Customer Behavior Analysis Report

Goal Statement:

The goal of this analysis is to derive actionable insights from customer transaction data in order to optimize business decisions. The key objectives include understanding purchasing patterns, detecting anomalies, classifying customers, behavioral trends, and improving overall business strategies.

Analysis Approach:

1. **Data Preparation:** Ensure data consistency, apply sampling techniques to Structure and prepare the data for further processing and analysis.
2. **Data Cleaning:** Fixing columns, handling missing values and outliers, perform feature engineering.
3. **Exploratory Data Analysis:** Identifying purchasing trends and key metrics using statistic graphs and plot.
4. **Customer Segmentation (RFM Analysis) and Business Insights:** Performing RFM analysis, studying sales patterns, and investigating anomalies
5. **Evaluation and Conclusion:** Providing actionable insights and recommendations for optimizing operations and customer engagement.

1. Data Preparation:

- We verified the data consistency and efficiency, also ensured all columns are correctly formatted.
- We found some missing values in columns like "**Shipping Address State**", "**Title**", "**Category**", "**ASIN/ISBN (Product Code)**" and "**Q-life-changes**".
- We structured and prepared the dataset for further processing, ensuring that relevant features are retained

2. Data Cleaning:

2.1 Handling Missing values:
- Prepare the data for further analysis by performing data cleaning such as missing value treatment, handle data schema, outlier analysis, and relevant feature engineering techniques.
- We handled the missing values by filling it with appropriate data like "**Unknown**", "**Not Specified**", etc. and therefore no missing values left.

2.2 Feature Engineering:
- Performed feature engineering on the dataset to extract relevant/ create new features as required and map specific data types.
- We checked for any negative or invalid entries , but there were none.
- We mapped categorical column like "**Q-demos-income**" and "**Q-demos-gender**" values to numeric values.

2.3 Data Cleaning:
- We handled the data using data cleaning techniques such as data duplication, dropping unnecessary values etc.
- We created the "**cleaned dataset**" with no duplicate and clean values and stored/saved it locally for further analysis

3. Exploratory Data Analysis:

3.1 Analyse purchases by hour, day and month:
- *"Examine overall trends in purchases over time and analyse the trends by hour, day, month."*

Key Findings:

- ❖ As shown in below graph, the highest purchases were mostly done on **Weekdays** as compared to Weekends.

- ❖ Significant seasonal spikes in purchases are observed in month of '**December**' reason can be during major holidays. And at year of **2021.**
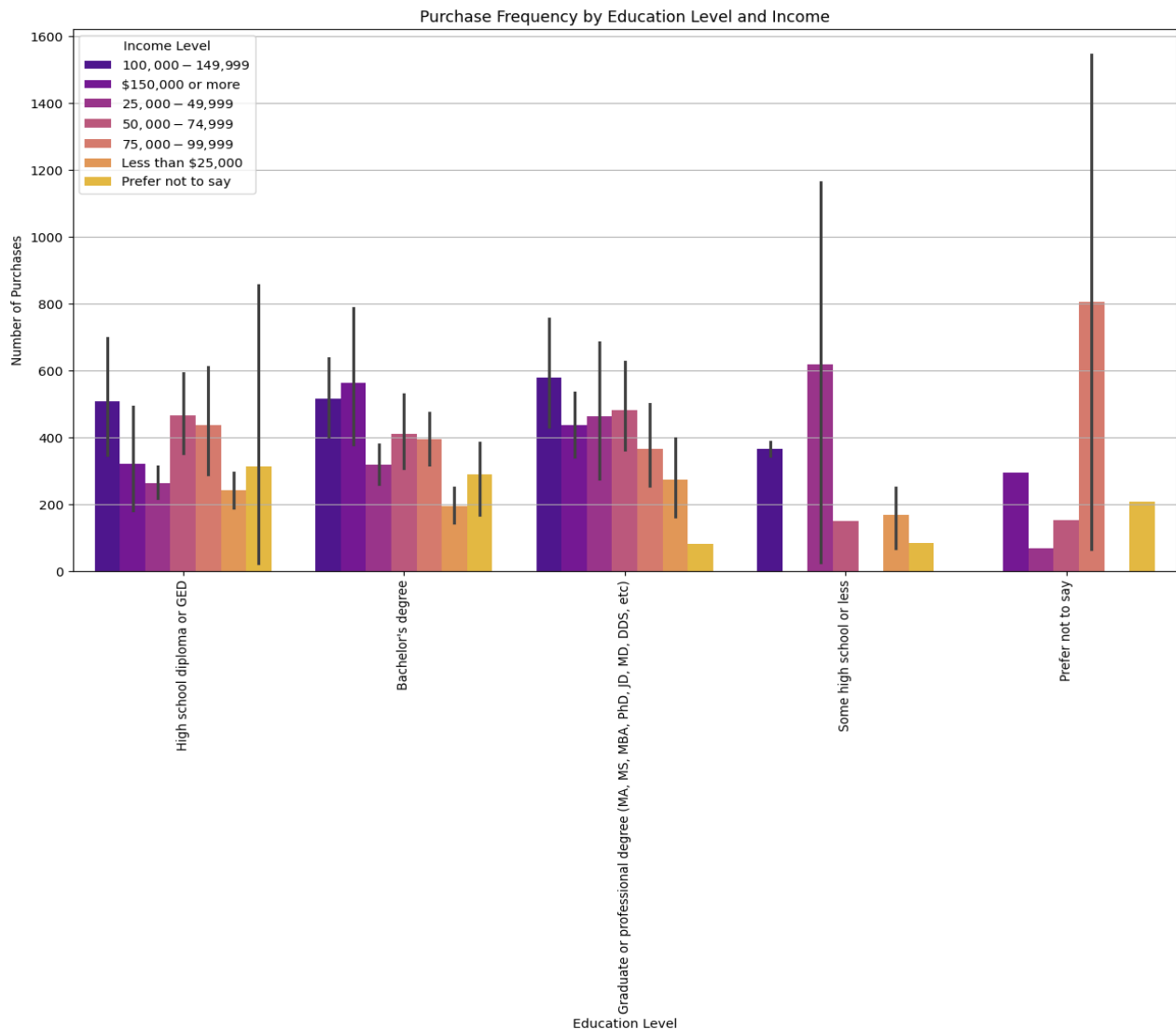- ❖ **Monday** exhibits higher purchase sales compared to other Days.

Plots Figure:

Purchases by Hour — Purchases by Day of the Week — Purchases by Month

Monthly Purchase Trends

Yearly Purchase Trends
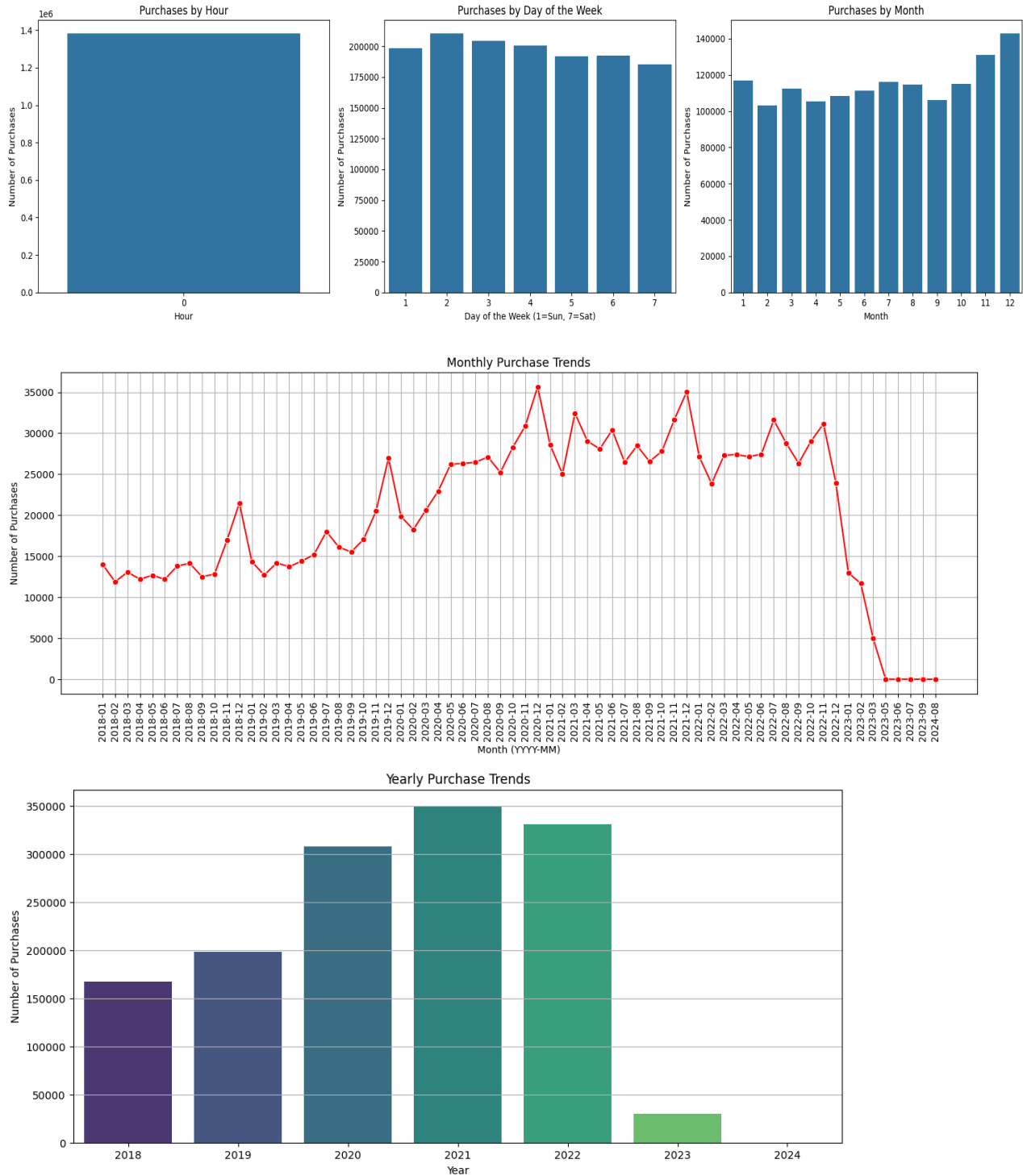
3.2 Customer Demographics vs Purchase Frequency:

- *"Analyse the trends between the customer demographics and the purchase frequency"*

Key Findings:

- ❖ As shown in above graph, Gender category **Females** of age **55-64 years** exhibits high purchase sales as compared to other category values.
- ❖ The people living in "**California**" and belongs to "**White or Caucasian**" showed great interest in purchasing products.
- ❖ The customers having "**prefer not to say**" about their education level with income Level of **$50,000-74,999** makes highest number of Purchases.

