Sales Data Analysis

1. Project Description

The project concerns the analysis of sales data from 2019, specifically focusing on product order details. Various aspects of sales were analyzed, including the number of orders, the total value of orders, product popularity, and regional distribution. The main goal of the analysis was to uncover key sales trends, identify popular products, analyze regional sales performance, and investigate sales seasonality.

2. Dataset Description

The dataset consists of 12 separate files, each containing data for a specific month in 2019. Upon merging these files into a single dataset, it was discovered that the data included six main columns:

- Order ID – Unique identifier for each order
- Product – Name of the product ordered
- Quantity Ordered – Number of units of the product ordered
- Price Each – Price per unit of the product
- Order Date – Date and time when the order was placed
- Purchase Address – Customer's location (including city, state, and zip code)

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

3. Data Cleaning

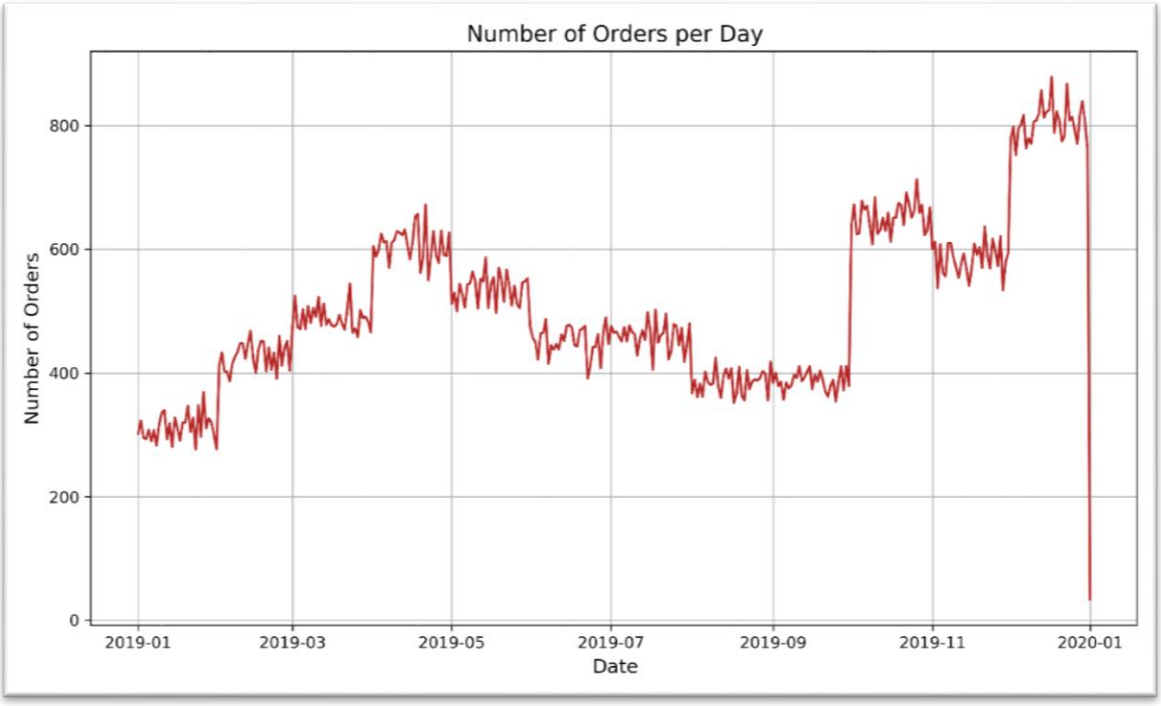
After combining all the data into a single file, a series of data cleaning tasks were performed. Empty rows and rows containing erroneous data were removed. The Order Date column was split into two separate columns: Date and Time, to facilitate temporal analysis. The Purchase Address column was also divided into three columns: City, State, and Zip Code. Duplicate entries were resolved by aggregating the quantities for the same product in the same order. Finally, all columns were checked for incorrectly entered or invalid values. The final cleaned dataset consists of 185,639 rows and 10 columns: Order ID, Product, Price per Item, Date, Time, City, State, Zip Code, Quantity Ordered, and Total Price.

