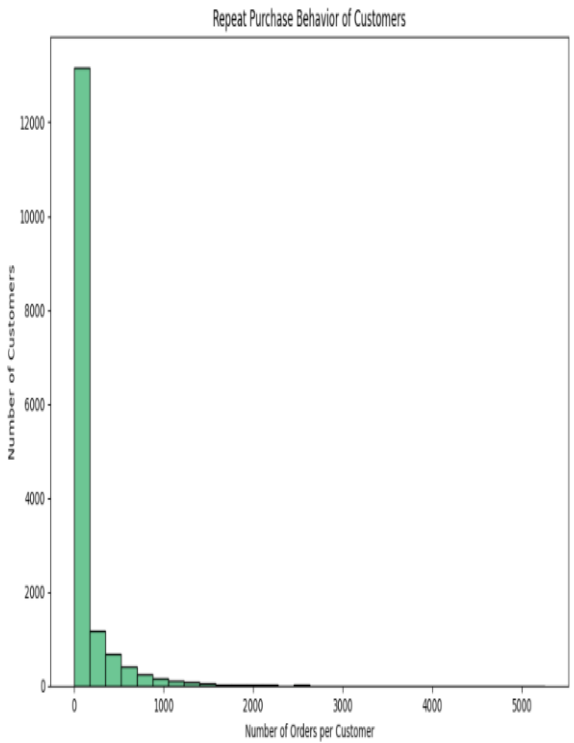
➢ 3.9 Analyse the spending KPIs

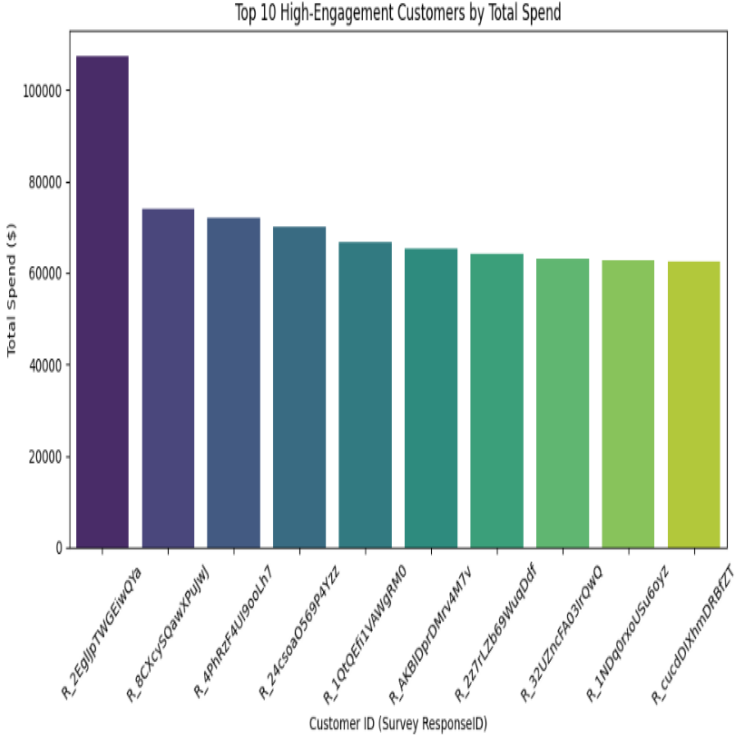


Exploratory Data Analysis – Data Cleaning

- 3.9 Analyse the spending KPIs
  - Analyse the Repeat Purchase Behavior of Customers
  - Analyse the top 10 high-engagement customers

