Customer Purchase Behavior – Samuel Suarez – E-commerce Data Analysis

This project analyzes simulated e-commerce transaction data to uncover key trends in customer purchasing behavior. Using tools like SQL, Google Sheets, and Tableau, it identifies top-performing product categories, seasonal sales patterns, and regional customer engagement differences. The insights lead to actionable recommendations to improve marketing efforts, checkout experience, and customer segmentation.

---

1. Project Overview

Business Task:

Analyze customer purchase behavior in an e-commerce environment to uncover trends that can help improve sales strategies and customer experience.

---

2. Data Sources & Tools

Data Source:

This project uses a simulated dataset representing customer transactions and behavior within an e-commerce platform. It includes variables such as order date, product category, state, coupon usage, payment method, sales amount, and cart abandonment.

Tools Used:

- **Google Sheets** – Used for organizing data, creating pivot tables, and building visualizations.
- **Google BigQuery (SQL)** – Used for querying large-scale data to extract specific metrics and patterns.
- **Tableau Public –** Used to create interactive and professional-level visualizations to support data storytelling and insight presentation.

---

3. Data Cleaning Process

The dataset was cleaned and prepared to ensure accurate analysis:

- Converted data types (e.g., date fields and numeric values) for proper aggregation.
- Removed duplicates and irrelevant records.
- Handled missing or invalid entries by excluding them from calculations where appropriate.
- Standardized naming conventions for product categories and U.S. state labels.
- Ensured sales amounts were properly cast to numeric format using SQL.

---

4. Analyze Phase – Exploration & Queries

To explore and interpret the data, **Google Sheets**, **Google BigQuery (SQL)**, and **Tableau Public** were used in combination.

Using Google Sheets:

- Created pivot tables to summarize data by product category, payment method, and coupon usage.
- Developed bar and column charts to visualize:
    - Total sales by month
    - Number of orders by product category
    - Average amount spent by product category
    - Average spending: coupon users vs. non-users
    - Number of orders by payment method

Using Google BigQuery (SQL):

- Wrote queries to calculate:
    - Total orders by state
    - Shopping cart abandonment rate by state
    - Total monthly revenue
    - Average order value per state
- Queried and transformed raw data for advanced metrics, which were then visualized in Google Sheets.

Using Tableau Public:

- Created visually enhanced bar charts to showcase:
    - Total sales by month
    - Total orders by state
    - Shopping cart abandonment rate by state
    - Average amount spent by state

Each step provided a clearer understanding of customer behavior and revealed performance differences across segments like state, product category, and payment method.

To generate these insights, several SQL queries were written in Google BigQuery. Below are a few key examples:

1. Abandonment Rate by State:

```
SELECT
  string_field_4 AS State,
  ROUND(SUM(CASE WHEN string_field_7 = '
FROM
  ecommerce_customer_data
WHERE
  string_field_4 != 'State'
GROUP BY
  string_field_4
ORDER BY
  AbandonmentRate DESC;
```

2. Total Orders by State:

```
SELECT
  string_field_4 AS State,
  COUNT(*) AS TotalOrders
FROM
  ecommerce_customer_data
GROUP BY
  State
ORDER BY
  TotalOrders DESC;
```

3. Total Sales by Month:

```
SELECT
  FORMAT_DATE('%B', PARSE_DATE('%m/%d/%Y
  ROUND(SUM(SAFE_CAST(string_field_8 AS
FROM
  ecommerce_customer_data
WHERE
  SAFE_CAST(string_field_8 AS FLOAT64) I
GROUP BY
  Month
ORDER BY
  TotalSales DESC;
```