Purchase Frequency by Education Level and Income

Purchase Frequency by State and Race

State (rows) × Race (columns). Column abbreviations:
- C1 = American Indian/Native American or Alaska Native
- C2 = Asian
- C3 = Asian,Native Hawaiian or Other Pacific Islander
- C4 = Asian,Other
- C5 = Black or African American
- C6 = Black or African American,American Indian/Native American or Alaska Native
- C7 = Black or African American,American Indian/Native American or Alaska Native,Asian
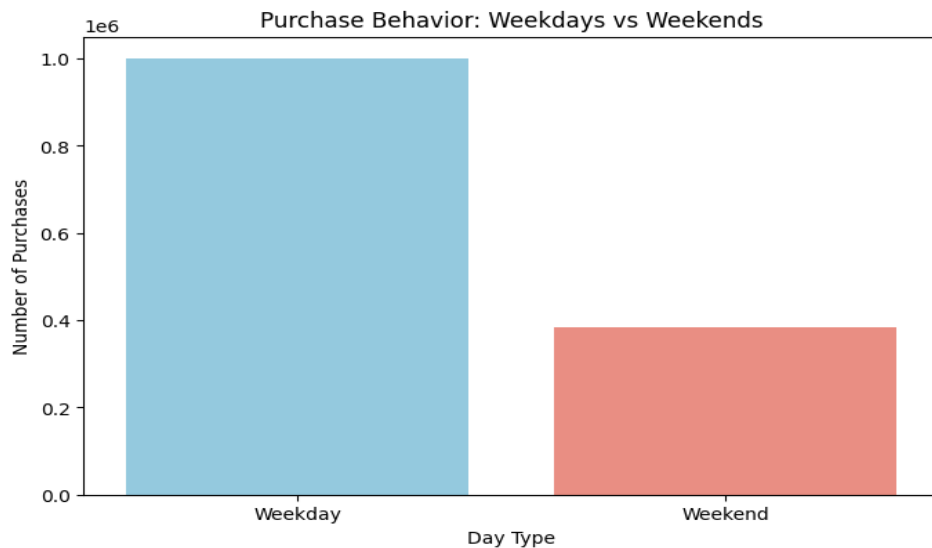- C8 = Black or African American,Asian
- C9 = Black or African American,Other
- C10 = Native Hawaiian or Other Pacific Islander
- C11 = Other
- C12 = White or Caucasian
- C13 = White or Caucasian,American Indian/Native American or Alaska Native
- C14 = White or Caucasian,American Indian/Native American or Alaska Native,Asian
- C15 = White or Caucasian,American Indian/Native American or Alaska Native,Other
- C16 = White or Caucasian,Asian
- C17 = White or Caucasian,Asian,Native Hawaiian or Other Pacific Islander
- C18 = White or Caucasian,Black or African American
- C19 = White or Caucasian,Black or African American,American Indian/Native American or Alaska Native
- C20 = White or Caucasian,Black or African American,American Indian/Native American or Alaska Native,Asian
- C21 = White or Caucasian,Black or African American,Asian,Native Hawaiian or Other Pacific Islander
- C22 = White or Caucasian,Black or African American,Other
- C23 = White or Caucasian,Native Hawaiian or Other Pacific Islander
- C24 = White or Caucasian,Other

State	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
Alabama					707						270	12826	277											151
Alaska	6											991	53										25	
Arizona	379	896			853						175	17816		35		1972	589	1505		229				
Arkansas	28				483						1122	8947	113						61					
California	915	23422	627	56	4589					41	5571	100578	3398			2020		1133		773				562
Colorado	823	203									289	23812	718			410								
Connecticut					1578						140	9808												
Delaware					1350							2507												
District of Columbia		95			426							6525												
Florida		2760	448		5024	5	229	255	18			79441				232		716	157					203
Georgia		454	475		10301			31			1063	29819	160			141		1252						
Hawaii		660	301				685		33	30		1418				1131	93							
I did not reside in the United States												665				87								
Idaho												4800												
Illinois	1522	2609			6939						125	43721	179			801			64					5
Indiana		668			941							34757				679		619						359
Iowa	31	107			190						25	7912												
Kansas		408										11425	308			212								
Kentucky		263			358			324				19134				293								
Louisiana		67			1400							15637	372			339				357				
Maine												3587												
Maryland	498	1889	475		1756		1440		9			22674				32							260	257
Massachusetts		2592			443				73			27537	462			1277								
Michigan	772	2409			3139	156						37992				725		590						
Minnesota		236			152						368	18147				942								
Mississippi		180			742							6614	41											
Missouri	67	370			1170						499	12874					9							
Montana												2573												
Nebraska		1278										6581				82								
Nevada		3700			10							9095	231			1535		652						747
New Hampshire					215							9304				705								
New Jersey		2520			1983						207	26443	1703			1427		209						895
New Mexico		522										5422	485			123		525						
New York		7696			5761	175		20			1743	62568	302	54		6		1978	646	230		690		156
North Carolina	29	1699			7151				536	146		47150	426	17		213							289	
North Dakota												720												
Ohio		2033			3132							62473						523	121					
Oklahoma	429	370			114							13125	837											
Oregon		536									939	25155	2575			243		1626	128				198	67
Pennsylvania		1706			4743						920	70291	616			292		234	56			69		640
Rhode Island												1754				696								
South Carolina	465	340			1543			75				11193				71								
South Dakota												1527												
Tennessee	164	555			200			67				19783	1935			1232								
Texas	62	5372			10610	23					3840	80629	947			3149		1611	119					1948
Utah		989										7100				847								
Vermont												3537												
Virginia		2256			2251						213	33286				44		1781						
Washington		3804			402						326	30641	565			1026		958		1726				
West Virginia					3							4723	462											
Wisconsin	397	724			2							28178				459								
Wyoming												334												

Purchase Frequency by Age and Gender

3.3 Purchase behavior weekend vs weekday:

- *"Compare the purchase behavior of customers on weekdays vs. weekends."*

Key Findings:

- ❖ As shown in above graph, **Weekday** exhibit higher sales compared to Weekend.

Purchase Behavior: Weekdays vs Weekends

3.4 Frequently purchased product pairs:

- *"Analyze how frequently products are purchased together (also known as Market Basket Analysis)"*
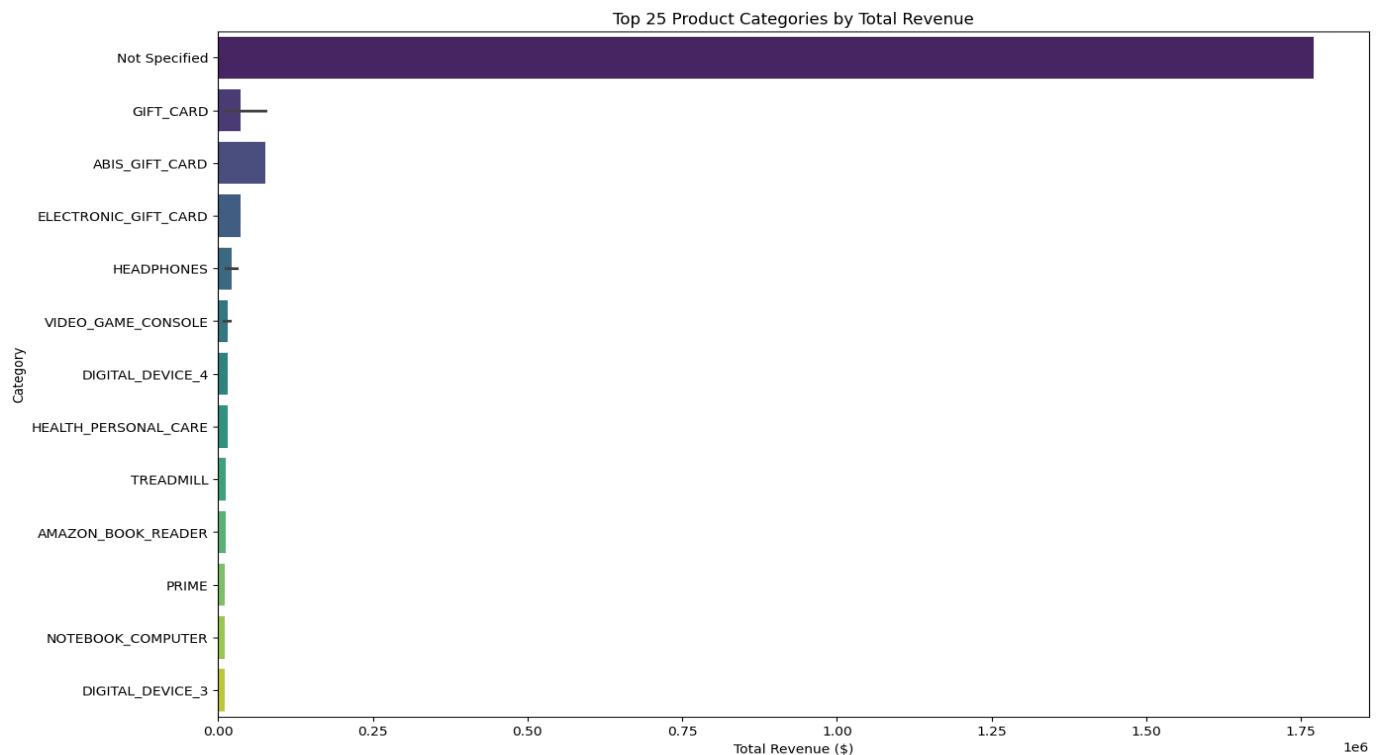
Key Findings:

❖ Specific pairs of products showed highest demand and can help in revenue growth by increasing their sales.

3.5 Examine Product Performance:

- *"Examine the performance of products by calculating revenue and item popularity."*

Key Findings:

❖ Products like **ABIS Book, Gift Card, Electronic Gift cards, Headphones** and many more were high in both Popularity and Total Revenue.

❖ Items as "**Not Specified**" spike the revenue and were top popular items.