	Order ID	Product	Price Each	Date	Time	City	State	Zip Code	Quantity Ordered	Total Price
0	141234	iPhone	700.00	2019-01-22	21:25:00	Boston	MA	02215	1	700.00
1	141235	Lightning Charging Cable	14.95	2019-01-28	14:15:00	Portland	OR	97035	1	14.95
2	141236	Wired Headphones	11.99	2019-01-17	13:33:00	San Francisco	CA	94016	2	23.98
3	141237	27in FHD Monitor	149.99	2019-01-05	20:33:00	Los Angeles	CA	90001	1	149.99
4	141238	Wired Headphones	11.99	2019-01-25	11:59:00	Austin	TX	73301	1	11.99

4. Data Exploration

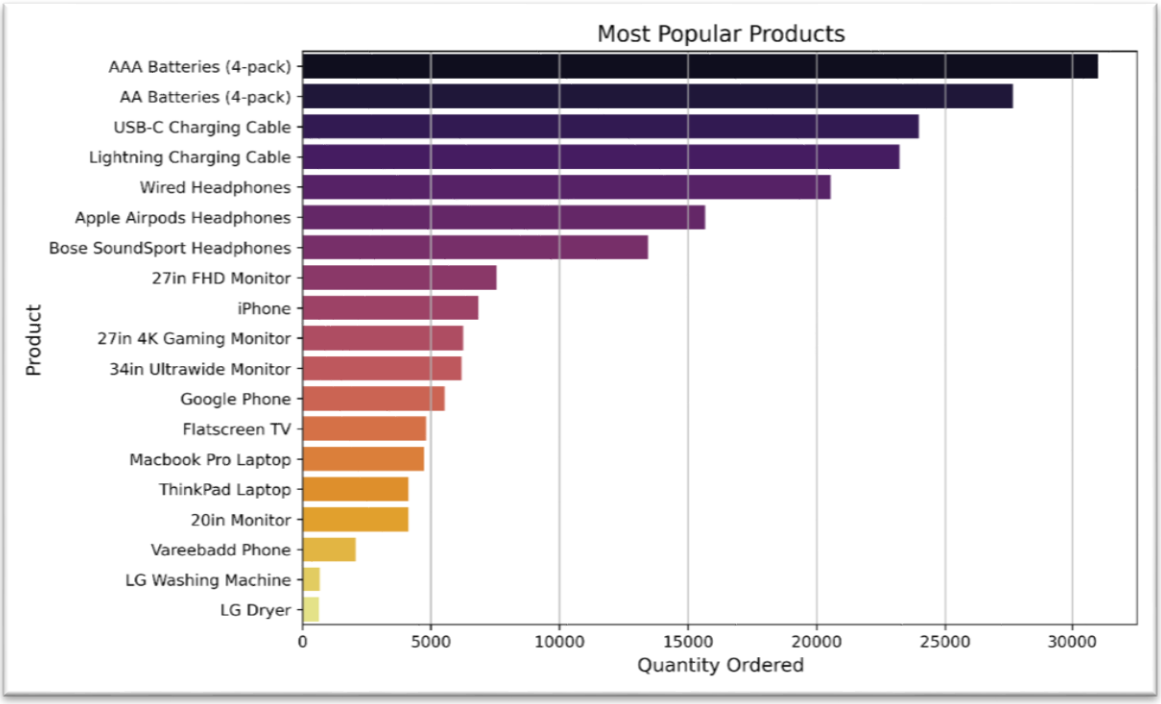
During the exploratory data analysis, several interesting observations were made. The Product column contains 19 unique products, while the State and City columns include 8 and 9 unique values, respectively. Additionally, the average shopping cart contained 1.17 products, and the average total price per cart was $193.30. Building on these observations, several visualizations were created to provide deeper insights into the data and highlight key trends. The following charts present a closer look at various aspects of the sales, including the popularity of products, sales distribution across different regions, and trends over time.

To begin the analysis, a time-based chart was created to observe daily order volume throughout the year.



