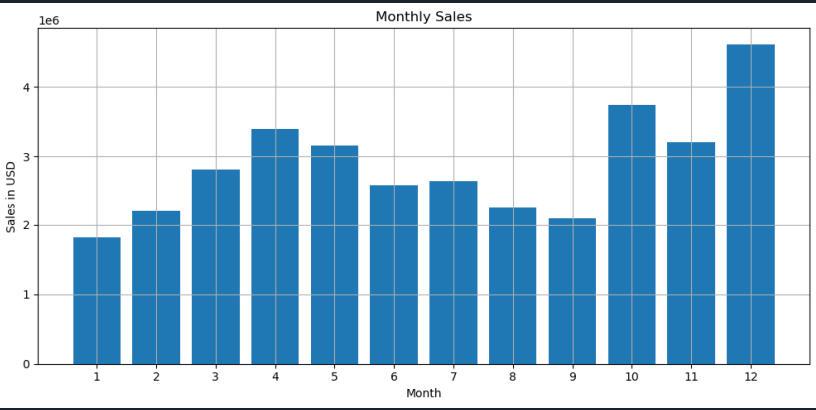
Sales Data Analysis – Exploratory Data Visualization Project

This project presents an exploratory data analysis (EDA) of a sales dataset using Python. The goal of this analysis is to extract meaningful insights from raw transactional data through visualizations. The project emphasizes trends in sales volume, consumer behavior, top-selling products

The dataset used in this analysis consists of sales records from multiple cities, capturing information such as purchase time, product, quantity, and price. Using Python libraries such as Pandas and Matplotlib, we cleaned, processed, and visualized the data to answer several key business questions:

1. Monthly Sales Distribution



Analysis:

This bar graph displays total monthly sales (in USD) across the year. The x-axis represents months (1 to 12), and the y-axis shows sales revenue in millions.

Observations:

- December recorded the highest sales, followed by October and April.

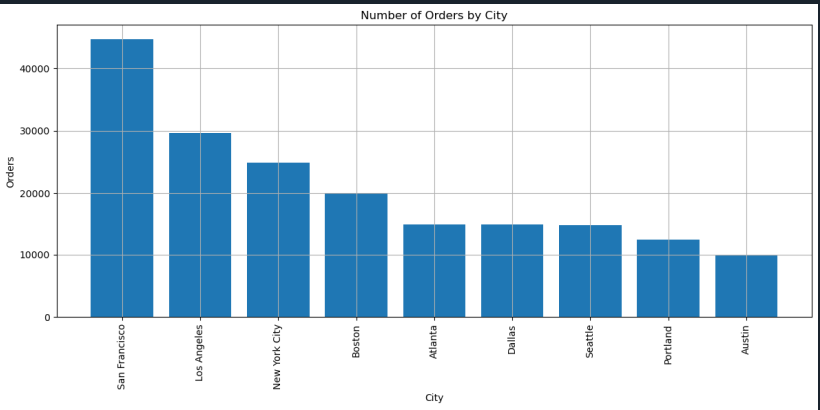
- A significant spike during Q4(October – December) suggests strong holiday-season buying behavior.

- February and September showed relatively low sales, indicating off-peak months.

Conclusion:

There is a strong seasonal component to customer spending. Businesses can utilize this trend for strategic planning like increasing ad budgets, launching promotions, or managing inventory during high-sales periods (e.g., holidays).

2. Number of Orders by City



Analysis: This bar chart highlights the total number of orders placed per city. Each bar represents a city where the business operates.

Observations:

- San Francisco leads with the highest order count, significantly ahead of other cities.

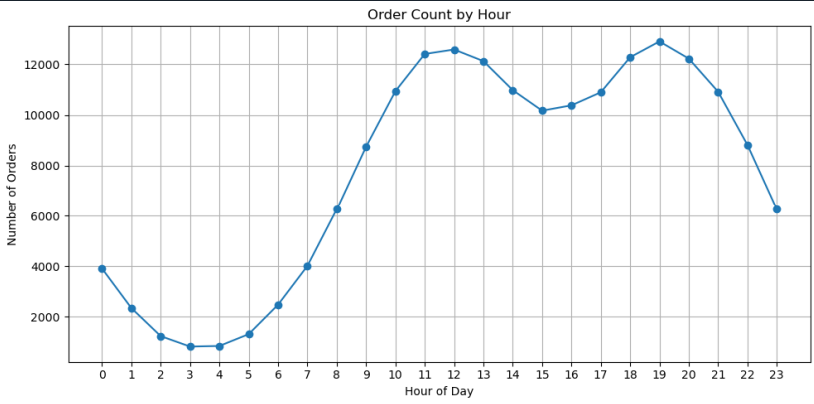
- Los Angeles and New York City follow next.