Top 25 Product Categories by Total Revenue

Top 25 Product Categories by Total Popularity
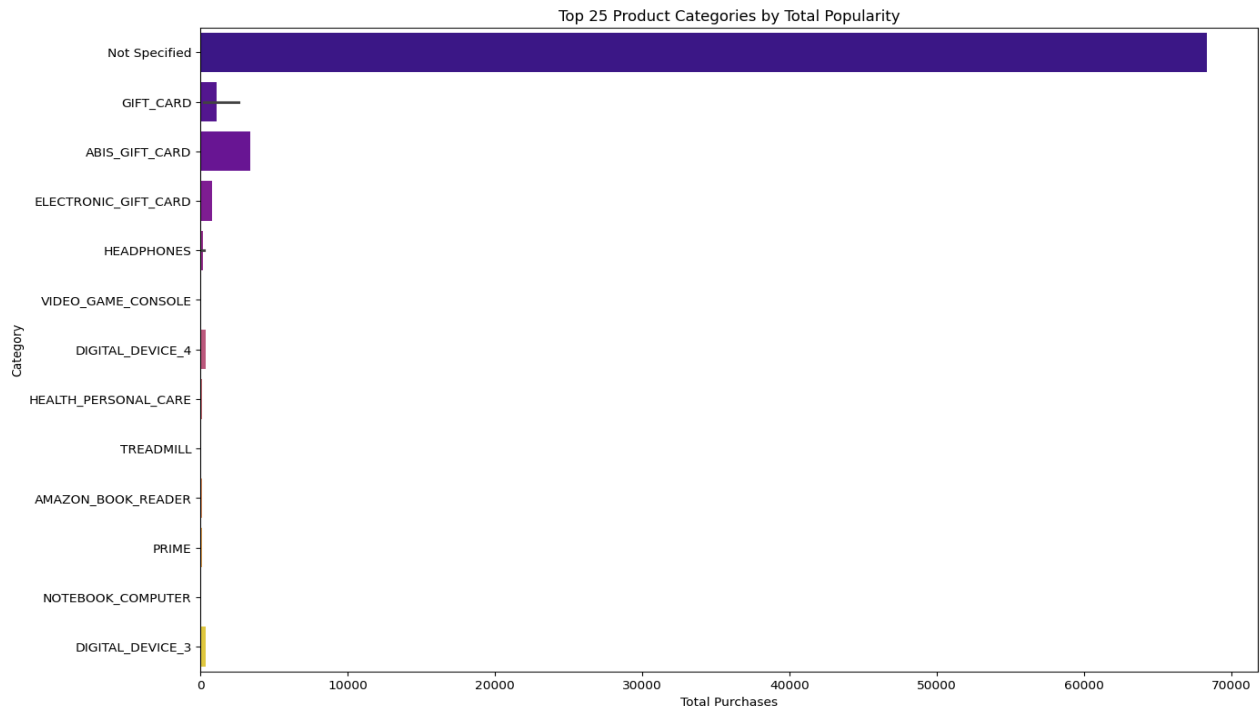
3.6 Top products by quantity:

- *"Identify the most frequently purchased products."*

Key Findings:

- ❖ Topmost frequently purchased Products by customer were most named "**Unknown**" but some of were **Amazon.com eGift Cards, Amazon.com Gift Card Balance Reload, Amazon Reload, Photo Prints** and many more.

Top 10 Products by Quantity
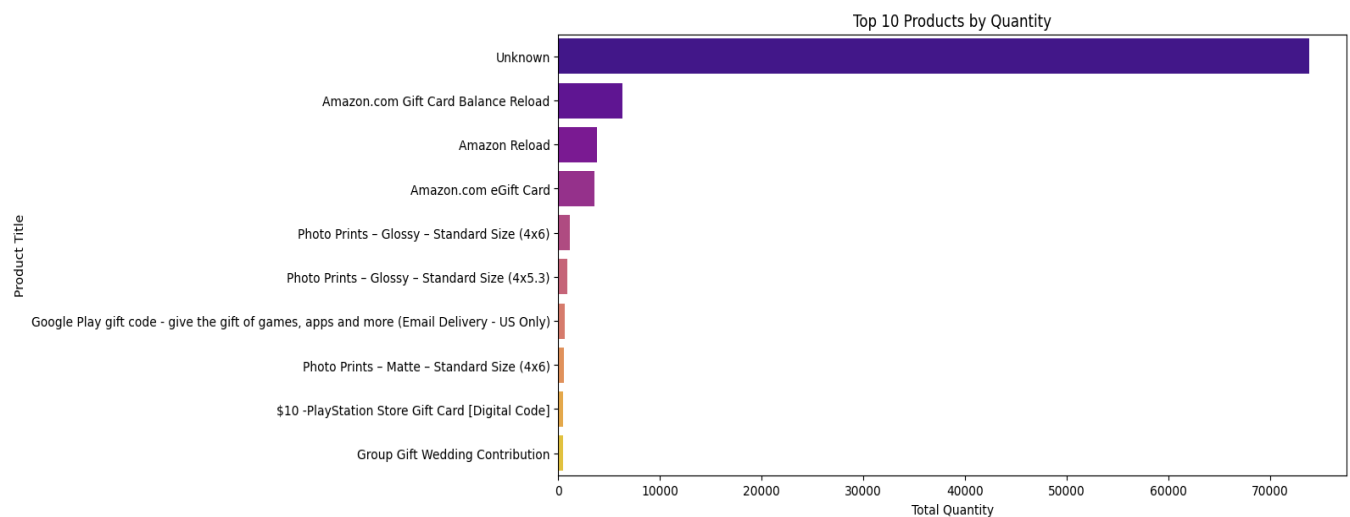
3.7 Distribution of Purchases by State:

- *"Analyze the distribution of purchases across states and categories."*

Key Findings:

- ❖ Topmost State by Purchase Distribution was **CA** (California) followed by **TX** (Texas), **FL** (Florida) and many more.

Top 25 States by Purchase Distribution
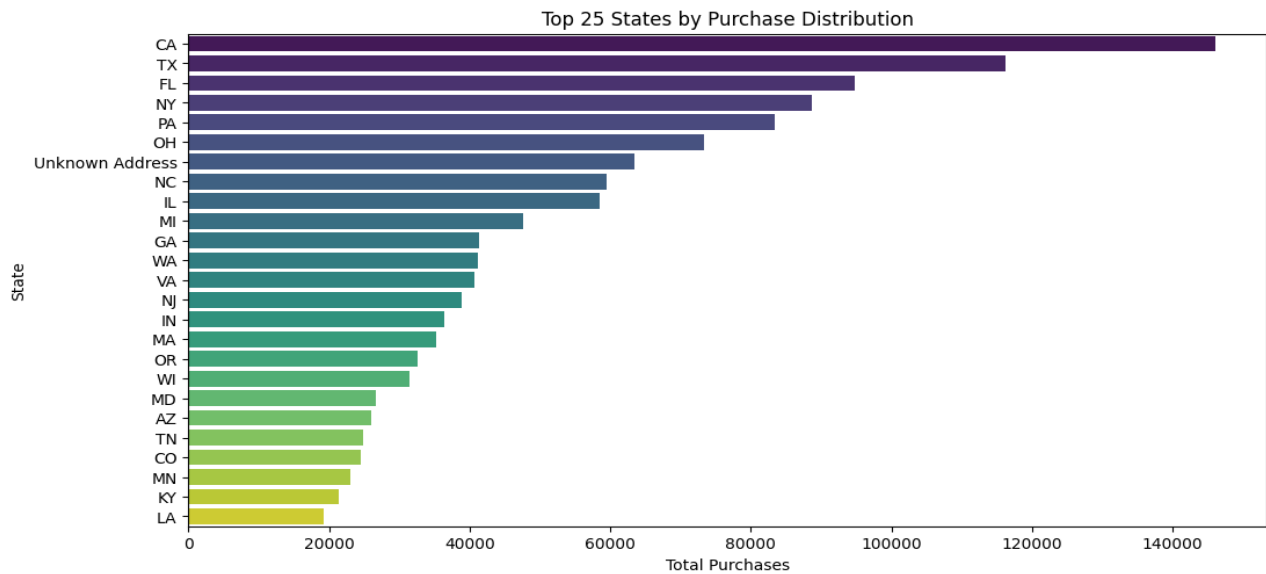
3.8 Price vs Product Quantity:

- *"Identify the Relationship between Price and Quantity"*
- ❖ Below pair-plot graph shows the relationship between Price and Quantity.

Relationship Between Product Price and Quantity Purchased
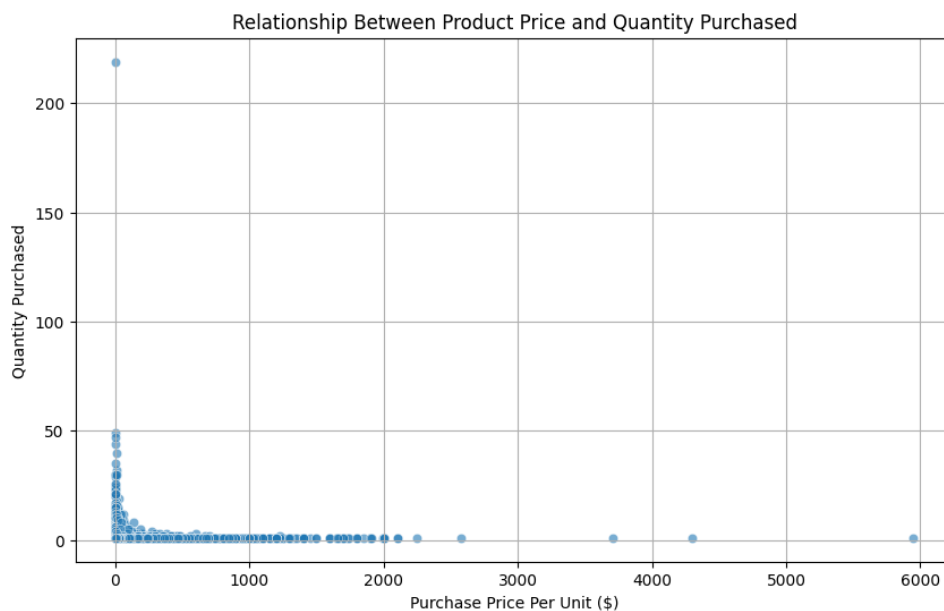
3.9 Analyse the spending KPIs:

- *"Calculate KPIs as the ratio of total transaction amount from non-recurring payments divided by the total number of customers who made a purchase"*

3.9.1 Analyse KPIs:

❖ The spending **KPIs** (**Average Spending Per Customer**) for E-commerce Company was found to be around **8222.85**.

