Exploring The Sales of Company

Objective

This project focuses on analysing a sample dataset to explore key trends and relationships among variables such as city, time, and performance indicators. The goal is to practice essential data analysis techniques including data cleaning, visualization, and statistical evaluation. The objective of this analysis is to identify patterns in sales performance in USA over the given time period, which later help us to find the factors that influent the sales.

Dataset Description

As I mentioned above, this is the sample dataset in Kaggle (https://www.kaggle.com/datasets/shantanugarg274/sales-dataset). This dataset, contains 12 columns and around 1200 rows, displays the sales from 2020 to 2025 years, in the different states of the USA.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Order ID	Amount	Profit	Quantity	Category	Sub-Category	PaymentMo	Order Date	CustomerName	State	City	Year-Month	
2	B-26776	9726	1275	5	Electronics	Electronic Games	UPI	27/06/2023	David Padilla	Florida	Miami	2023-06	
3	B-26776	9726	1275	5	Electronics	Electronic Games	UPI	27/12/2024	Connor Morgan	Illinois	Chicago	2024-12	
4	B-26776	9726	1275	5	Electronics	Electronic Games	UPI	25/07/2021	Robert Stone	New York	Buffalo	2021-07	
5	B-26776	4975	1330	14	Electronics	Printers	UPI	27/06/2023	David Padilla	Florida	Miami	2023-06	
6	B-26776	4975	1330	14	Electronics	Printers	UPI	27/12/2024	Connor Morgan	Illinois	Chicago	2024-12	
7	B-26776	4975	1330	14	Electronics	Printers	UPI	25/07/2021	Robert Stone	New York	Buffalo	2021-07	
8	B-26942	1525	185	12	Office Supplies	Pens	Debit Card	11/05/2024	John Fields	Florida	Orlando	2024-05	
9	B-26942	1525	185	12	Office Supplies	Pens	Debit Card	09/10/2021	Clayton Smith	Florida	Miami	2021-10	
10	B-26640	883	117	10	Electronics	Laptops	EMI	18/11/2022	Richard Kelley	California	Los Angele	2022-11	
11	B-26640	8127	3551	16	Furniture	Tables	Credit Card	18/11/2022	Richard Kelley	California	Los Angele	2022-11	
12	B-25890	2516	734	19	Furniture	Chairs	UPI	12/11/2023	Jacqueline Hubbard	New York	New York (	2023-11	
13	B-25102	2975	462	14	Office Supplies	Pens	Credit Card	23/03/2020	Jessica Anderson	New York	New York (	2020-03	
14	B-25426	6851	2812	1	Office Supplies	Markers	Debit Card	04/05/2020	Mary Taylor	Florida	Orlando	2020-05	
15	B-25426	7626	1046	15	Furniture	Sofas	Credit Card	04/05/2020	Mary Taylor	Florida	Orlando	2020-05	