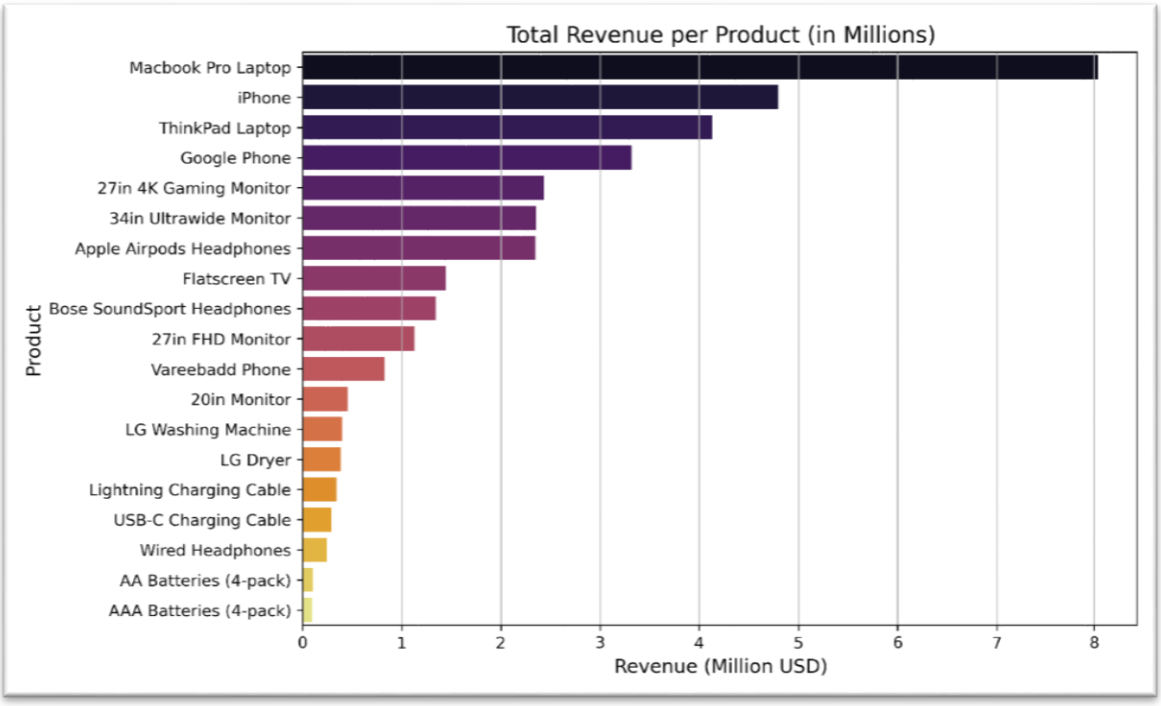
The graph illustrates the daily number of orders, highlighting variations in sales across different months.

Next chart visualizes the most frequently ordered products, highlighting the number of units ordered for each product. This provides a clear view of which products dominated the sales in 2019, with the most popular products being easily identifiable based on their quantity ordered.



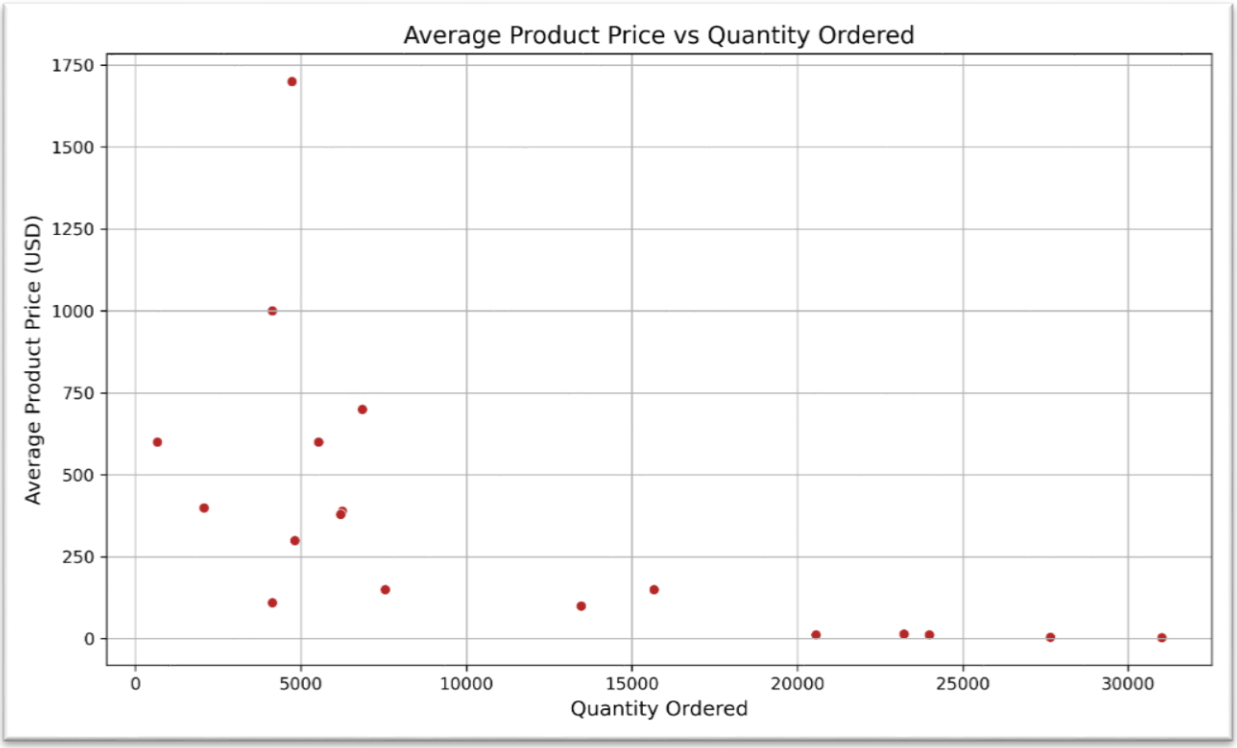
As shown in the chart, batteries were the most frequently ordered products, followed by chargers and headphones. This trend suggests that accessories for more expensive devices are frequently purchased alongside the main products.