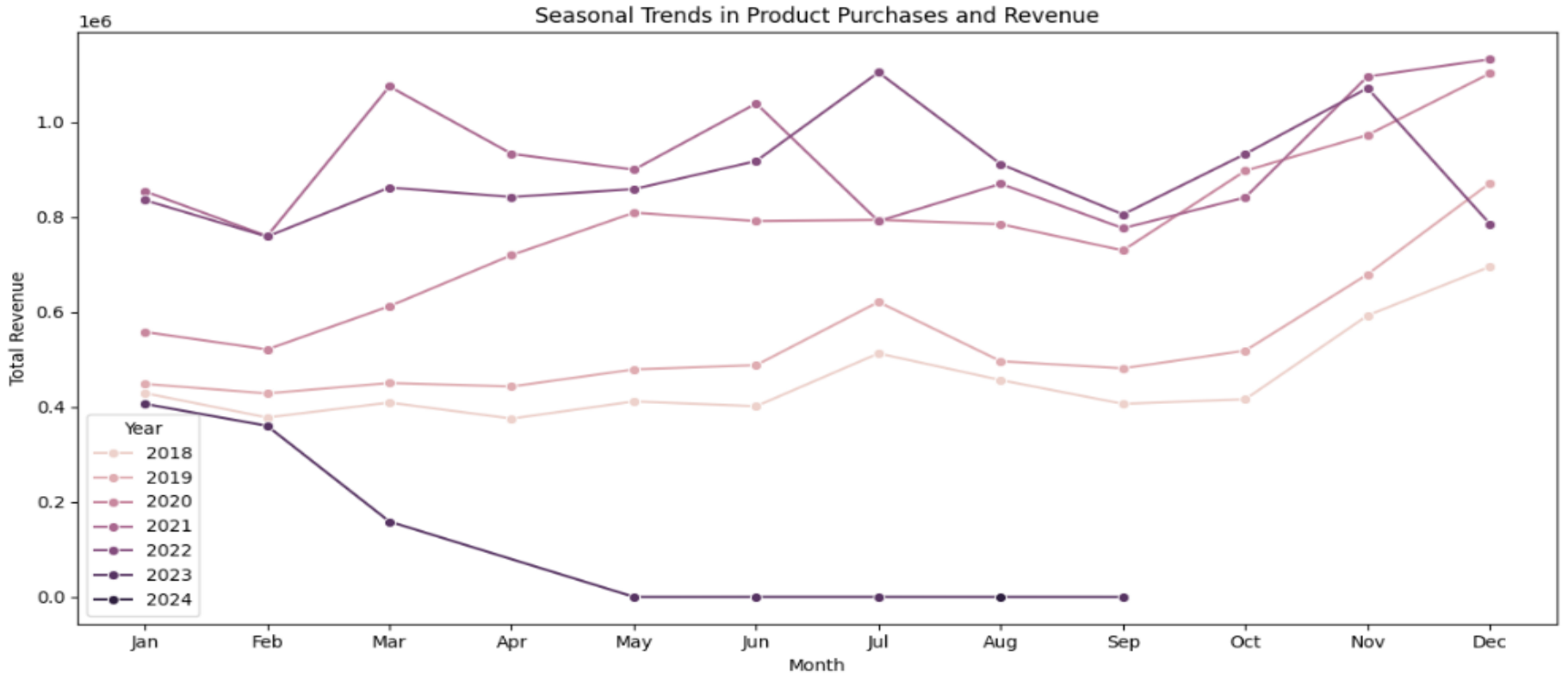






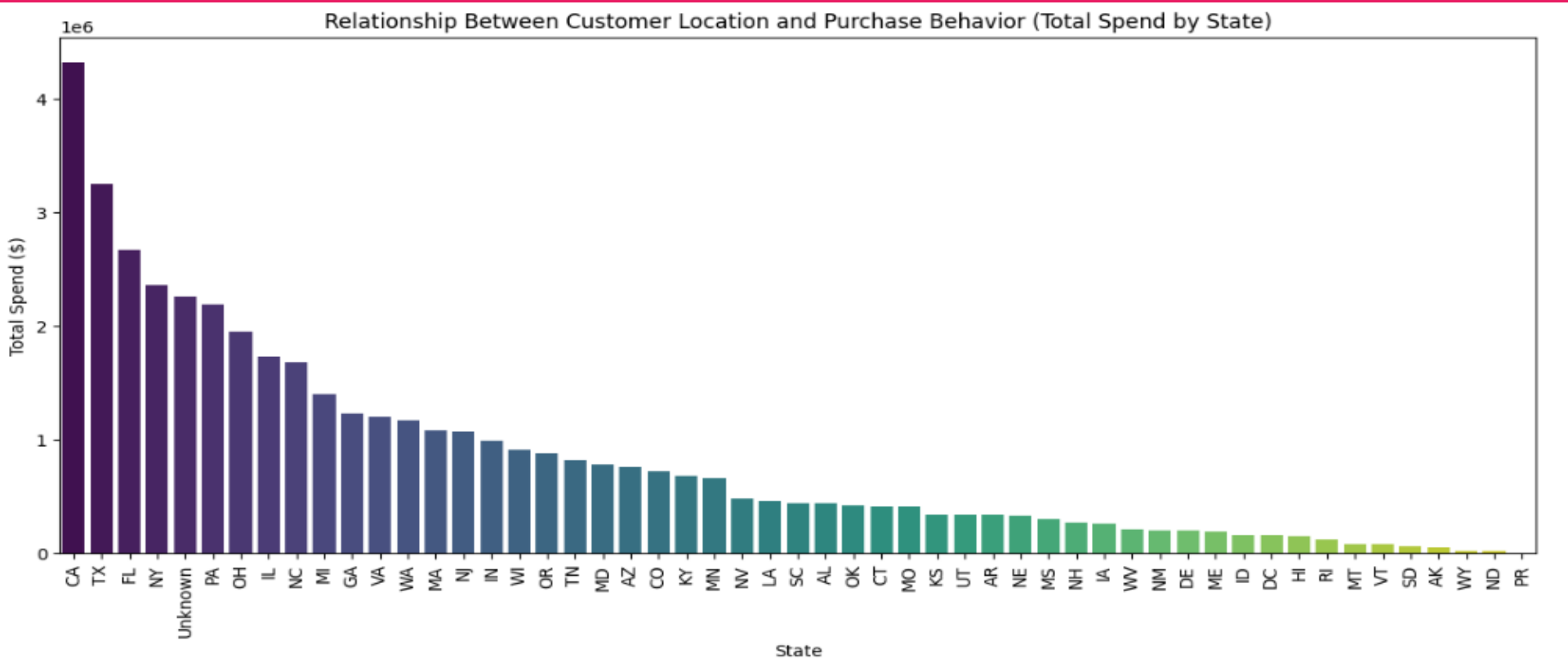
🔍 Exploratory Data Analysis – 🧹 Data Cleaning

➢ 3.10 Seasonal trends in product purchases and their impact on revenues



🔍 Exploratory Data Analysis – 🧹 Data Cleaning

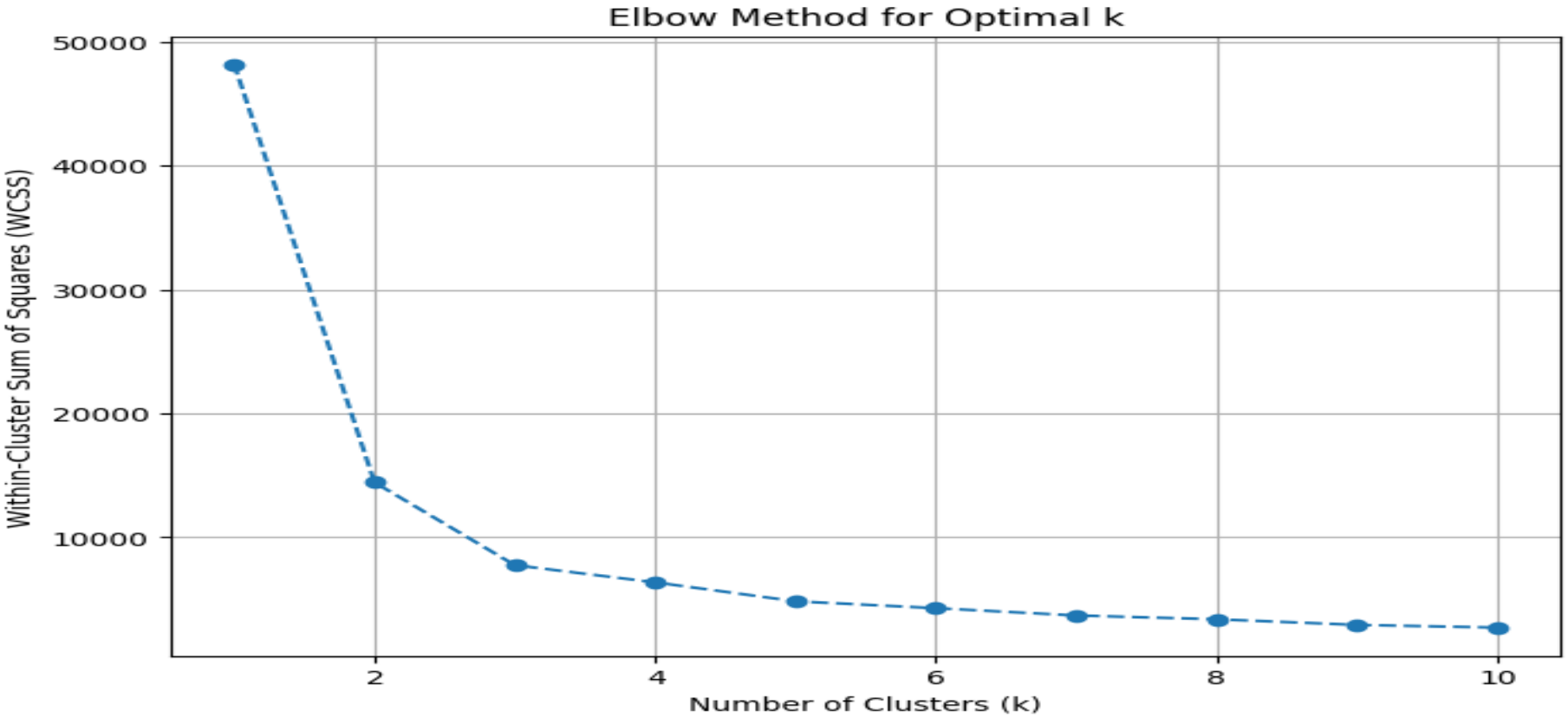
➢ 3.11 Customer location vs purchasing behavior