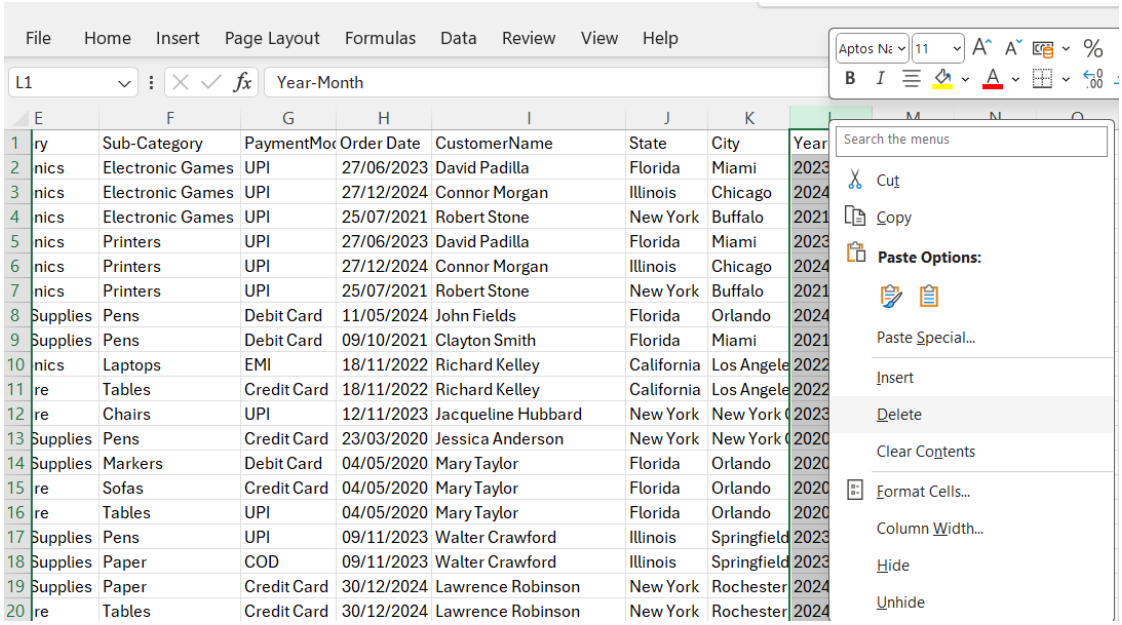
Data cleaning

Next step of analysis is data cleaning and preparation which contains removing duplicates of null values, converting data types or renaming columns for clarity. After carefully analysing the dataset, I found that the columns Category and Sub-Category, as well as City and State, convey similar information in a hierarchical manner. However, for the purpose of detailed analysis and comparison at different levels, I decided to keep both columns for now:

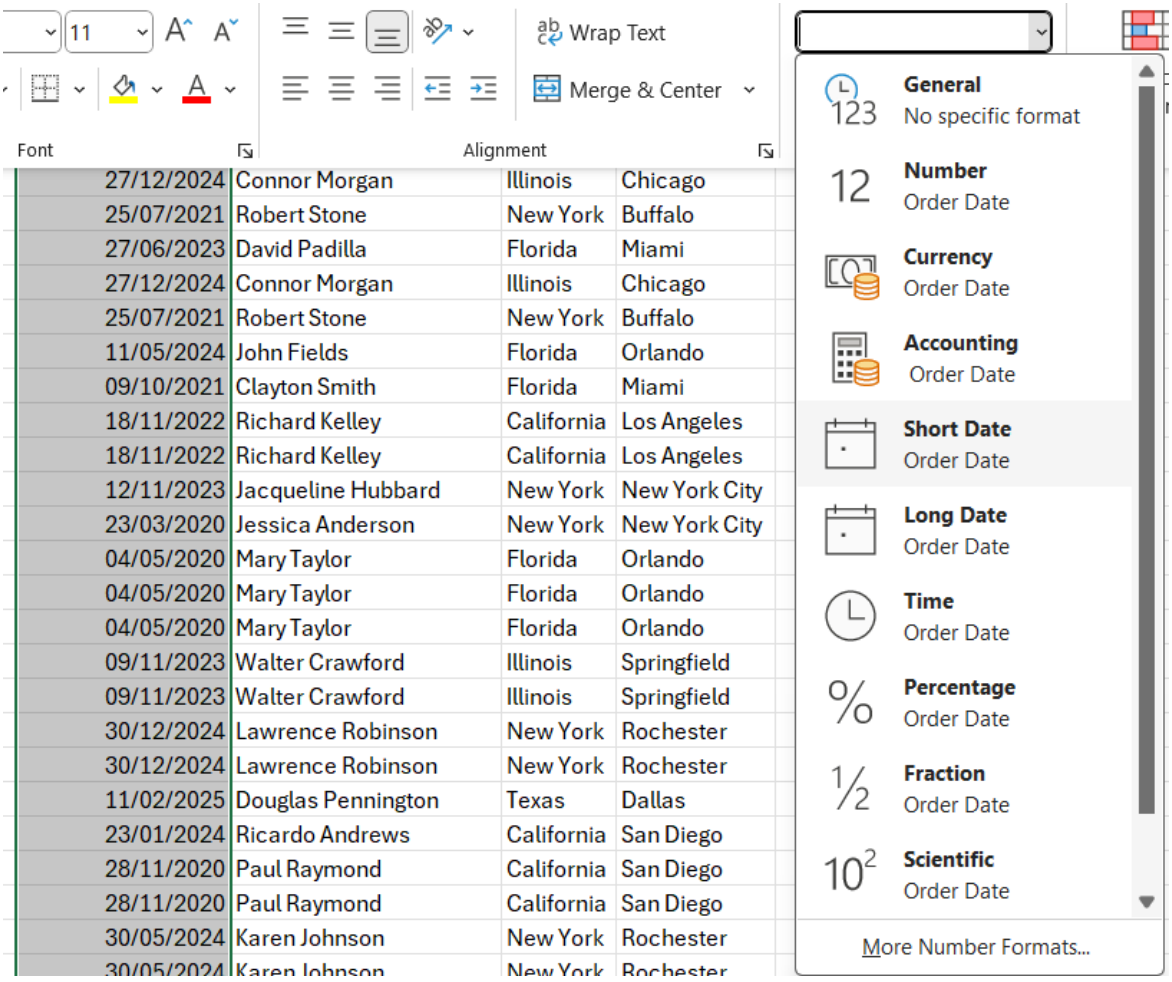
E	F
Category	Sub-Category
Electronics	Electronic Games
Electronics	Electronic Games
Electronics	Electronic Games