The next chart illustrates the total revenue generated by each product.



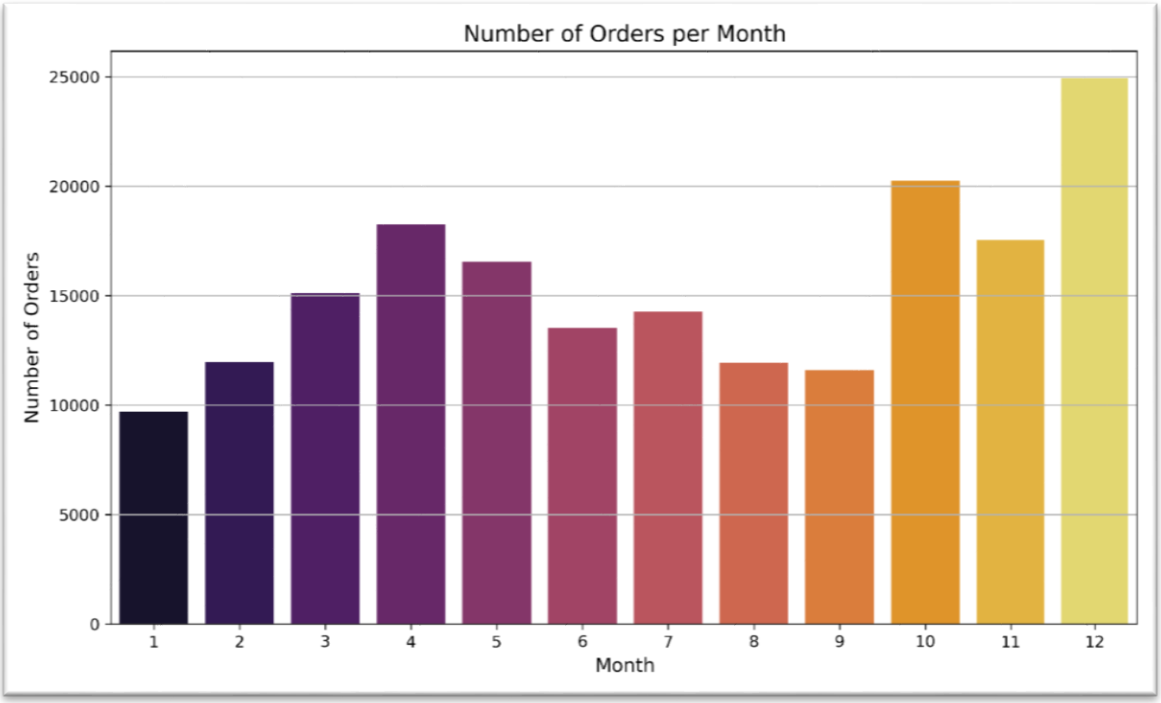
The chart clearly shows that the highest revenue came from the MacBook Pro laptop, with the top four positions also occupied by laptops and smartphones.

Additionally, the third chart illustrates the relationship between the average product price and the quantity of products ordered.