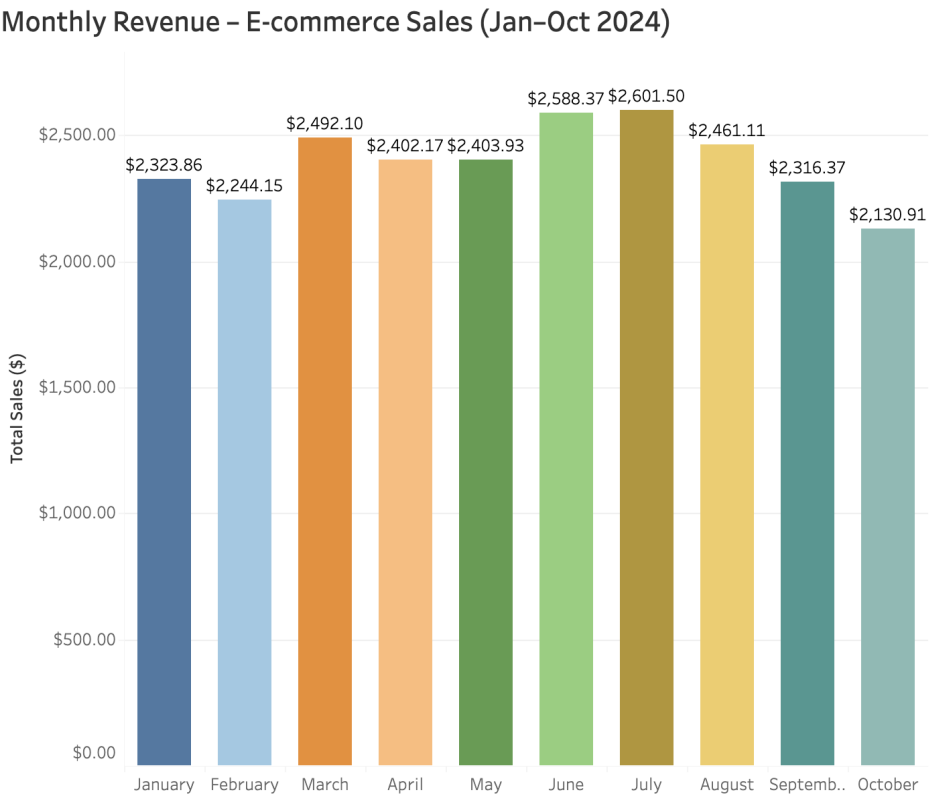
4. Average Amount Spent by State:

```
SELECT
  string_field_4 AS State,
  ROUND(AVG(SAFE_CAST(string_field_8 AS
FROM
  ecommerce_customer_data
WHERE
  SAFE_CAST(string_field_8 AS FLOAT64) I
GROUP BY
  State
ORDER BY
  AverageAmountSpent DESC;
```

5. Share Phase – Visualizations & Insights

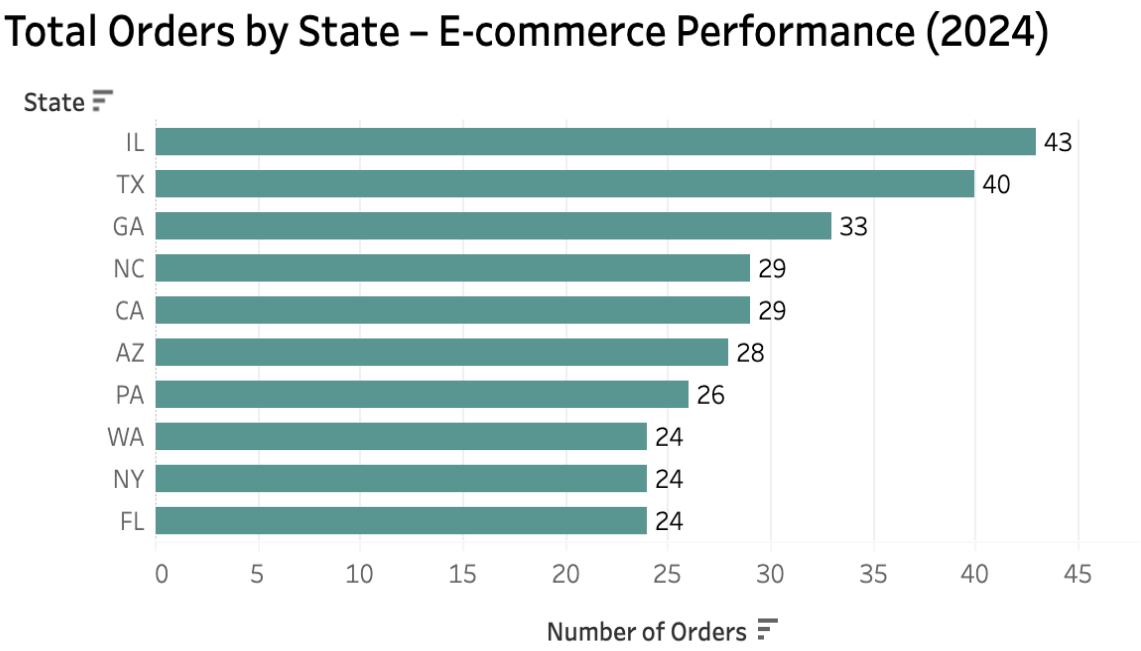
The following visualizations were created using Google Sheets:

- Total Sales by Month



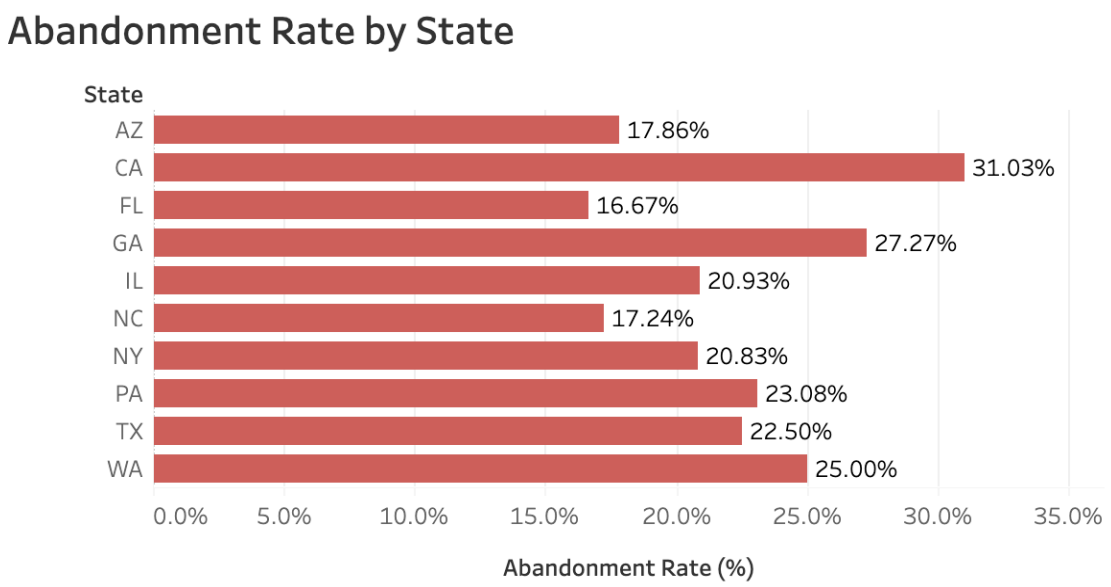
Findings: The month with the highest revenue was **July**, with a total of **$2,601.50**, followed closely by **June** ($2,588.37). **March** also performed strongly with **$2,492.10**, while **October** recorded the lowest revenue at **$2,130.91**. There is a clear upward trend from February to July, possibly driven by seasonal campaigns or increased user activity during that period. Starting in August, revenue shows a gradual decline, which may require further analysis related to seasonality or changes in marketing strategy.