🧠 Customer Segmentation & Insights – 📊 RFM Analysis

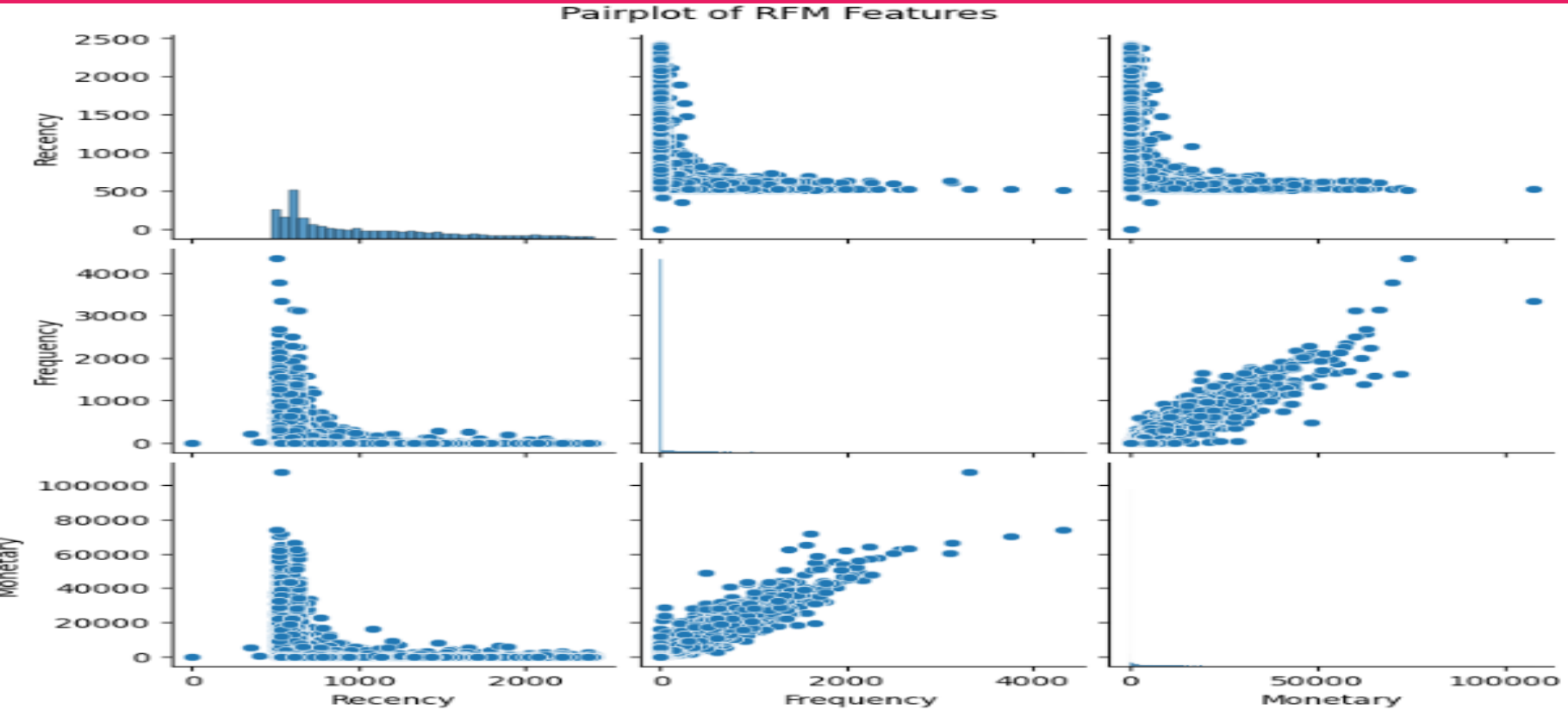
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