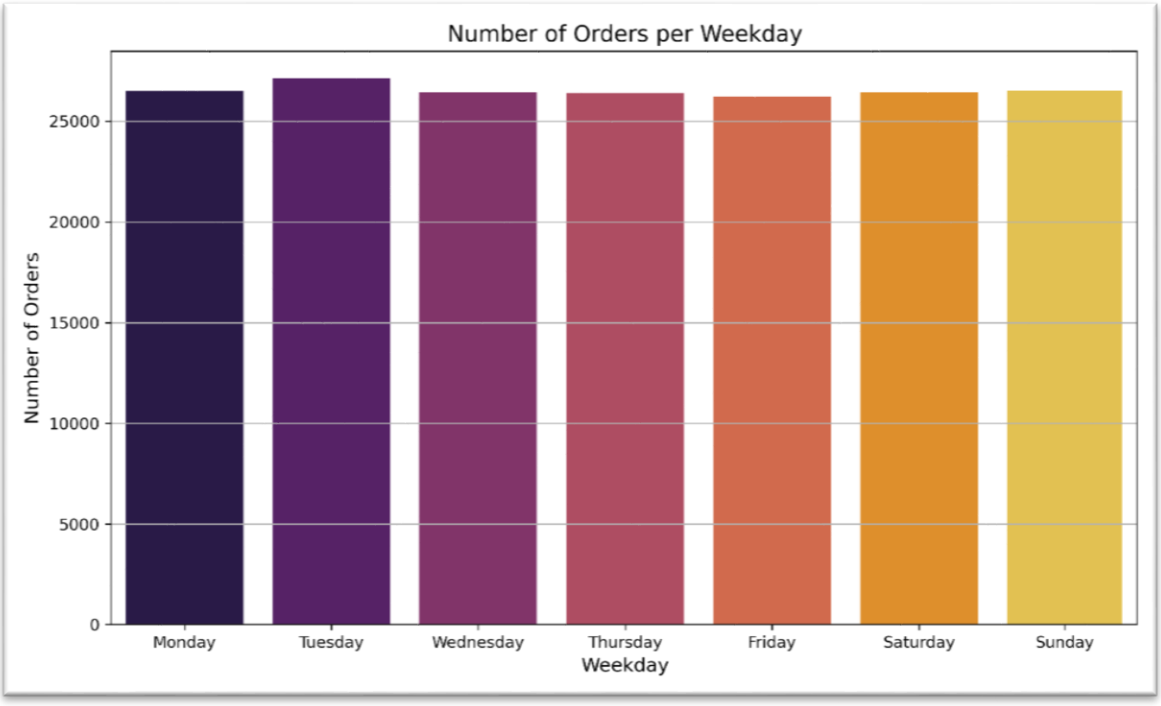
It shows that the cheaper the product, the higher the quantity ordered.

The next part of the report focuses on the analysis of data over time. The fourth chart illustrates the number of orders in each month.



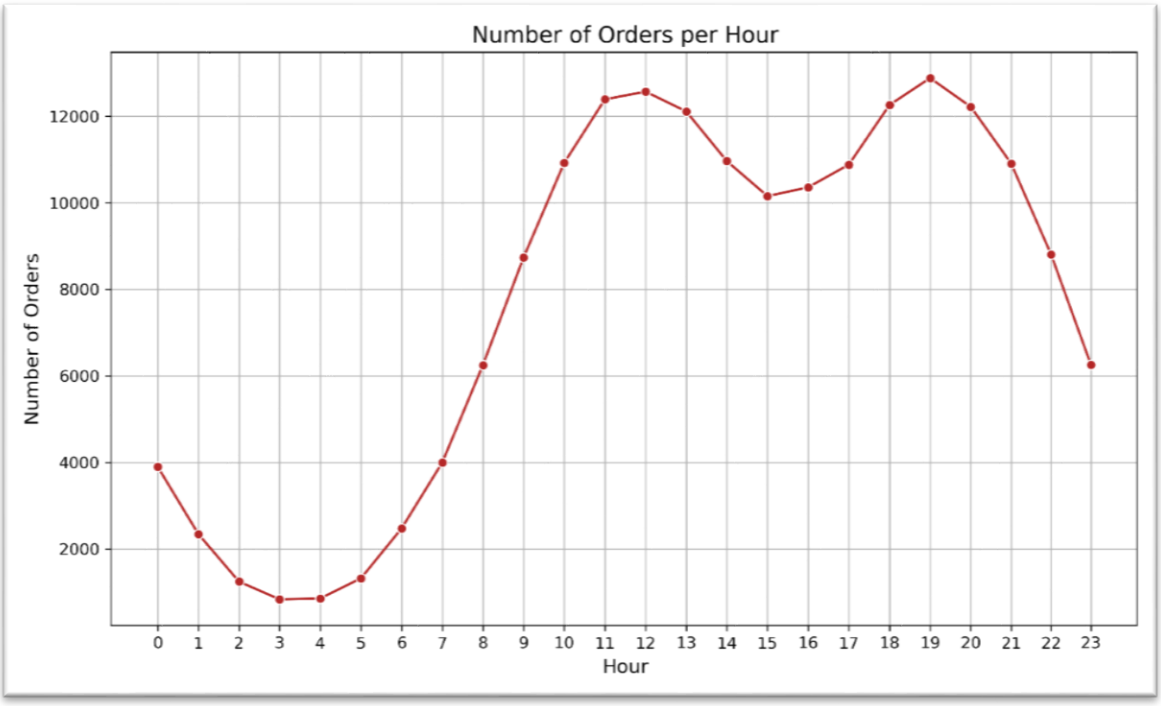
It shows that the highest number of orders occurred in December, followed by October, November, and April.