- Total Orders by State



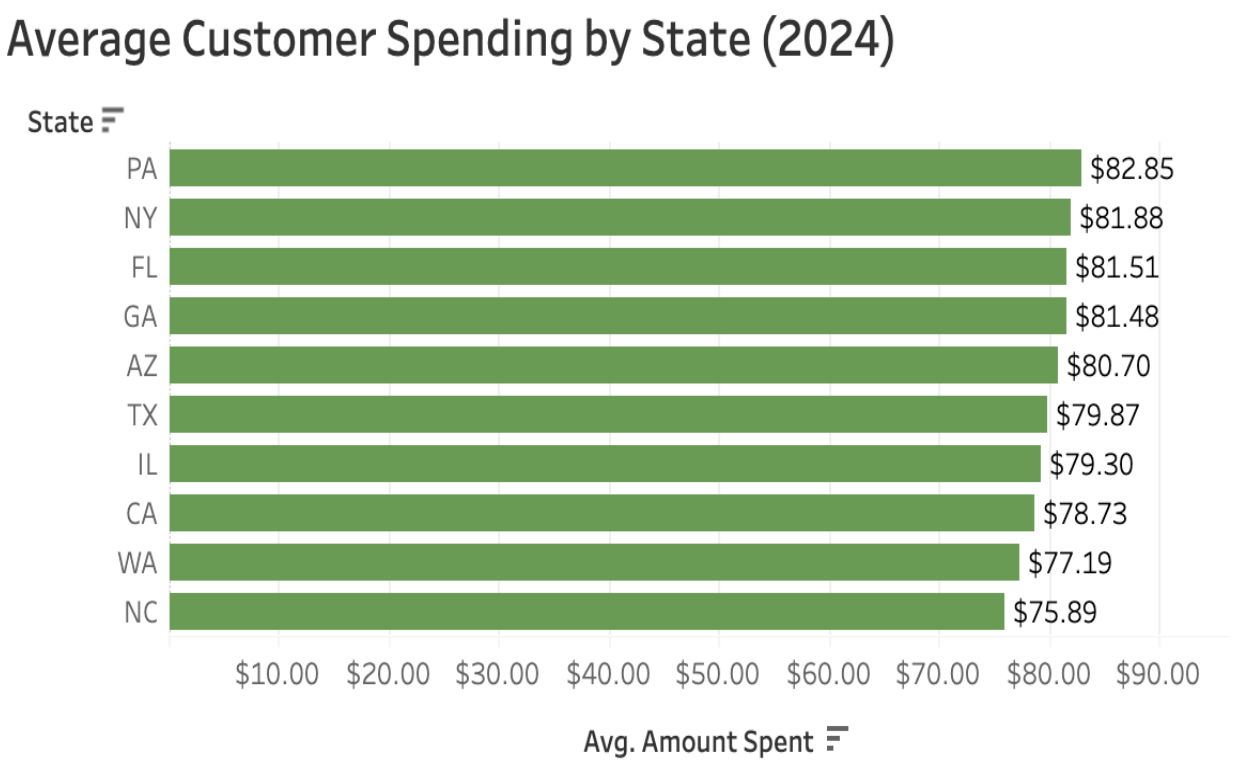
Findings: Illinois had the highest number of orders with **43**, followed by Texas (**40**) and Georgia (**33**), indicating strong customer engagement in these states. States like Florida, New York, and Washington had the lowest counts, each with **24 orders**, suggesting potential opportunities for regional marketing or improved outreach. The data highlights a concentration of e-commerce activity in the Midwest and South, while other areas show a more moderate performance.

- Shopping Cart Abandonment Rate by State



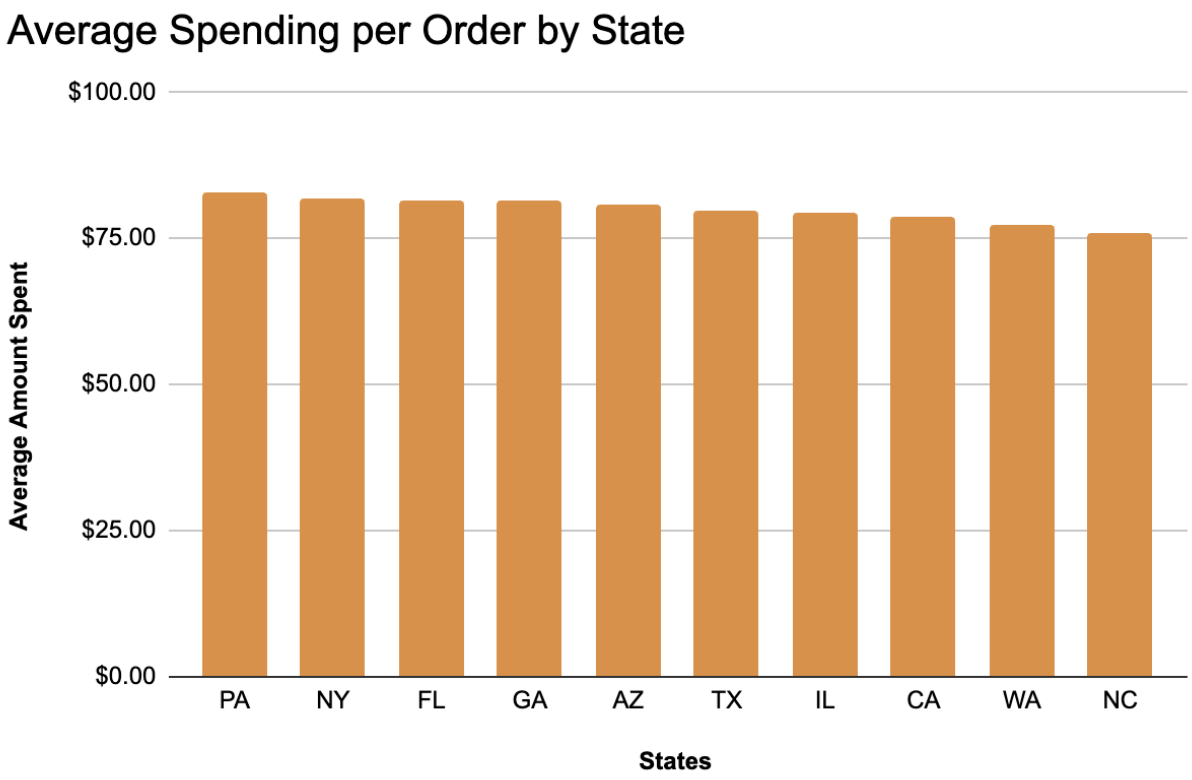
California recorded the highest shopping cart abandonment rate at 31.03%, while Florida had the lowest at 16.67%. This suggests notable regional differences in customer behavior, which could be influenced by factors such as user experience, localized marketing strategies, or consumer intent in each state.