➢ 4.1 Perform RFM Analysis



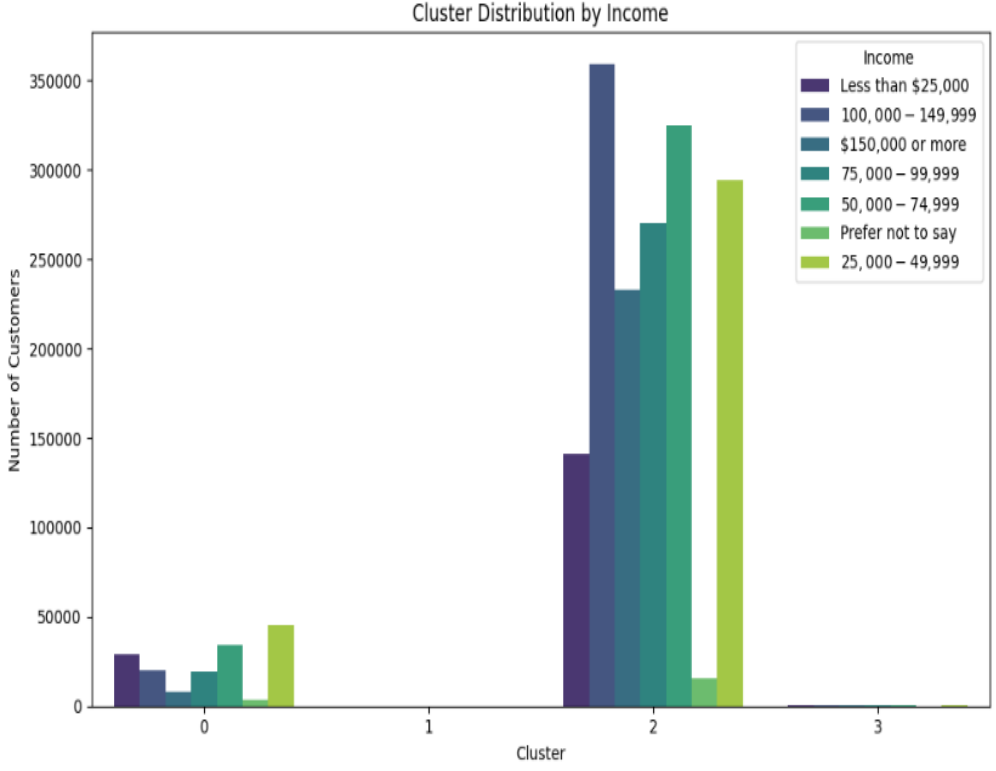
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

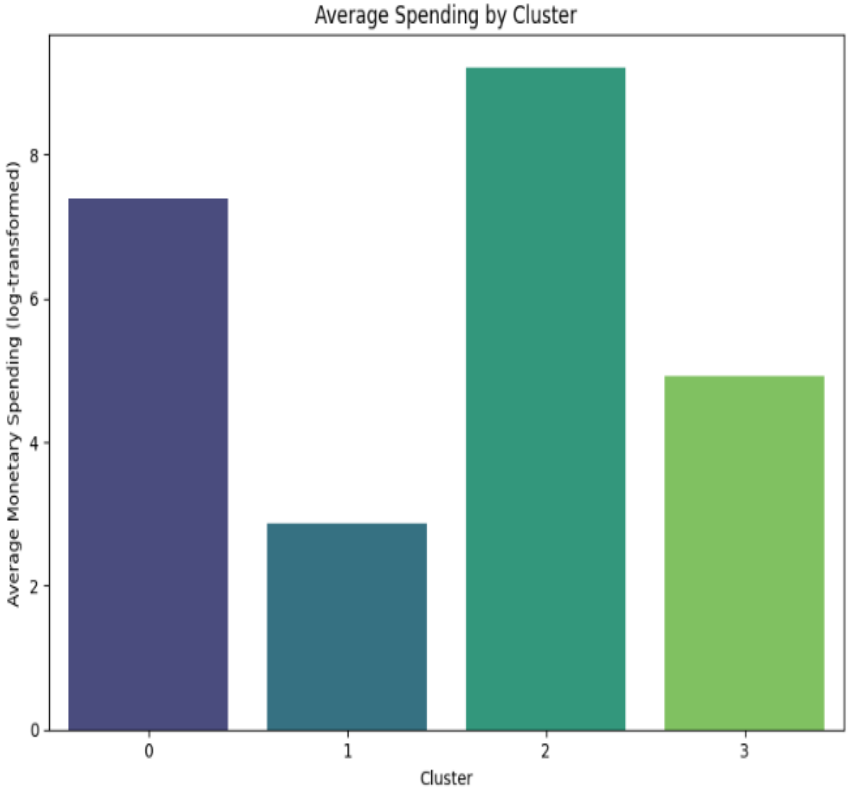
- **# Fit the K-Means model using the optimal number of clusters obtained after understanding the elblow plot**



🧠 Customer Segmentation & Insights – 📊 RFM Analysis

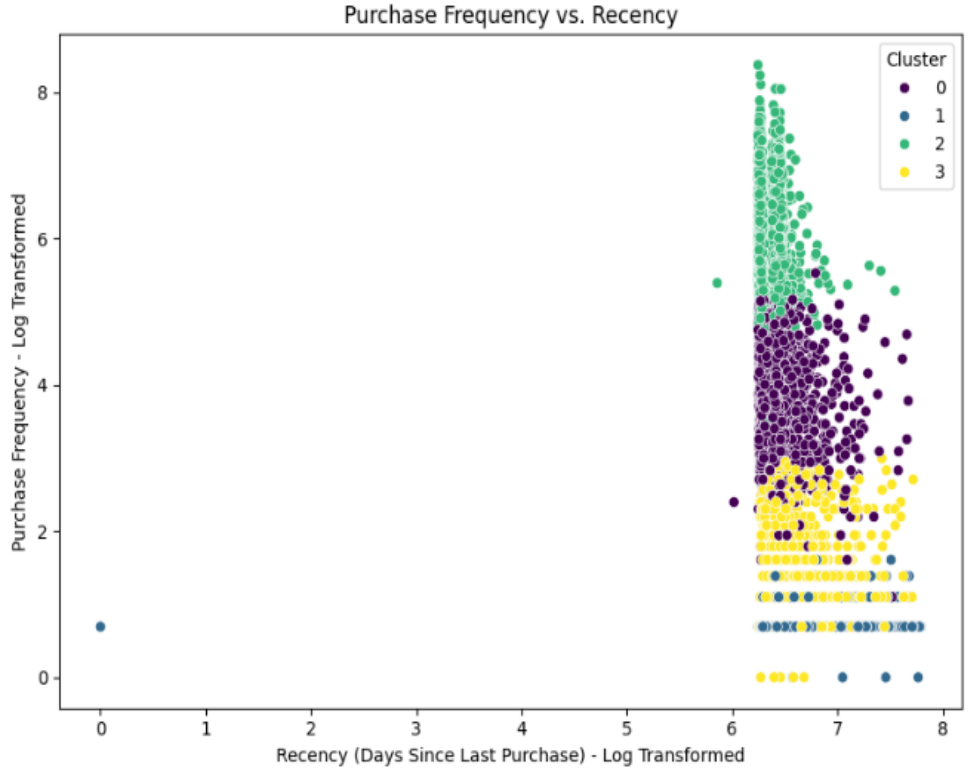
- Analyse the Cluster Distribution by Income
- Analyse the Average Spending by Cluster