The next chart illustrates the number of orders broken down by weekdays.



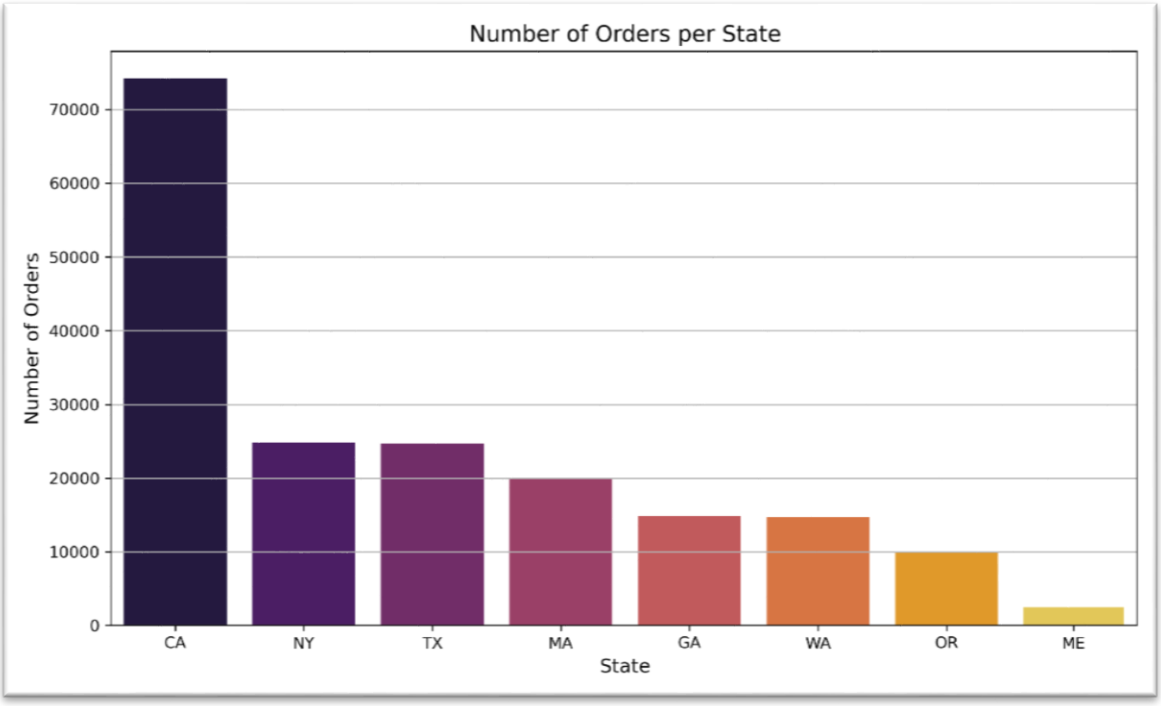
It shows that the highest number of orders occurred on Tuesdays, although overall, the number of orders was quite similar across all days of the week.