Distribution of Total Spend per Customer

Distribution of Average Spend per Customer
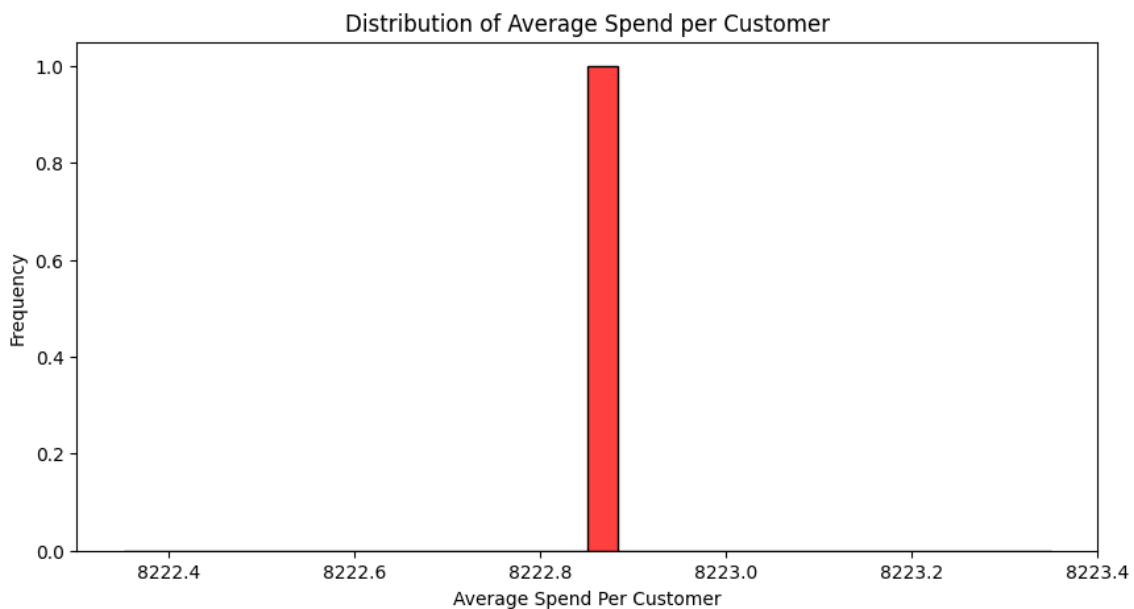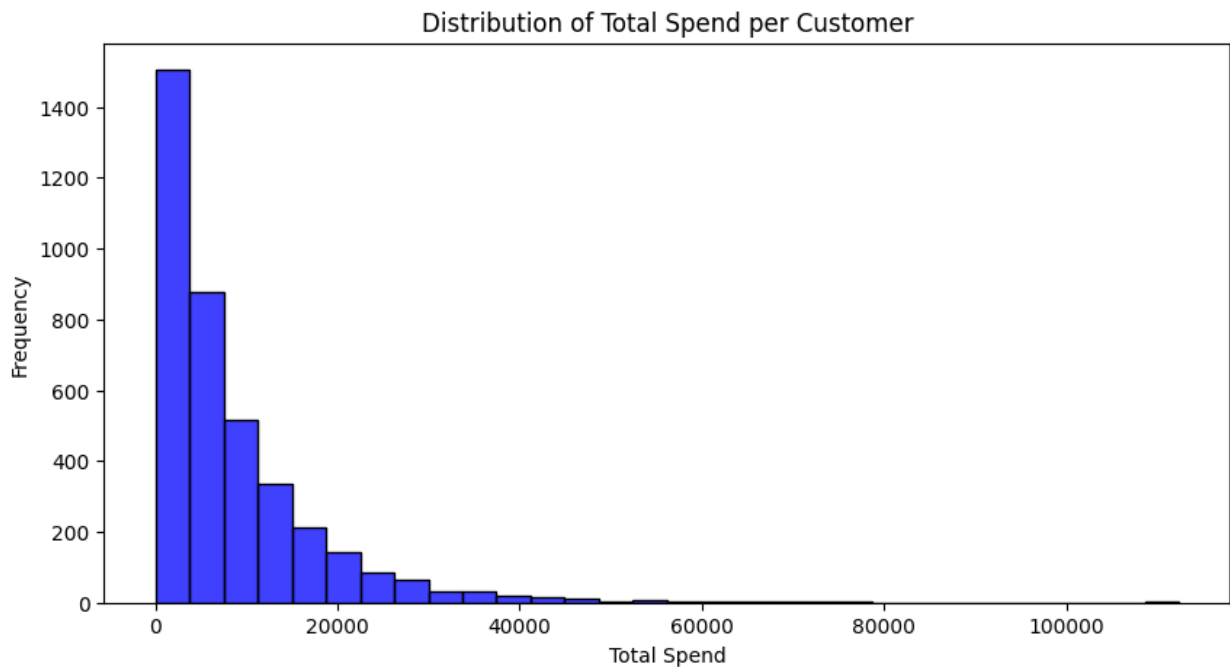
3.9.2 Analyse the Repeat Purchase Behavior of Customers:

❖ A significant portion of customers exhibited **Repeat purchase** behavior to enhance retention initiatives.
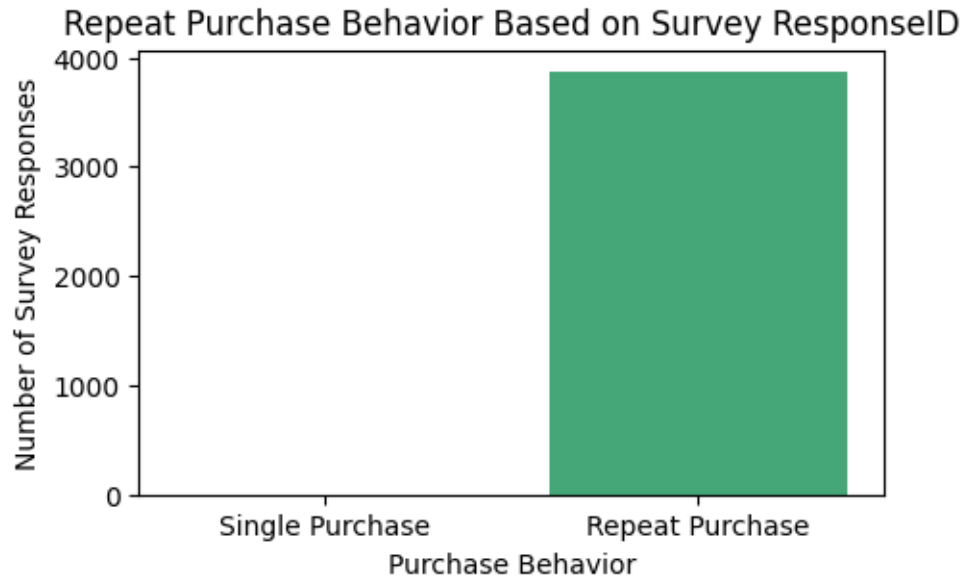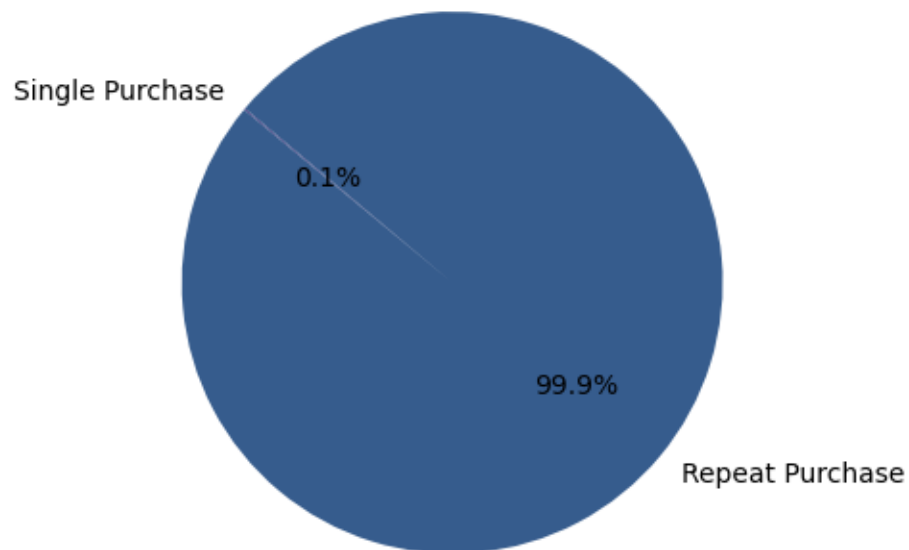
3.9.3 Analyse the top 10 high-engagement customers:

❖ Some top customers had high Purchase frequency and must be provided with regular discount and offers for being loyal customer to company.
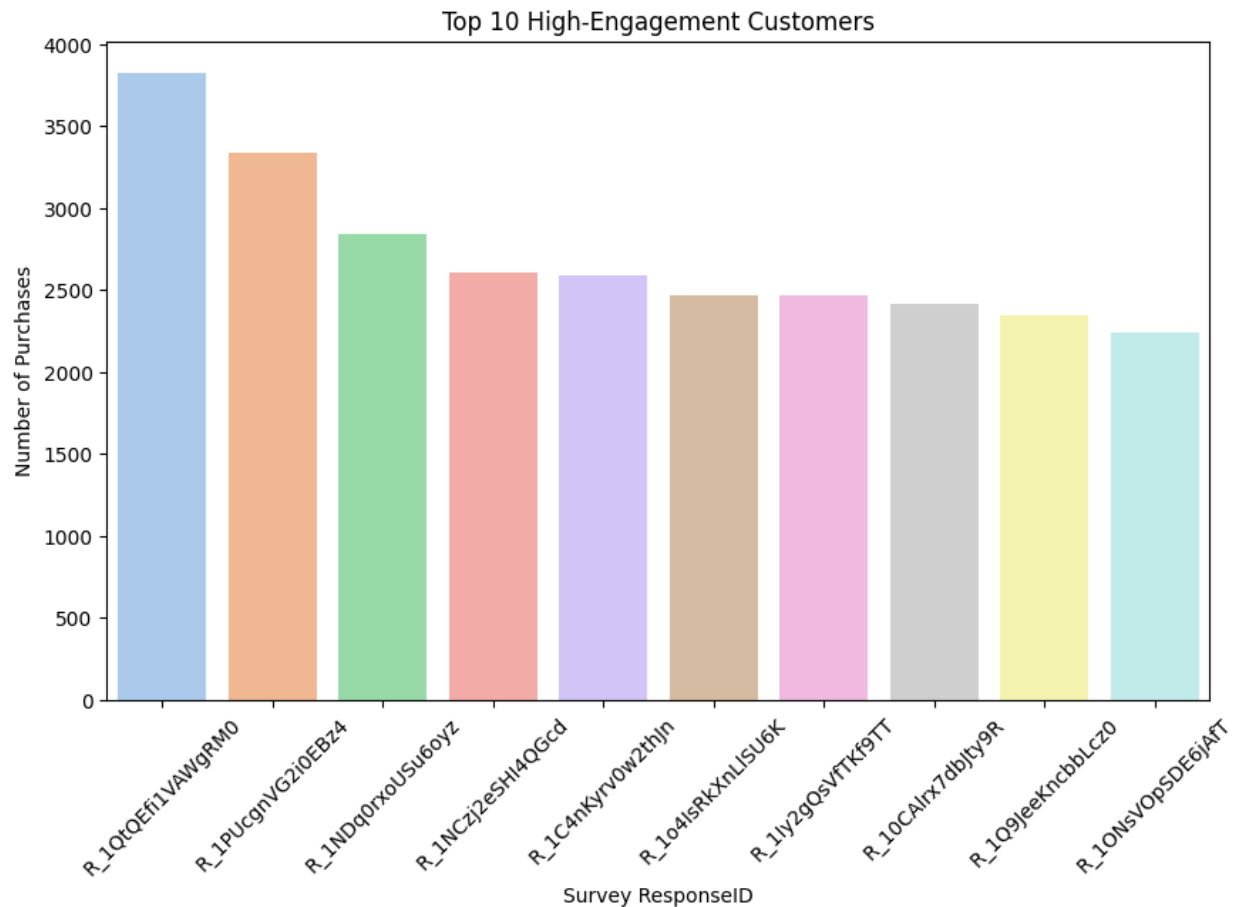
3.10 Seasonal trends in product purchases and their impact on revenues:

● *"Investigate the seasonal trends in product purchases and their impact on the overall revenue."*

Key Findings:

❖ The overall Seasonal Purchase trends were mostly found high in Year **'2021'** and **'2020'**.