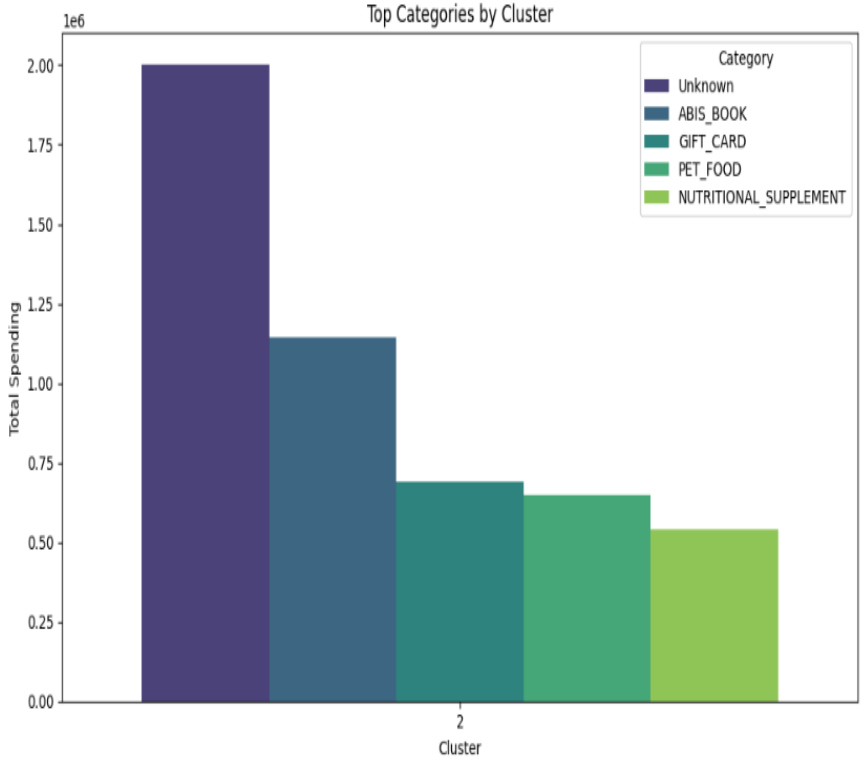




🧠 Customer Segmentation & Insights – 📊 RFM Analysis

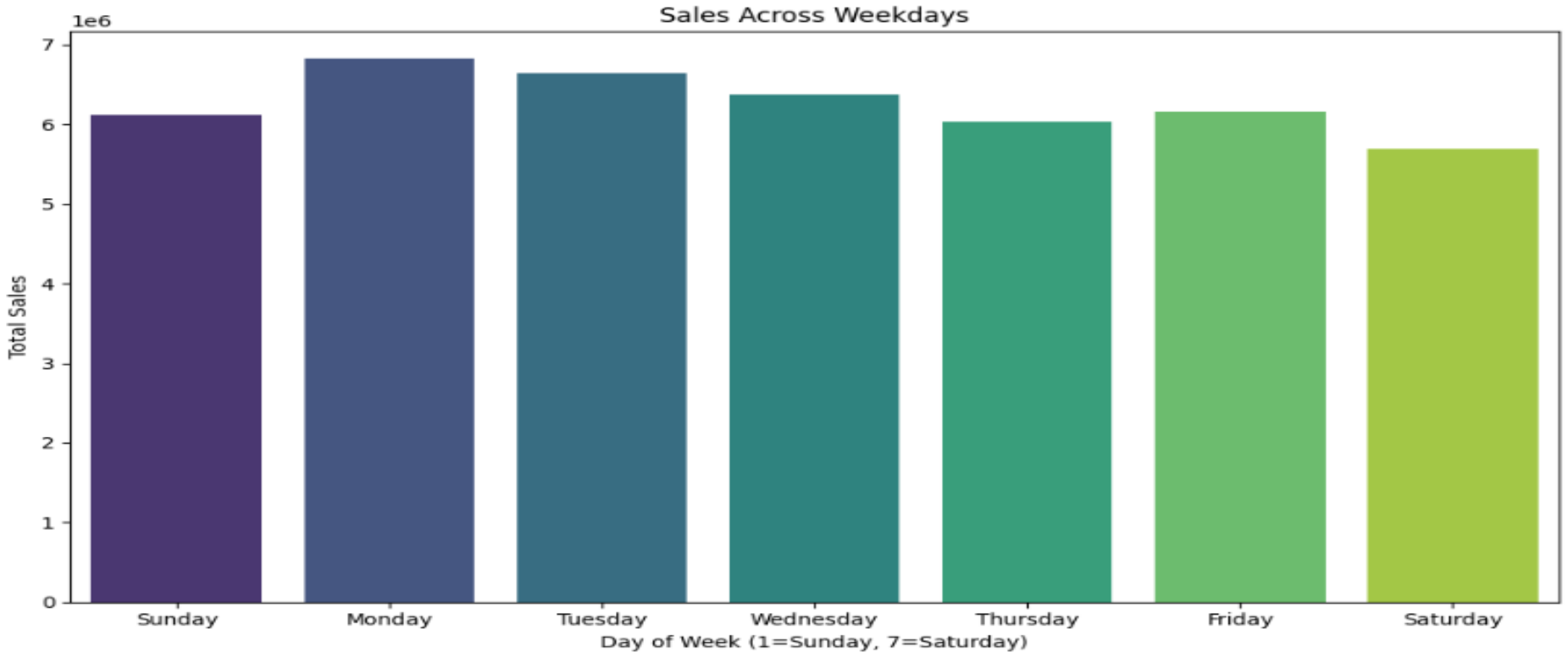
- Analyse the Purchase Frequency vs. Recency
- Analyse the top categories by clusters