- Average Amount Spent by State



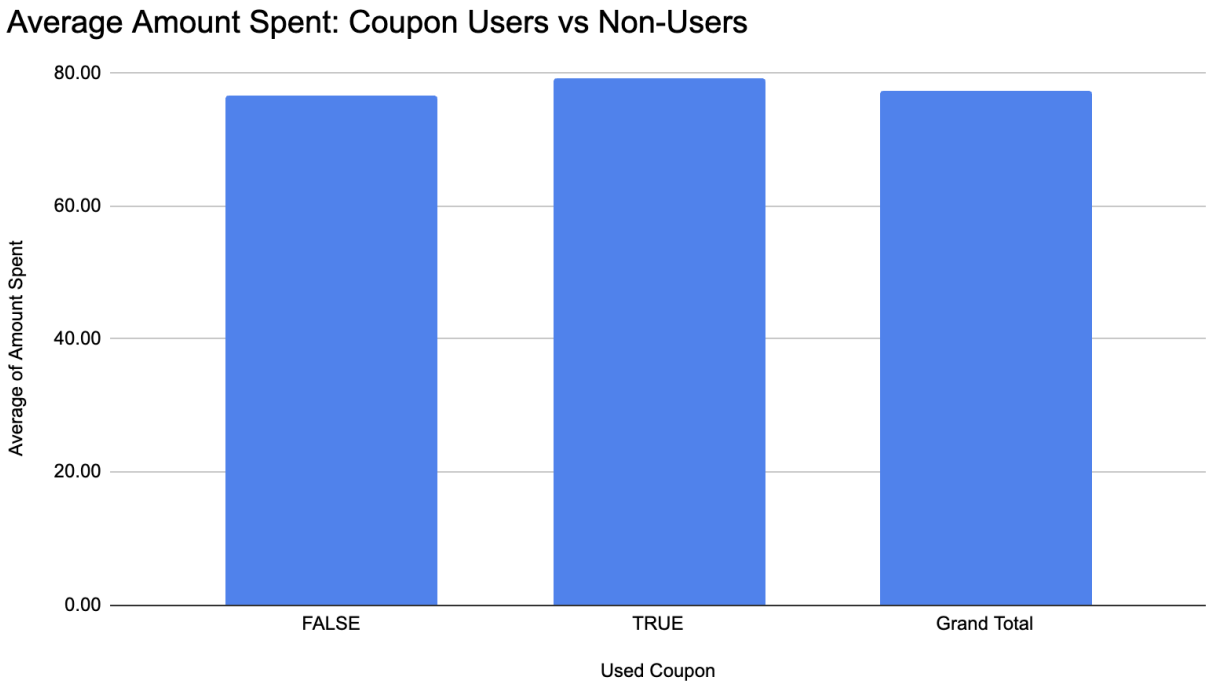
On average, customers in Pennsylvania ($82.85), New York ($81.88), and Florida ($81.51) spent the most per order, while North Carolina ($75.89) and Washington ($77.19) had the lowest average spending, highlighting notable differences in consumer purchasing behavior across states.