Seasonal Trends in Product Purchases and Revenue
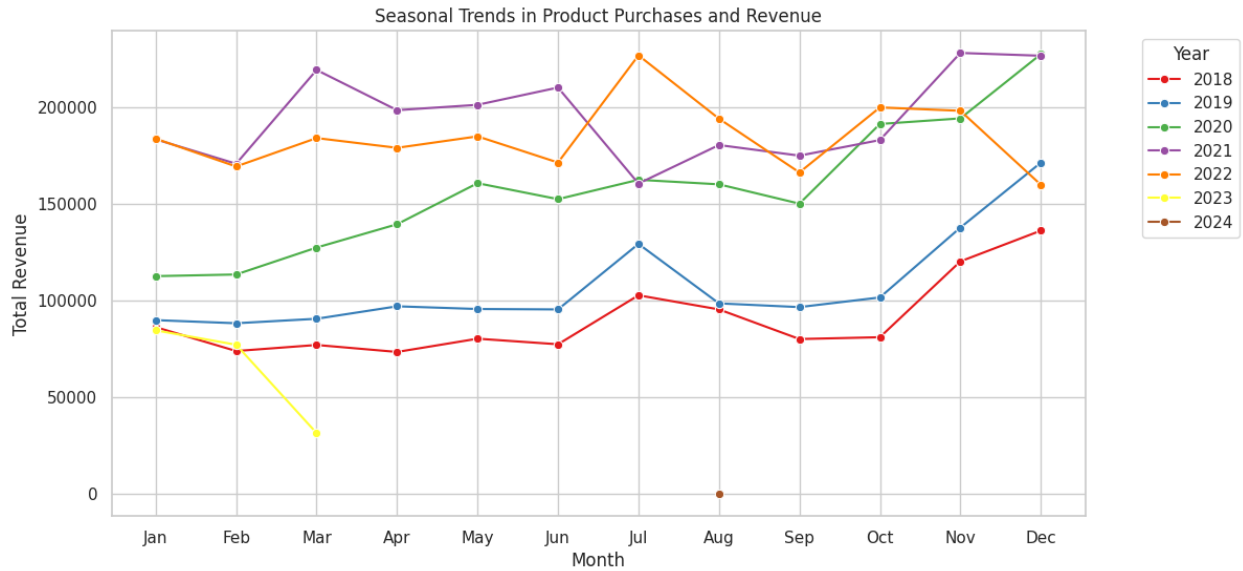
3.11 Customer location vs purchasing behavior:

- *"Examine the relationship between customer's location and their purchasing behaviors"*

Customer Location vs Purchasing Behavior
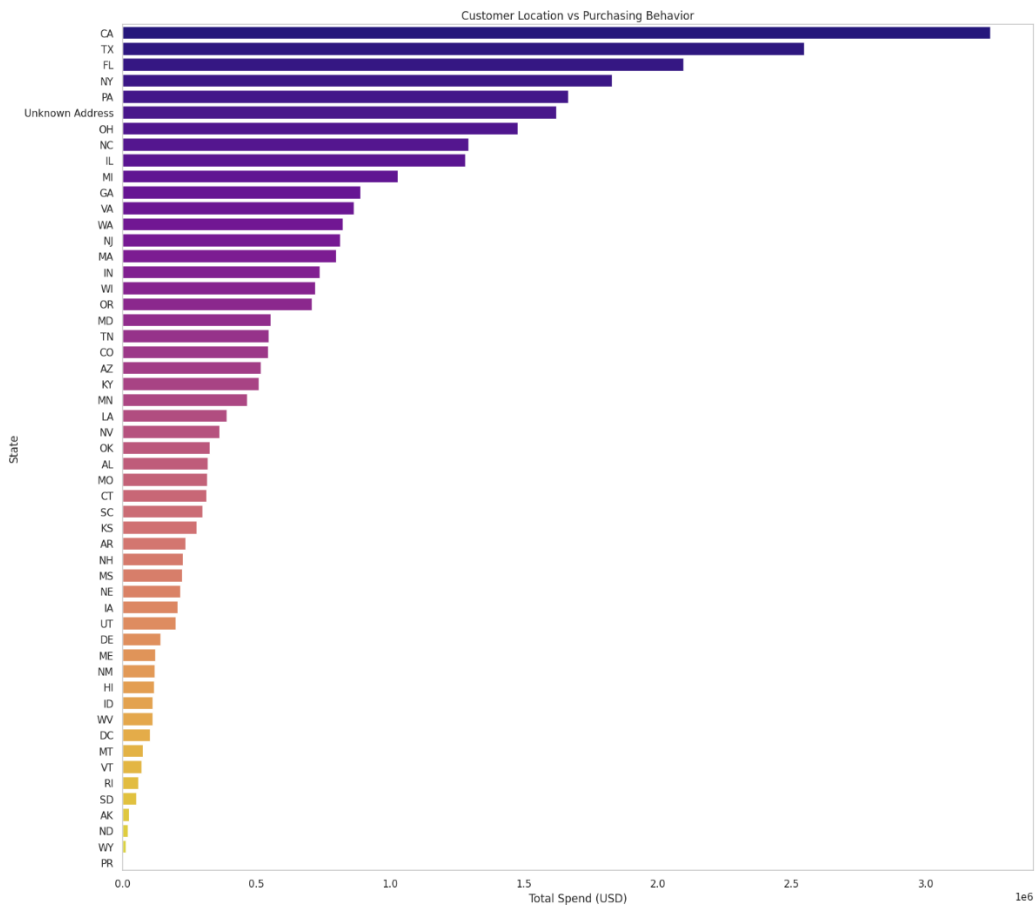
Key Findings:

❖ From above graph, Purchasing Behavior as per Location was found highly in State of **CA** (California) followed by **TX** (Texas) many more.

4. Customer Segmentation and Insights:

- RFM analysis identified high-value customers, dormant users, and frequent buyers.

- Personalized engagement campaigns can be designed to target each segment effectively.

4.1 Perform RFM Analysis:

- *"Classify customers using RFM analysis to tailor marketing strategies."*

Key Findings:

RFM Analysis is a powerful customer segmentation technique used to evaluate and quantify customer value based on three key dimensions:

❖ **Recency**,
❖ **Frequency**,
❖ **Monetary**.

Elbow Method for Optimal Clusters
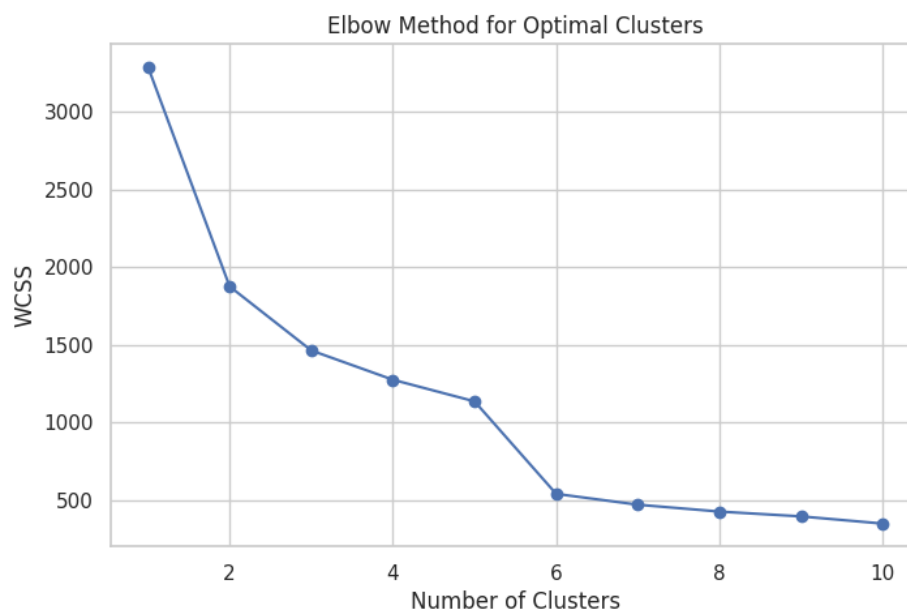
RFM Analysis: Customer Segmentation by Recency, Frequency, and Monetary
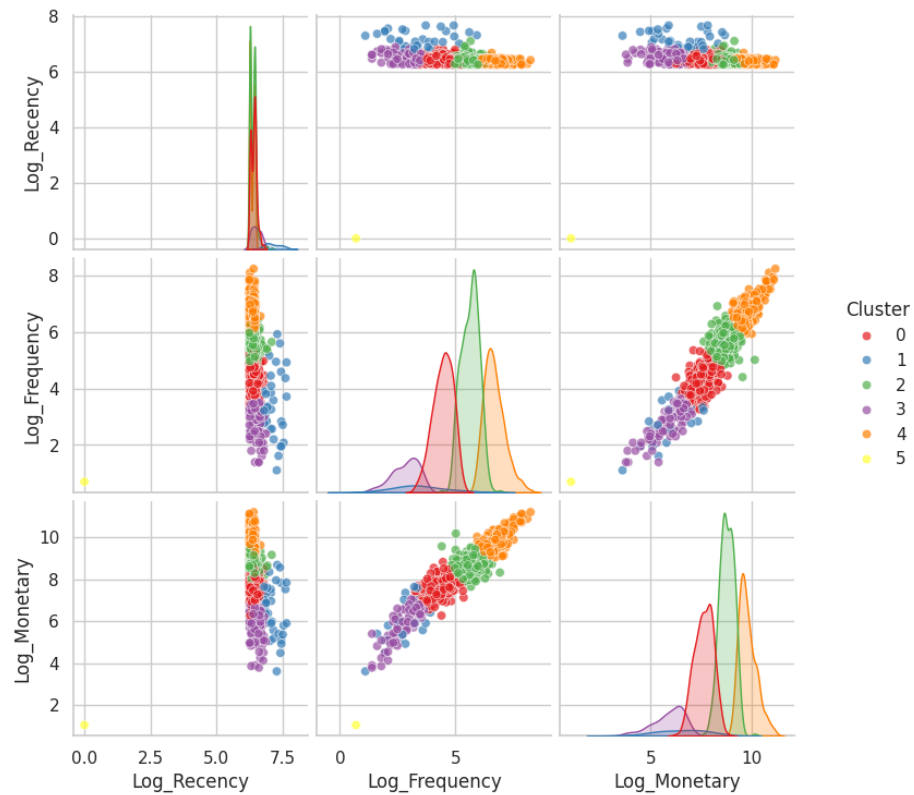
4.1.1 Analyse the Cluster Distribution by Income:

❖ Cluster Distribution by Income was high for **Cluster 0** in which Highest number of customers were of Income range **$25,000-49,999.**

Cluster Distribution by Income
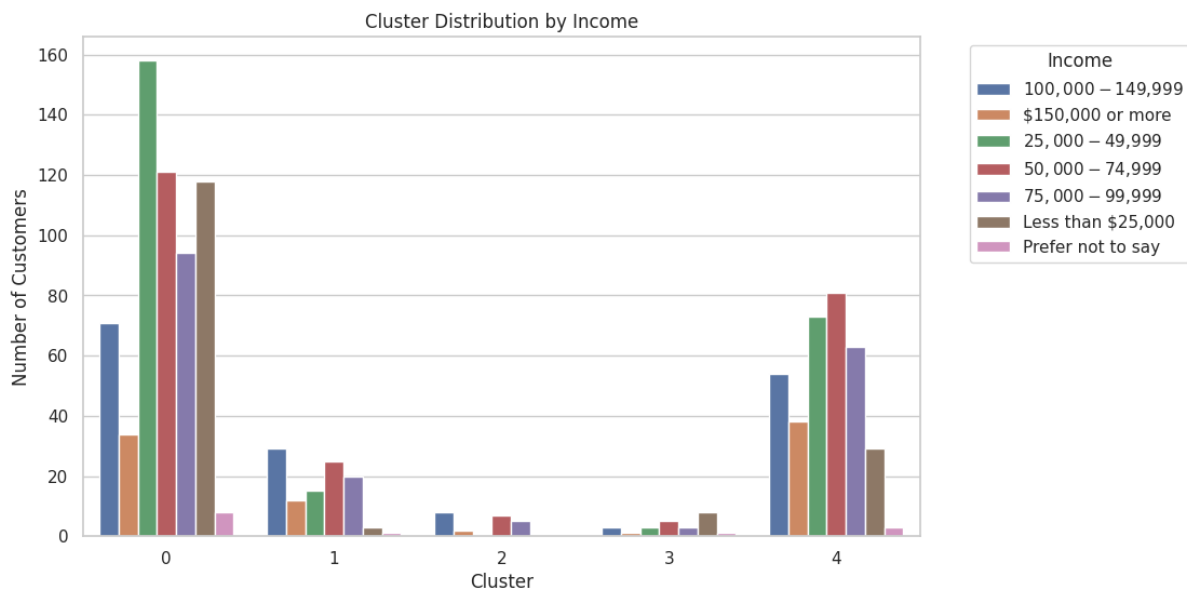
4.1.2 Analyse the Average Spending by Cluster:

❖ The Average Monetary Spending by Cluster was high in Cluster 2.
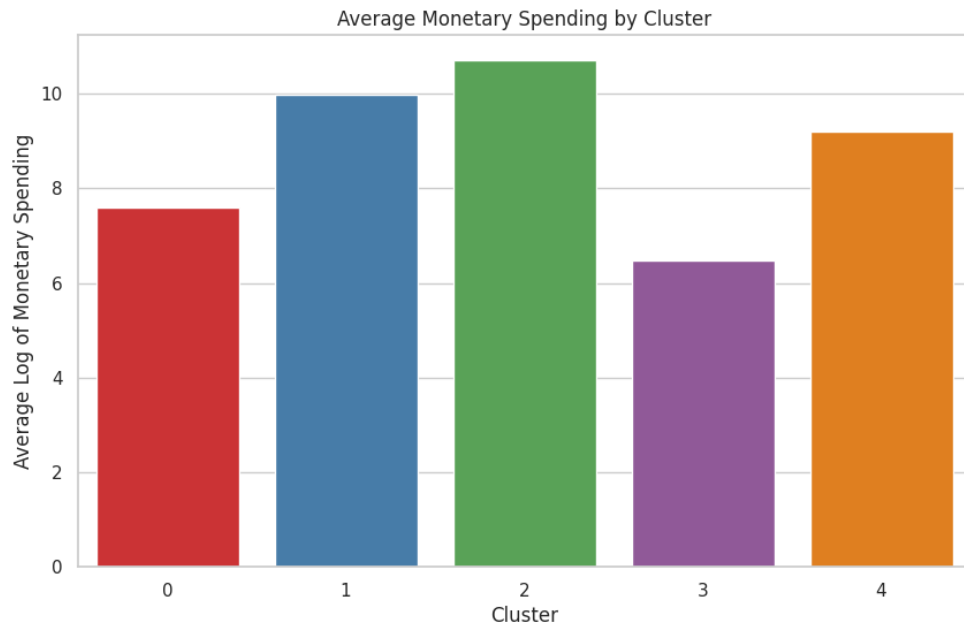
4.1.3 Analyse the Purchase Frequency vs. Recency:

❖ Below pair-plot shows the relationship between Purchase Frequency and Recency.
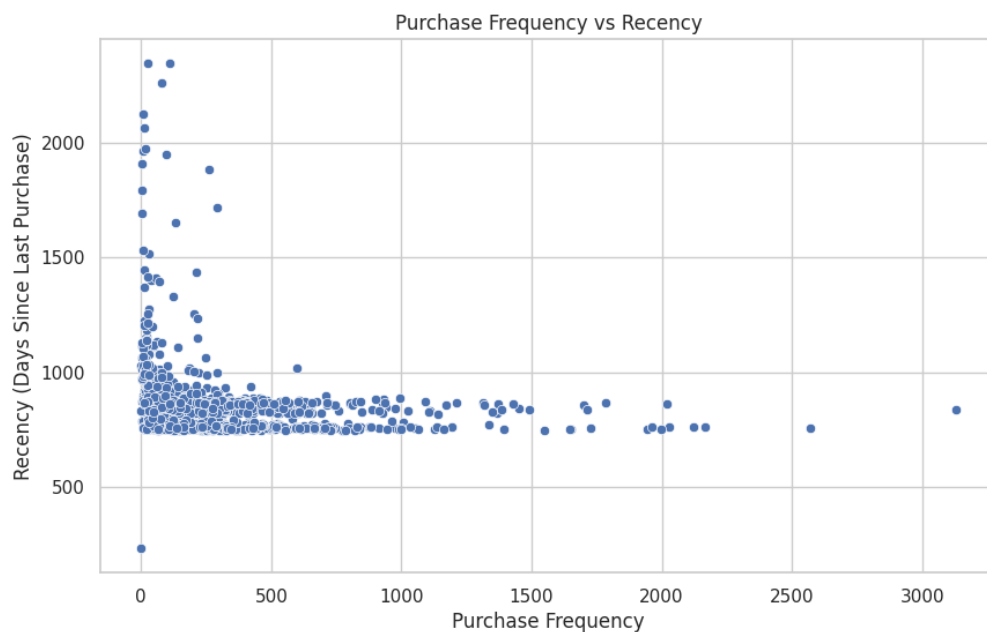
4.1.4 Analyse the top categories by clusters:

❖ The Product named "**Not Specified**" as a Top category showed highest total spending in **Cluster 4.**
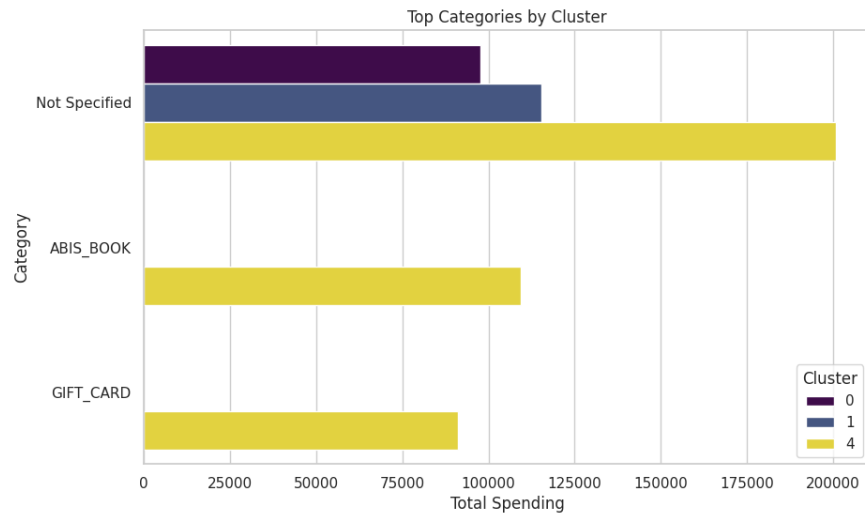
4.2 Insights:

4.2.1 When to schedule effective promotions:

- *"Compare sales across weekdays to schedule effective promotions"*

Key Findings:

❖ Sales were high on **Monday** followed by other weekdays; therefore, effective promotions must be scheduled moreover on weekdays.
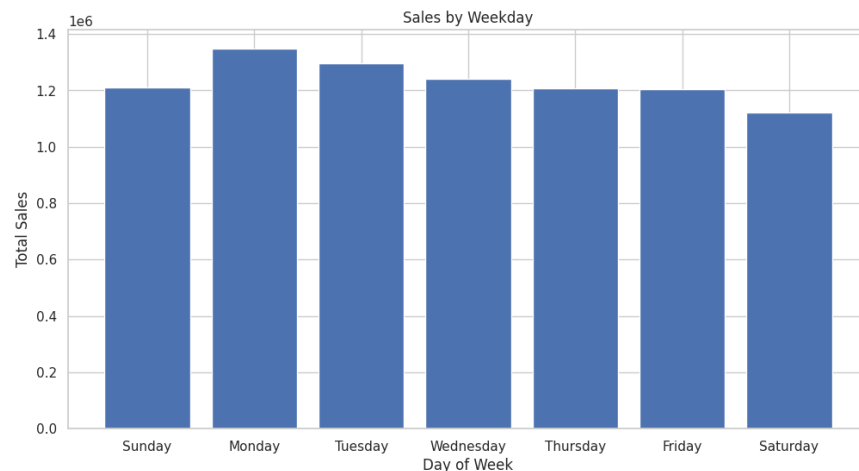
4.2.2 Top-selling Products:

- *"Identify top-selling products by considering revenue and engagement metrics"*

Key Findings:

❖ The Product code with **B00IX1I3G6** followed by some top 10 products can help in revenue growth by increasing their sales.