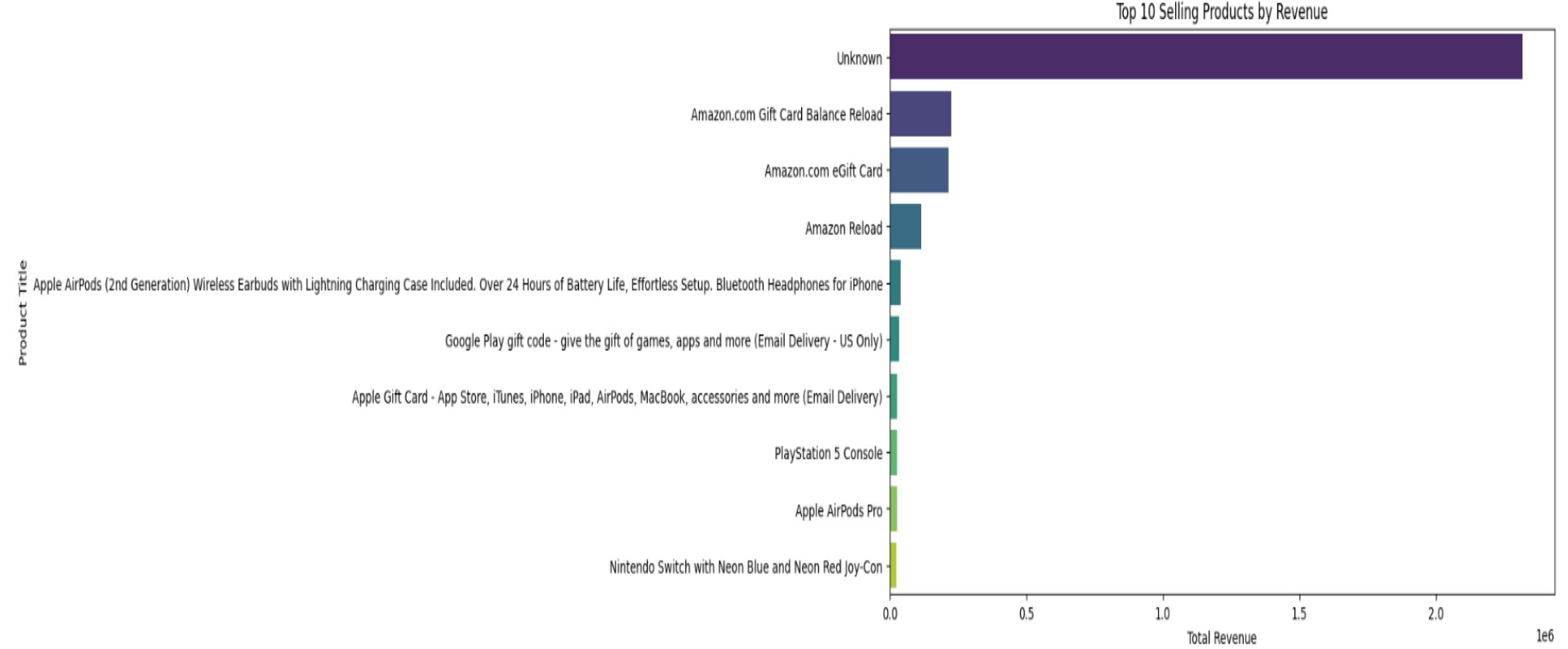
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- 4.2 Insights
- 4.2.1 When to schedule effective promotions



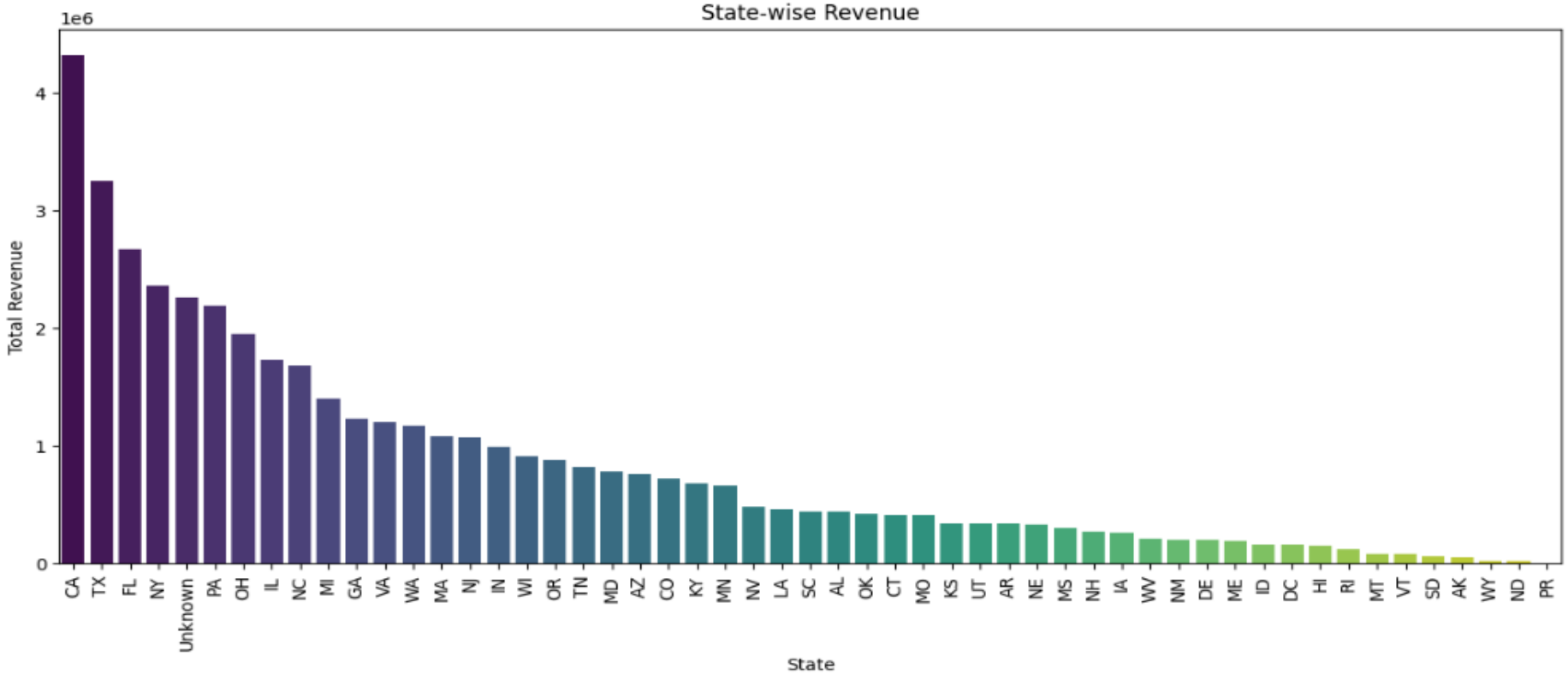
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- **4.2 Insights**
- **4.2.2 Top-selling Products**