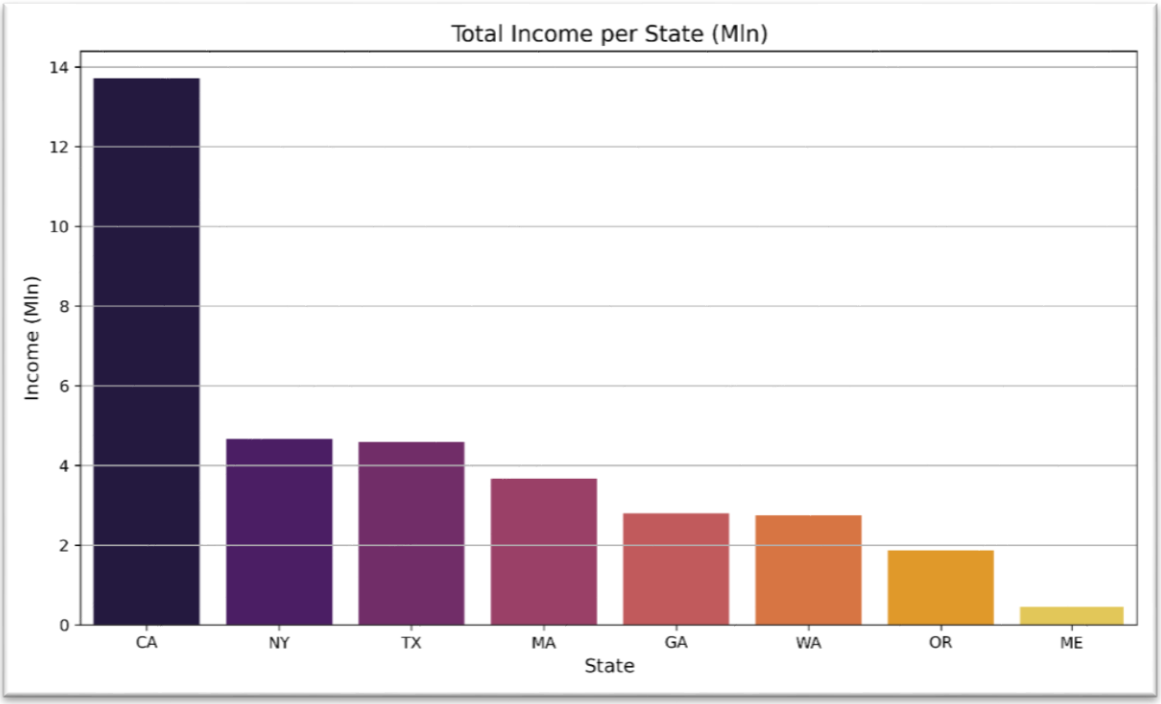
The next chart illustrates the number of orders broken down by hour of the day.



It shows two peaks in the number of orders around 12:00 PM and 7:00 PM. Additionally, most orders were placed between 8:00 AM and 11:00 PM, while the fewest orders occurred around 3:00-4:00 AM.

The next charts illustrate the number of orders and revenue in each state.