Top 10 Products by Revenue
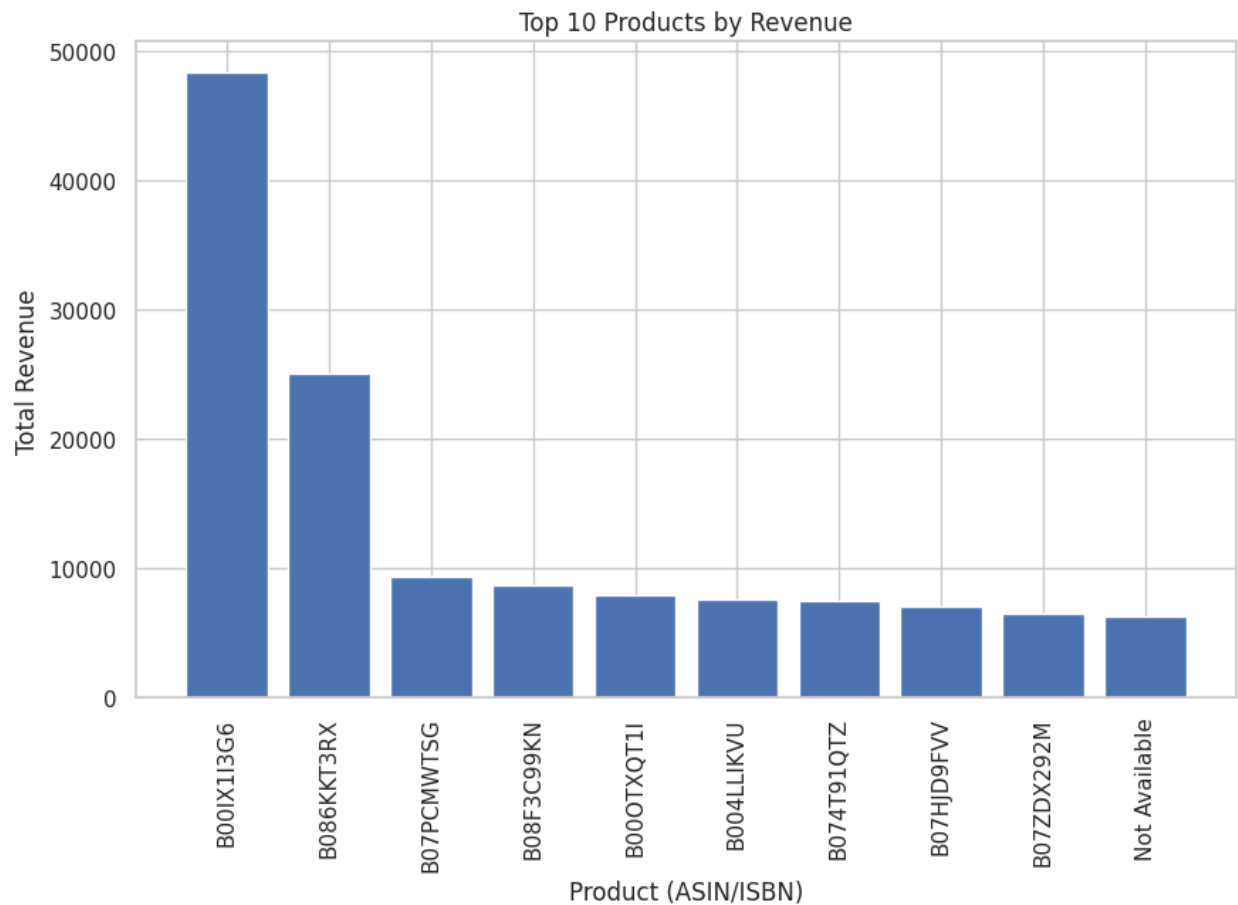
4.2.3 State-wise revenue Distribution:

- *"Assess state-wise revenue to focus on high-growth areas"*

Key Findings:

❖ The Highest growth area was found to be State "**California**" with profitable revenue.

State-wise Revenue Distribution
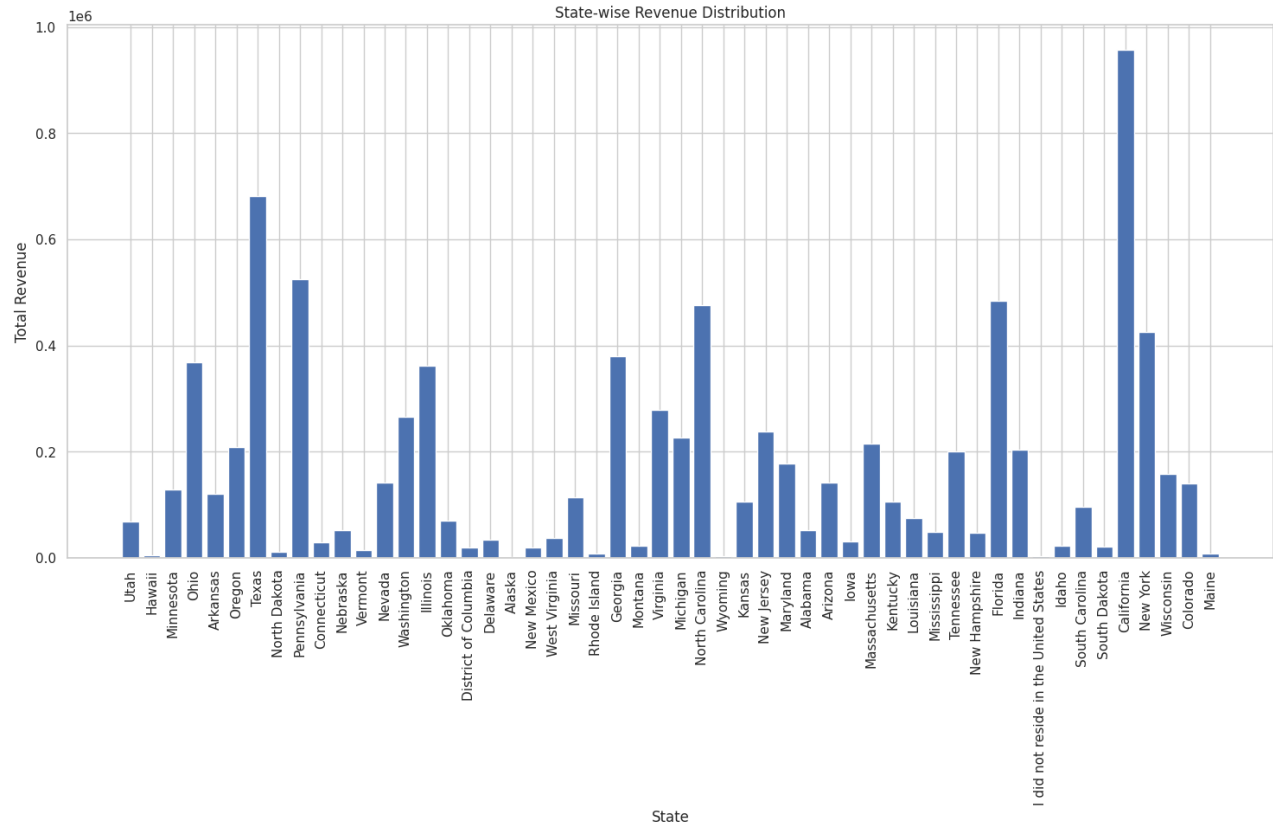
4.2.4 Repeat Purchase Behavior:

- *"Examine repeat purchase behavior to enhance retention initiatives."*

Key Findings:

❖ A significant portion of customers exhibited **Repeat purchase** behavior to enhance retention initiatives.

4.2.5 Flagging Potential Fraud:

- *"Identify irregular transaction patterns to flag potential fraud."*

Key Findings:

❖ We tried to detect some irregular transaction patterns to flag potential fraud. Based on these findings can mitigate financial risks.

4.2.6 Demand Variations across product categories:

- *"Perform inventory management by monitoring demand variations across product categories."*

Key Findings:

❖ Some Top-Category Product like **ABIS Book**, **Cellular Phone**, **Toy Figure**, **Gift Cards** and many mores has more demand for purchase.


Revenue Trends for Top 25 Categories
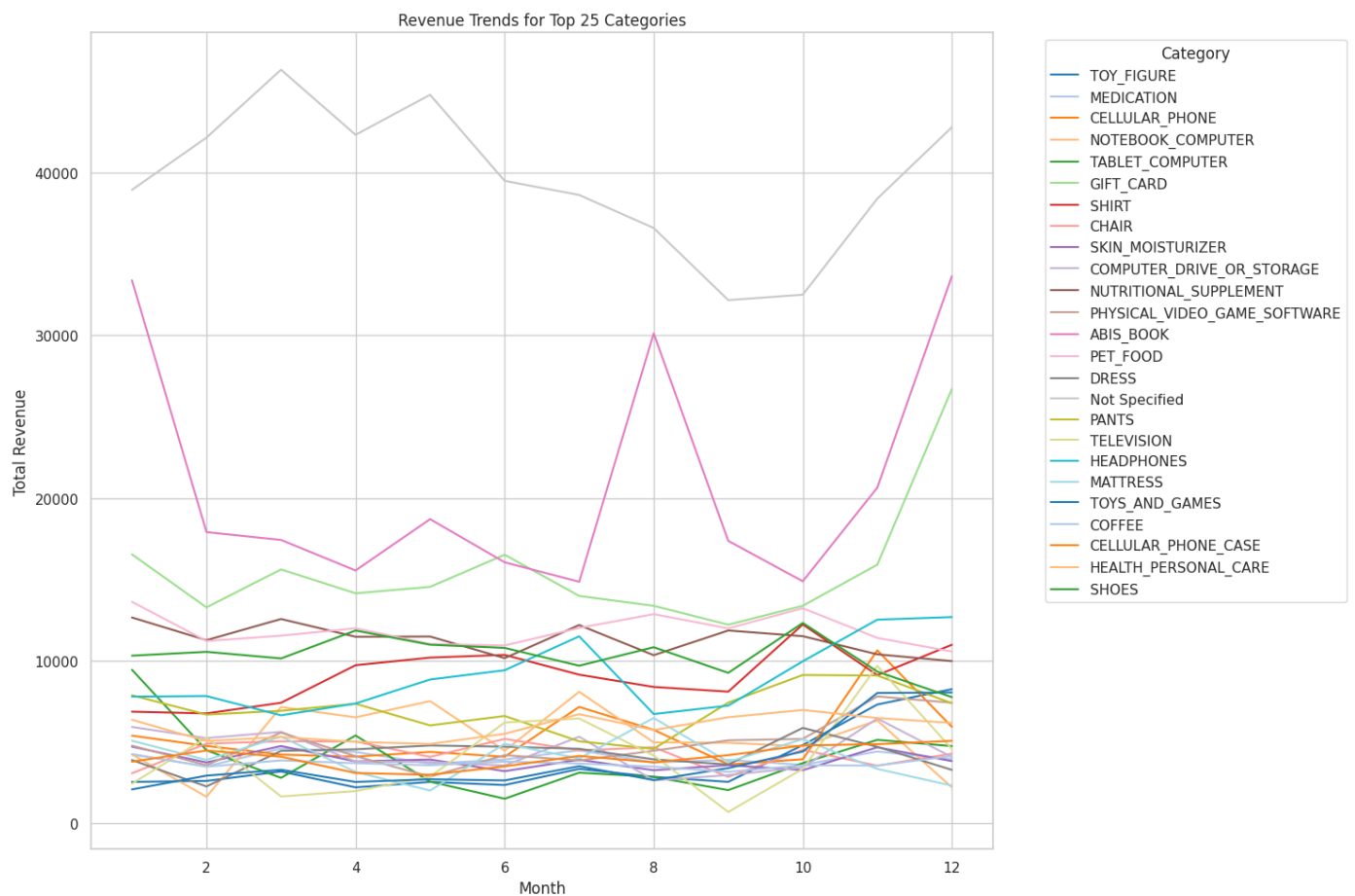
4.2.7 Assess how bulk purchases affect revenue and supply chain operations:

- *"Analyse the impact of how bulk purchasing behavior affects revenue and the overall supply chain operations"*

Key Findings:

- ❖ Categories mostly "**Grocery**" following with **Sugar Candy, Popcorn** and **Snack Food Bar** are some **top Bulk Purchases**.
- ❖ Negotiating better supplier deals and managing inventory levels can further improve margins.