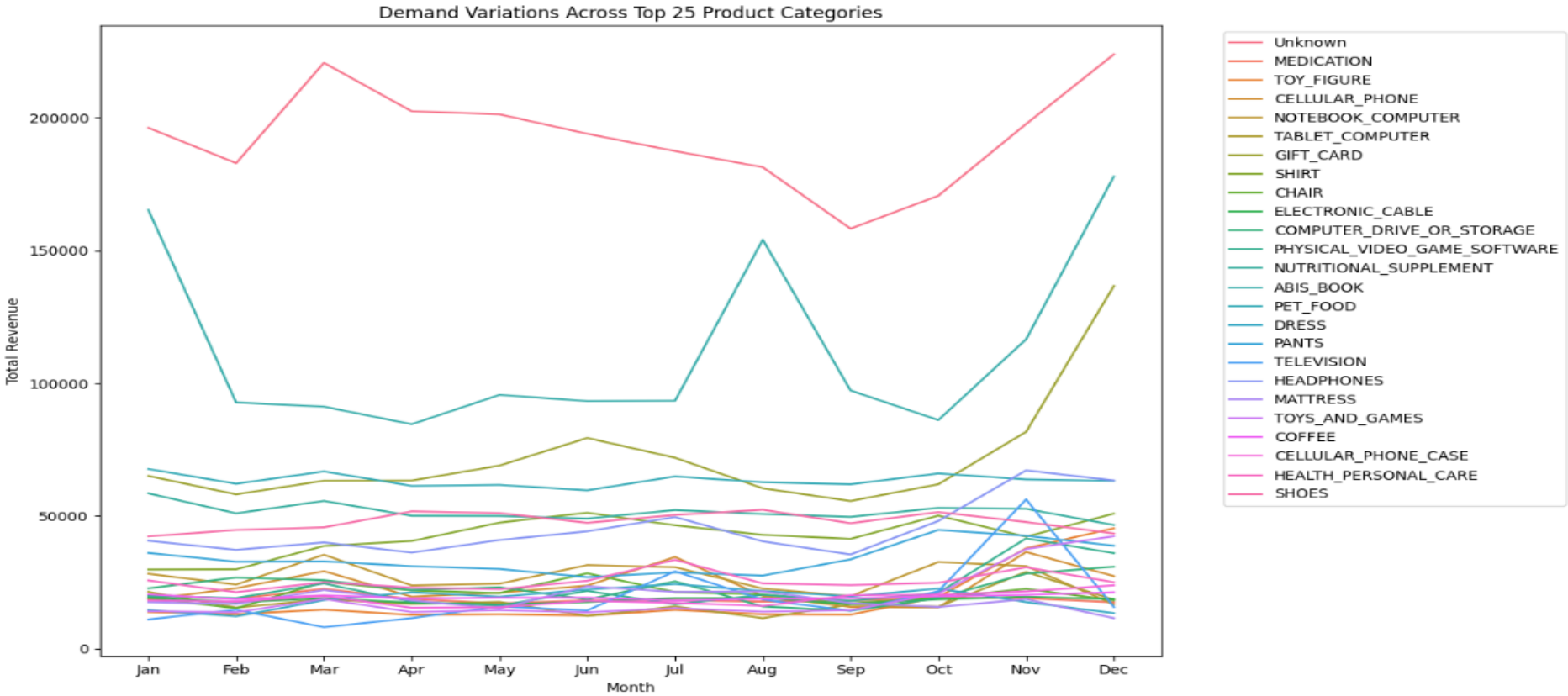
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- 4.2 Insights
- 4.2.3 State-wise revenue Distribution



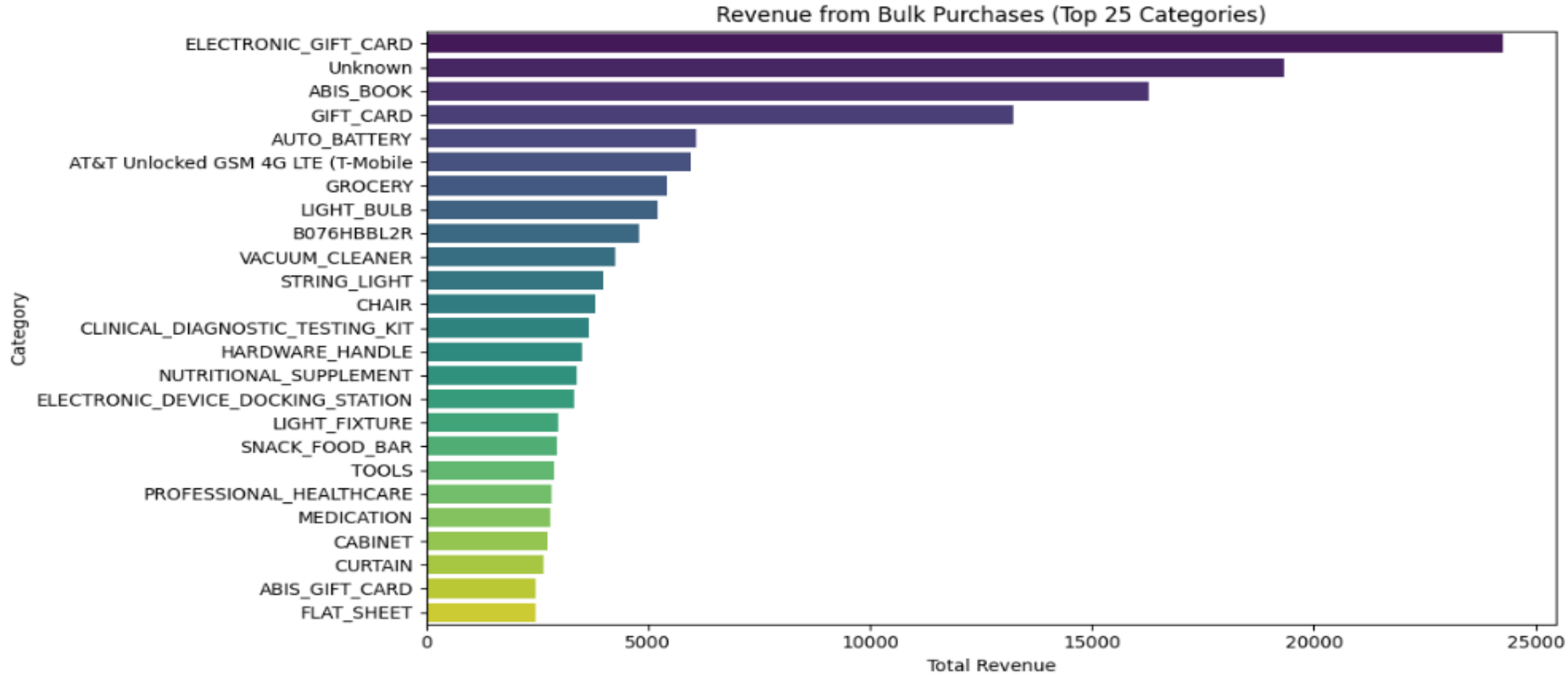
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- 4.2 Insights
- 4.2.6 Demand Variations across product categories