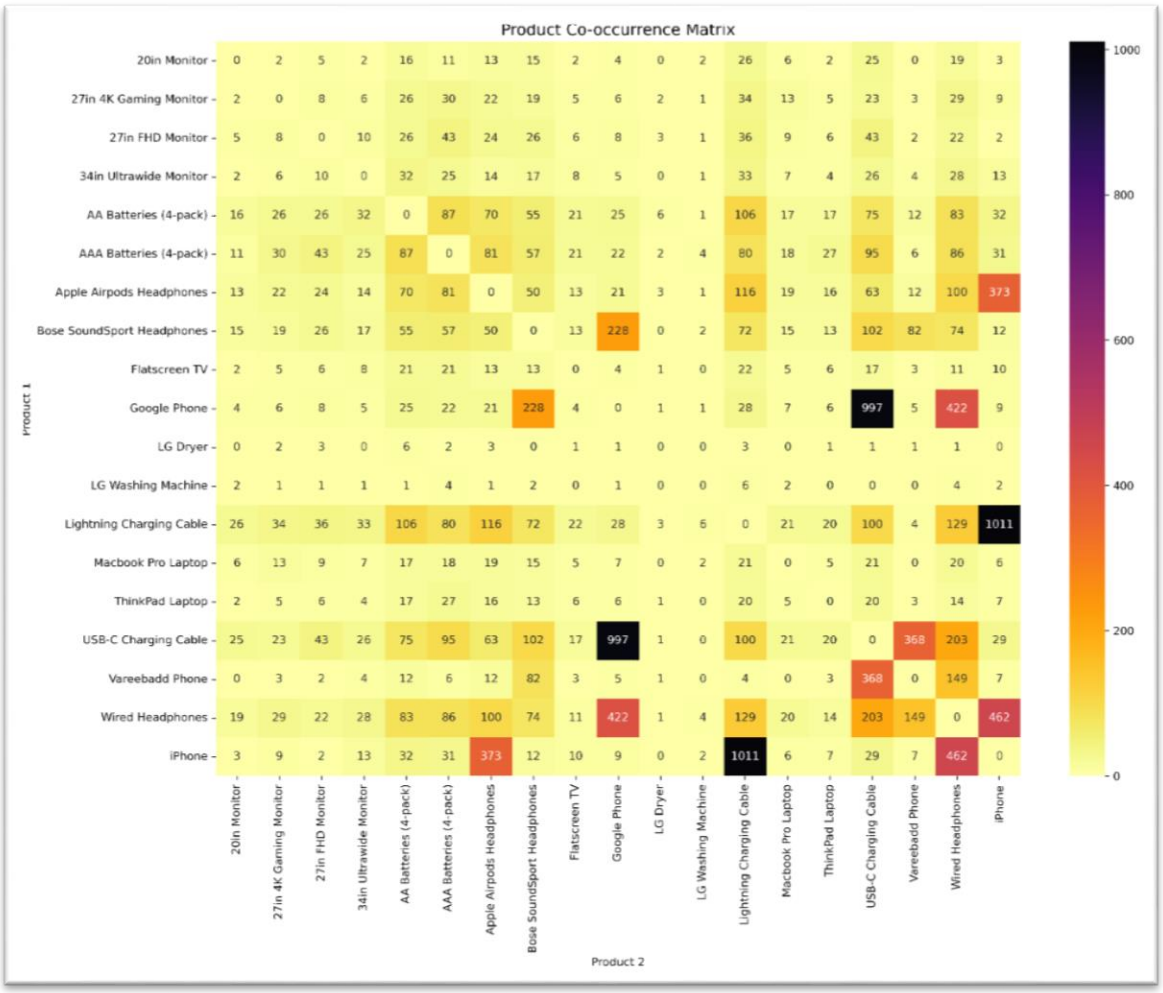
These charts are quite similar, showing that the number of orders directly correlates with revenue. States with the highest number of orders also had the highest revenue. California stands out significantly, surpassing all other states in both orders and revenue. On the other hand, Maine had the lowest revenue and number of orders.

The following table summarizes how many unique products were found in individual shopping carts, how often such carts occurred, and what percentage of the total they represented:

Unique Products in Cart	Number of Carts	Percentage of Total Carts
1	171605	96,17%
2	6479	3,63%
3	337	0,19%
4	15	0,008%
5	1	0,0006%

Clearly, the majority of shopping carts contained just one product.

From the chart, we can observe that certain product pairs appear frequently in orders, indicating that customers tend to purchase these items together more often than others. This information could be useful for targeted marketing or product bundling strategies.



The matrix reveals that the most frequent product pairs are the iPhone and Lightning Charging Cable, as well as the Google Phone and USB-C Charging Cable. Additionally,

phone headsets were also quite popular. These products could be recommended as complementary items for phones, potentially increasing sales through product bundling or targeted marketing.

5. Summary

Based on the analysis of sales data from 2019, several key trends and interesting insights can be observed. The product popularity analysis shows that the most frequently ordered items were accessories such as batteries, chargers, and headphones, suggesting that consumers often choose accessories to complement more expensive devices.

In terms of revenue, laptops, particularly the MacBook Pro model, had the most significant impact on total income. The analysis of the relationship between product price and the number of orders reveals that cheaper products received more attention and had higher order volumes.

From a seasonality perspective, data shows the highest number of orders occurred in December, October, November, and April, which may correlate with the holiday season and the beginning of the school year.

Regarding the timing of orders, the majority of transactions occurred around 12:00 PM and 7:00 PM, suggesting increased user activity during lunch breaks and evenings. The least amount of orders was recorded between 3:00 AM and 4:00 AM.

Regional analysis revealed that California was the clear leader in terms of both order volume and revenue, while Maine recorded the lowest revenue and order volume. Additionally, the average shopping cart consisted of 1.17 products, and the average revenue per cart was $193.30, providing insights into typical consumer spending.

An interesting finding was the analysis of product pairs. The most frequently ordered sets were iPhone and Lightning Charging Cable, as well as Google Phone and USB-C Charging Cable. Headphones for phones were also popular. These data can serve as a basis for recommendations regarding product bundling, which could contribute to increased sales.

The insights from this analysis allow for a better understanding of consumer preferences and the identification of products that could be more effectively promoted together or offered in bundles, potentially leading to increased sales and optimization of marketing strategies.