- Average Amount Spent by Product Category



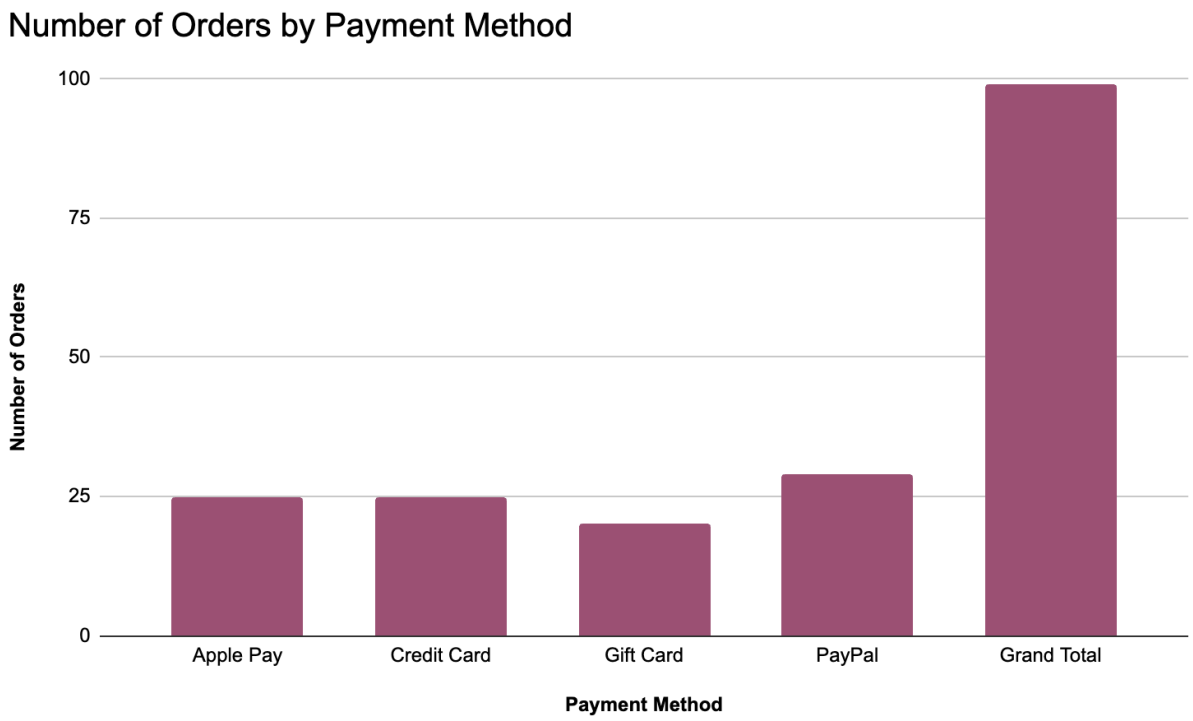
Pennsylvania, New York, and Florida lead in average customer spending per order, all exceeding $80, indicating stronger purchasing behavior in these states. In contrast, states like California, Washington, and North Carolina show slightly lower average spending, suggesting potential opportunities to enhance customer engagement or value per transaction in those regions.