- Smaller cities like Portland and Austin show the lowest activity.

Conclusion:

Urban hubs like San Francisco, LA, and NYC are key markets. These cities likely contribute the most to revenue and deserve priority in logistics, promotions, and customer service planning.

3. Order Count by Hour



Analysis:

This line graph illustrates the distribution of orders throughout the 24-hour day.

Observations:

- The number of orders sharply increases from 9 AM, peaking at 12 PM.

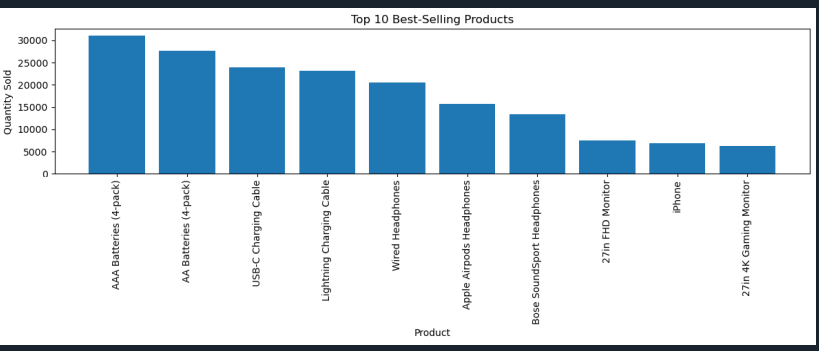
- A second peak occurs between 6 PM and 8 PM.

- Very few orders are placed between midnight and 6 AM.

Conclusion:

Consumer activity peaks around lunch and evening hours, indicating optimal times for sending promotional emails or ad campaigns. This also highlight when customer support and system scalability are most needed.

4. Top 10 Best-Selling Products



Analysis: This bar chart shows the top 10 products sold based on quantity.

Observations:

- AAA and AA batteries (4-packs) dominate the list, followed by charging cables.

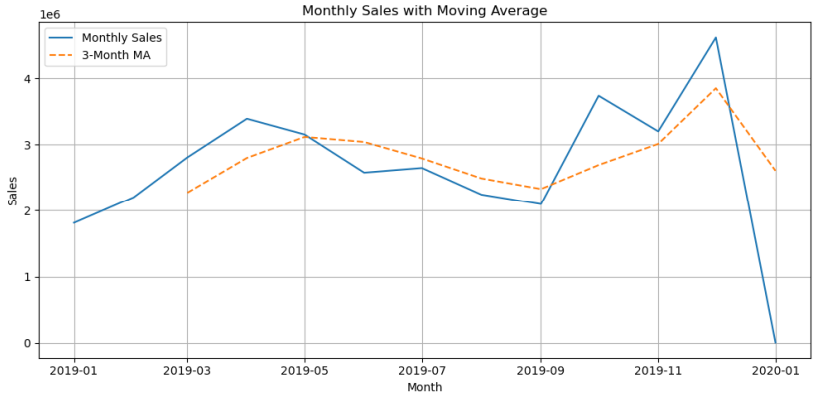
- Several tech accessories like wired headphones, Airpods, and monitors are also top sellers.

- High-priced items like phones and monitors appear less frequently than accessories.

Conclusion:

Smaller, low-cost and high-utility items sell more frequently. These items could be used for cross-selling or bundled deals. Emphasizing accessory products may improve total sales volume even with lower margins.

5. Monthly Sales with Moving Average



Analysis: This time series chart compares monthly sales against a 3-month moving average (MA) to reveal underlying trends.

Observations:

- The moving average smooths out short-term fluctuations.

- It confirms seasonal peaks in April, October, and December.

- The drop in January 2020 is exaggerated due to limited data or transition between years.

Conclusion:

Using a moving average helps highlight long-term trends and seasonality while minimizing short-term noise. This is crucial for forecasting sales, resource allocation, and strategic planning.