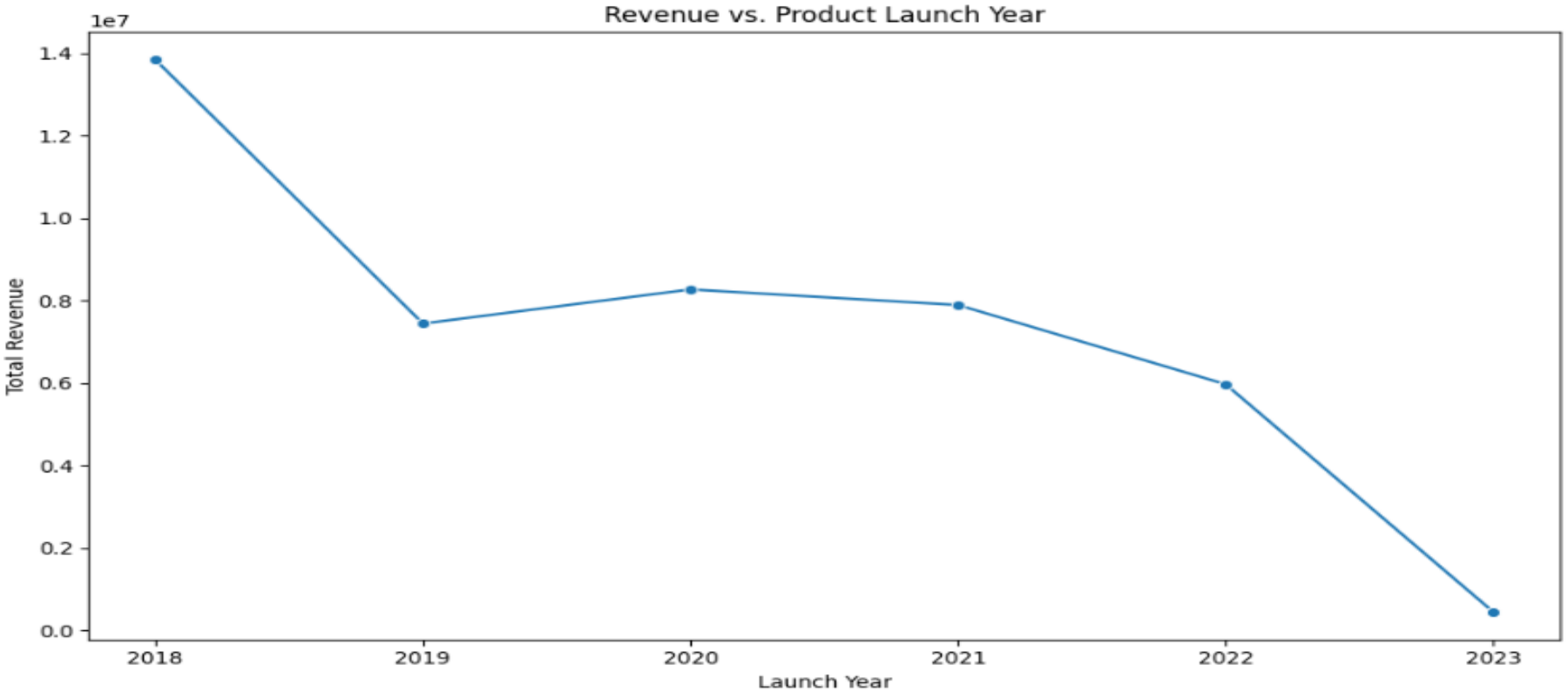
🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- **4.2 Insights**
- **4.2.7 Assess how bulk purchases affect revenue and supply chain operations**



🧠 Customer Segmentation & Insights – 📊 RFM Analysis

- 4.2 Insights
- 4.2.8 Compare lifecycle strategies



✅ Evaluation & Conclusion

✅ Evaluation & Conclusion

📝 Summary of Customer Behavior Analysis

- Uncovered purchase patterns, customer behavior, and product performance in the e-commerce domain.
- Applied RFM segmentation to identify high-value and at-risk customer groups for targeted marketing.
- Analyzed sales trends by time, category, and region to guide inventory and supply chain improvements.
- Detected behavioral differences across demographics and purchase frequencies.
- Identified trends in bulk buying, product lifecycle stages, and potential fraud activity.

Summary & Recommendations

📌 Key Recommendations

1. 🎯 Leverage customer segmentation for personalized marketing strategies.
2. 📦 Align inventory with demand cycles and seasonal trends.
3. 🚚 Enhance logistics to better serve bulk purchases.
4. 🔄 Optimize product lifecycles for timely innovation and cost efficiency.
5. 🛡️ Strengthen fraud detection using behavioral flags.
6. 💡 Invest in loyalty programs to improve retention.
7. 🌍 Focus on high-performing regions and segments to drive targeted growth.


THANK
YOU