- Average Amount Spent: Coupon Users vs. Non-Users



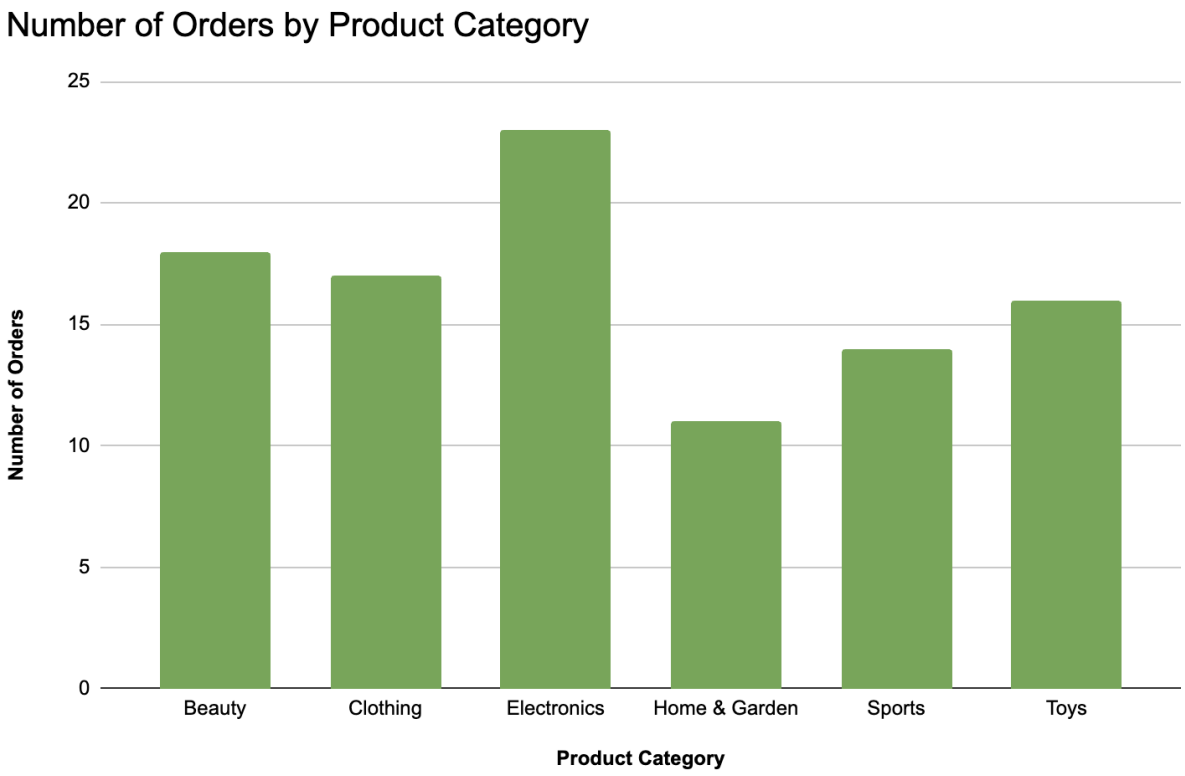
Customers who used coupons spent slightly more on average than those who did not, suggesting that coupon strategies may encourage higher spending behavior rather than just attracting discount-driven purchases.

- Number of Orders by Payment Method



PayPal was the most used payment method, followed closely by Apple Pay and Credit Card, which had the same number of orders. Gift Cards were the least used, indicating a preference for digital and traditional payment methods over prepaid options.

- Number of Orders by Product Category



Electronics had the highest number of orders, indicating strong consumer demand in that category, followed by Beauty and Clothing. In contrast, Home & Garden received the fewest orders, suggesting lower customer interest or seasonal factors affecting that segment.

Each chart includes a short description and insight to support decision-making. Example insights:

- Electronics had the highest number of orders.
- California had the highest cart abandonment rate.
- Coupon users spent slightly more on average.

6. Act Phase – Recommendations

Based on the analysis, the following recommendations can be made:

- Invest more in high-performing product categories such as Electronics.
- Review and improve the checkout process in states with higher abandonment rates.
- Optimize discount campaigns, as coupon users showed slightly higher spending.
- Ensure a smooth experience with PayPal and credit cards, the most-used payment methods.

7. Limitations

While this project provides valuable insights into customer purchase behavior, it is important to consider the following limitations:

● Simulated Dataset: The data used is not from a real-world e-commerce platform. While it reflects realistic patterns, the insights may not translate perfectly to actual business scenarios.

● No Customer Demographics: The dataset lacks key customer attributes such as age, gender, or acquisition channel, which limits the ability to perform segmentation or personalization strategies.

● Time Range Constraints: The analysis covers a limited time period (January–October), which may not fully capture year-round seasonal trends such as holiday spikes in November and December.

● No Product-Level Granularity: Product categories are used instead of individual product SKUs, which limits the depth of product-specific insights and inventory optimization opportunities.

● Assumed Data Quality: Although cleaning was performed, assumptions were made regarding data completeness and accuracy that could impact results.

These limitations should be considered when interpreting the findings and applying the recommendations to real-world strategies.

Appendix – Resources & Interactive Dashboards

Tableau Dashboard:
https://public.tableau.com/app/profile/samuel.suarez.abreu/viz/E-commerceSalesAnalysisJanOct2024/RevenuebyMonth

📋 Google Sheets Source File:
https://docs.google.com/spreadsheets/d/1CSIvQcqYNJGPaZyFzIPIsOA38jW8GdVsAjLAUF-L-8I/edit?usp=sharing

💻 SQL Queries (Google BigQuery):
Available upon request