Revenue from Bulk Purchases for Top 25 Categories
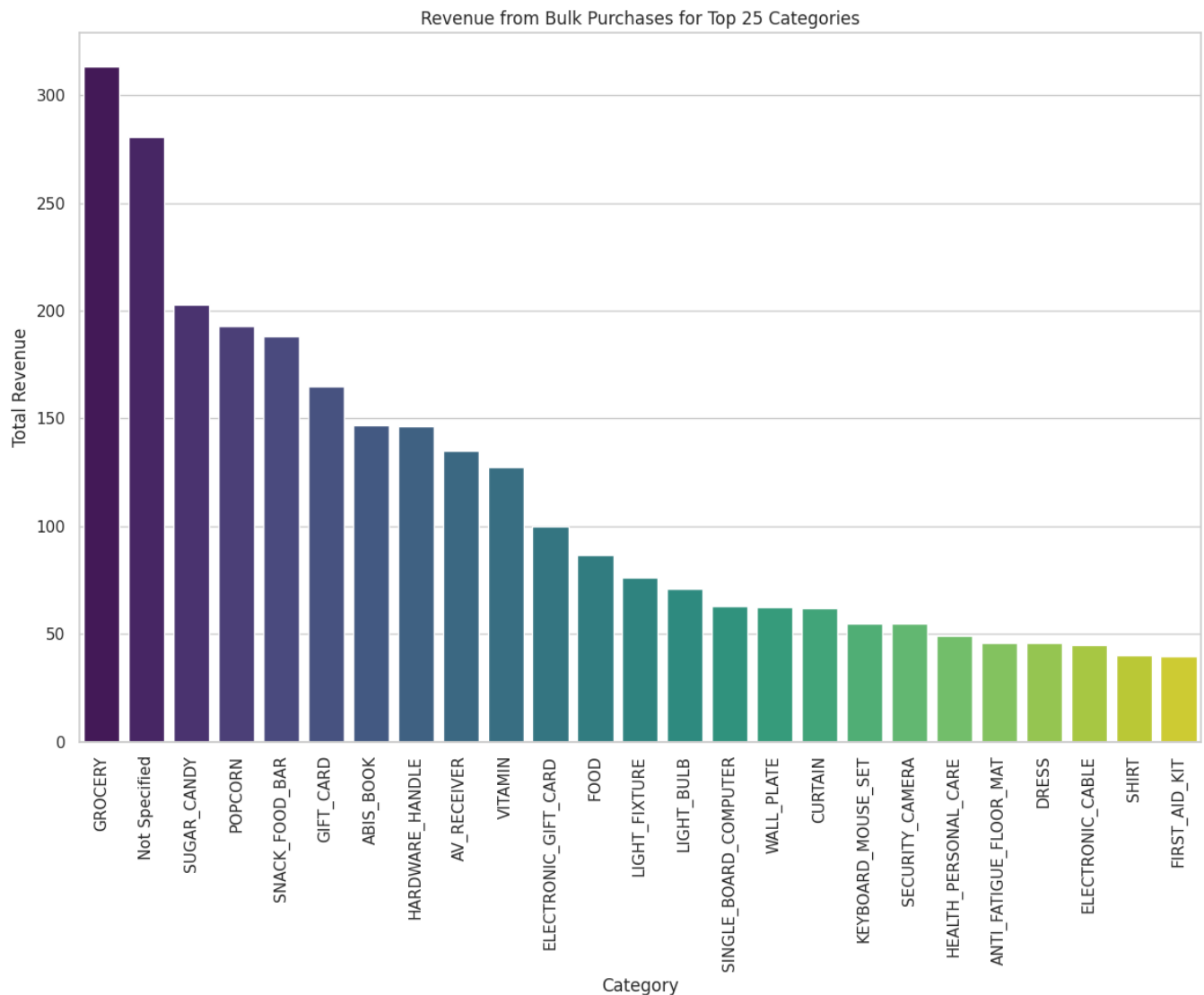
4.2.8 Compare lifecycle strategies:

- *"Compare new and established products to inform and compare lifecycle strategies to make informed decisions."*

Key Findings:

- ❖ The products Launched in Year **2021** showed high revenue eventually **declined** after coming to Year **2023**.
- ❖ Marketing strategies can be adjusted to support new product launches or phase out declining products.

Revenue vs. Launch Year for Products
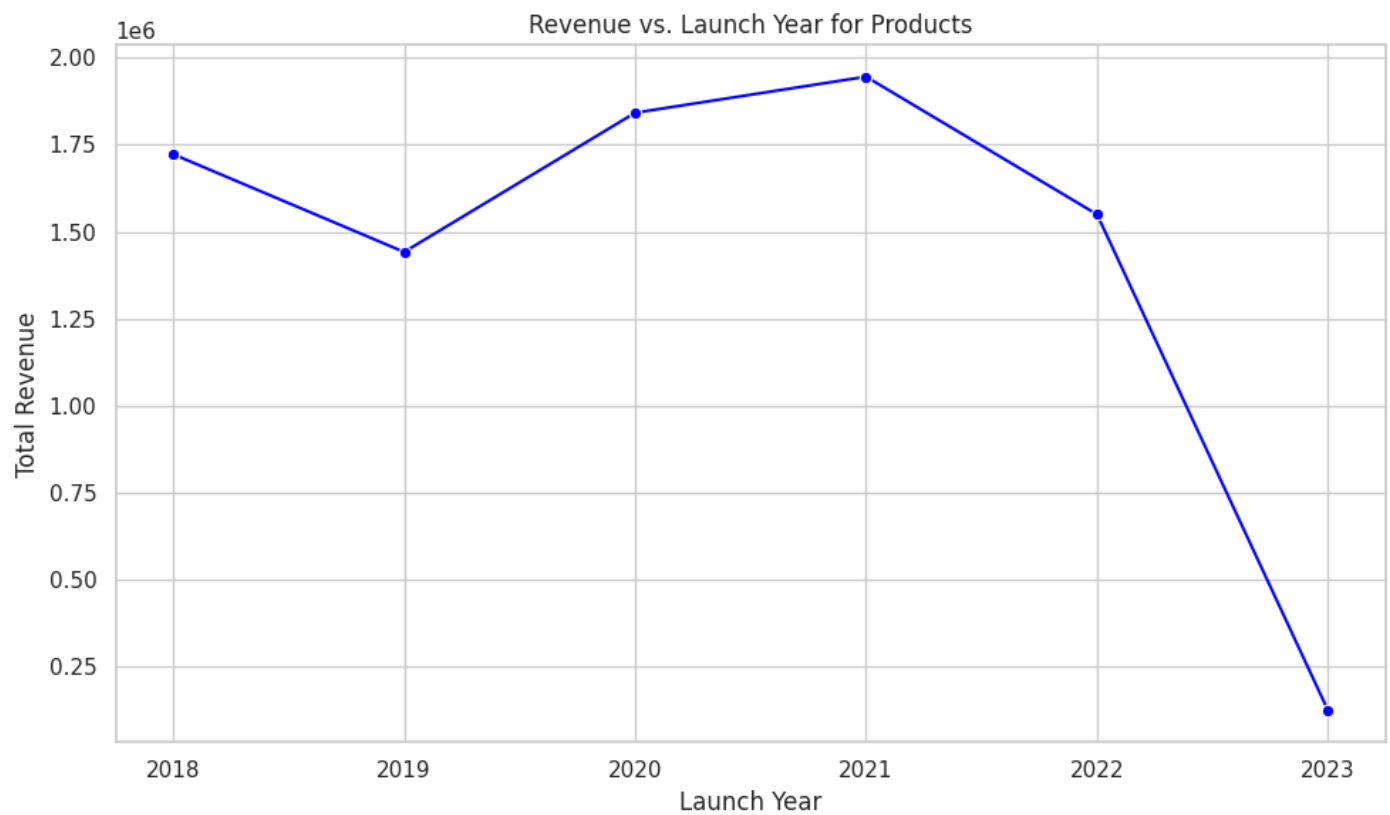
5. Conclusions:

This analysis provided comprehensive insights into customer behavior and identified actionable opportunities to optimize business operations. By leveraging the findings, businesses can improve customer engagement, enhance product performance, and increase overall revenue.

The findings from the study can be summarized as follows:

➢ **Seasonal Trends:** The overall Purchase trends were mostly high in Year **'2020'** and **'2021'** with Month of **December**.

➢ **Gender and Age:** We saw that **Females** with age group of **55-64 years** showed great purchase frequency as compared to other Gender categories.

➢ **Education and Income Level:** Company must prefer the customers having **"prefer not to say"** about their education level with income Level of $**50,000-74,999** makes highest number of Purchases.

➢ **State and Race:** The people living in **California** and belongs to **White or Caucasian** showed highly interest in purchasing products.

➢ **Weekdays vs. Weekends: Weekday** analysis revealed higher sales volumes on **Monday**, suggesting optimal days for promotional campaigns moreover on weekdays.

➢ The **spending KPIs** for E-commerce Company was found to be around **8222.85**.

➢ **Repeat Purchase:** A significant portion of customers exhibited **repeat purchase** behavior to enhance retention initiatives.

➢ **Top Engaged Customer:** Some top customers had high Purchase frequency and must be provided with regular discount and offers for being loyal customer to company.

- ➢ **Top-Selling Products:** The Product code with **B00IX1I3G6** followed by some top 10 products can help in revenue growth by increasing their sales.

- ➢ **Potential Fraud:** Detected some irregular transaction patterns to flag potential fraud. Based on these findings can mitigate financial risks.

- ➢ **Top-Category Product:** Product category like **ABIS Book, Cellular Phone, Toy Figure, Gift Cards** and many more has great demand for purchase.

- ➢ **Top-Bulk Purchase:** Categories mostly "**Grocery**" following with **Sugar Candy, Popcorn** and **Snack Food Bar** are some **top Bulk Purchases**. Negotiating better supplier deals and managing inventory levels can further improve margins.

- ➢ **Lifecycle Strategies:** Products Launched in Year **2021** showed high revenue eventually declined after coming to Year **2023.** Marketing strategies can be adjusted to support new product launches or phase out declining products.

This analysis provides actionable insights for improving business performance, increasing customer satisfaction, and driving revenue growth.