J	K
State	City
Florida	Miami
Illinois	Chicago
New York	Buffalo

Also, since the Year-Month column can be derived from Order Date, keeping only the Order Date column is sufficient for most analyses. In Excel, date-based grouping (by day, month, or year) can be performed directly through PivotTable features. Therefore, the Order Date column was retained as the primary date field for analysis:



Another thing data analysts should be careful is the data types. In my dataset, columns Amount, Profit, Quantity have to be numeric while Order date column should be date format. But by default, Excel see all columns as General. So, to get precise results I converted the datatypes as I said:



Exploratory Data Analysis (EDA)

The purpose of this part is to explore and analyse the dataset to uncover patterns, trends and relationships between key variables. By conducting EDA, I aim to gain a deeper understanding of the dataset and help company to detect the weak point of company.

Title: Sales Analysis

Purpose:

To detect which product categories, contribute the most to total sales.

Method:

By using Pivot Table and summed amount by each category inside Column Chart

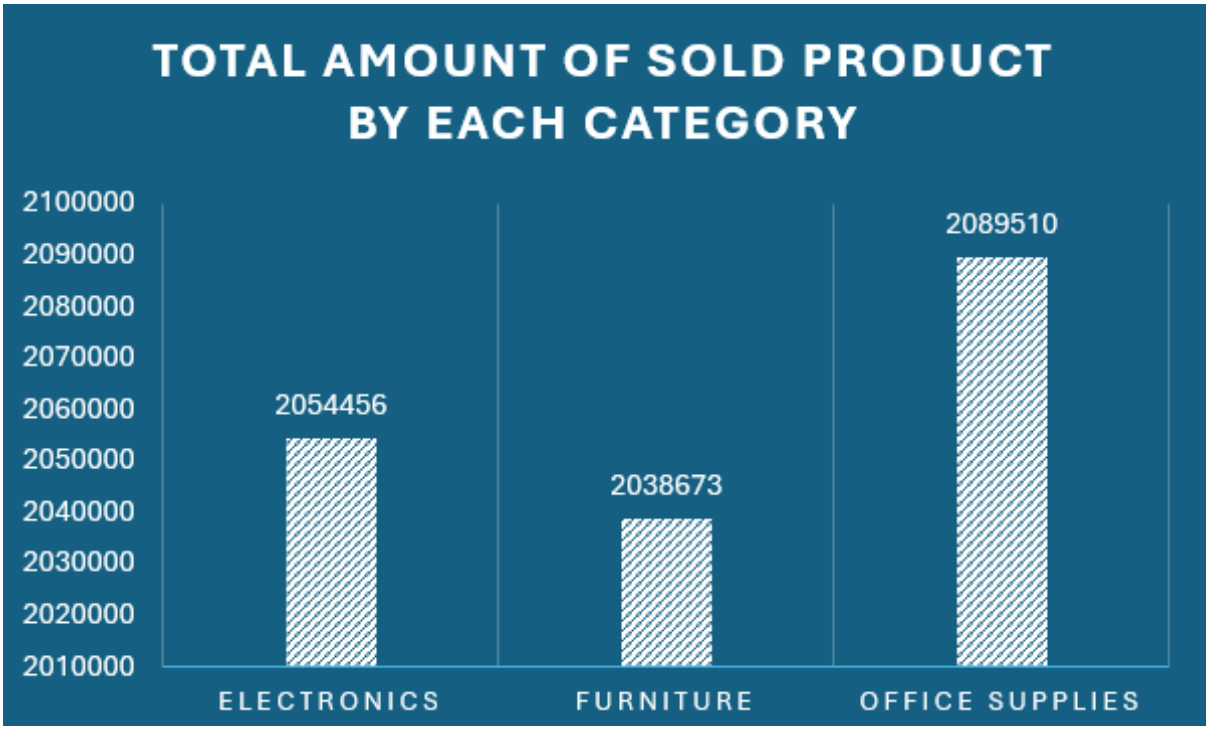
Observation:

Office Supplies Category account for the largest portion of number of total products sold (around 209000 number)

The product of *Electronics* and *Furniture* Categories were sold 2054456 and 2038673 respectively.

Recommendation:

This shows that customers tend to buy *Office Supplies* more frequently, likely because these items are everyday essentials and lower in price. On the other hand, *Electronics* and *Furniture* are sold in smaller quantities but probably generate more profit per sale due to their higher value. From a business point of view, this means it's important to maintain steady sales of office supplies while also focusing on promoting higher-value products to maximize overall profit.



Title: Profit by Payment Method

Purpose:

To evaluate which payment methods, generate the most profit, informing potential strategic promotions.

Method:

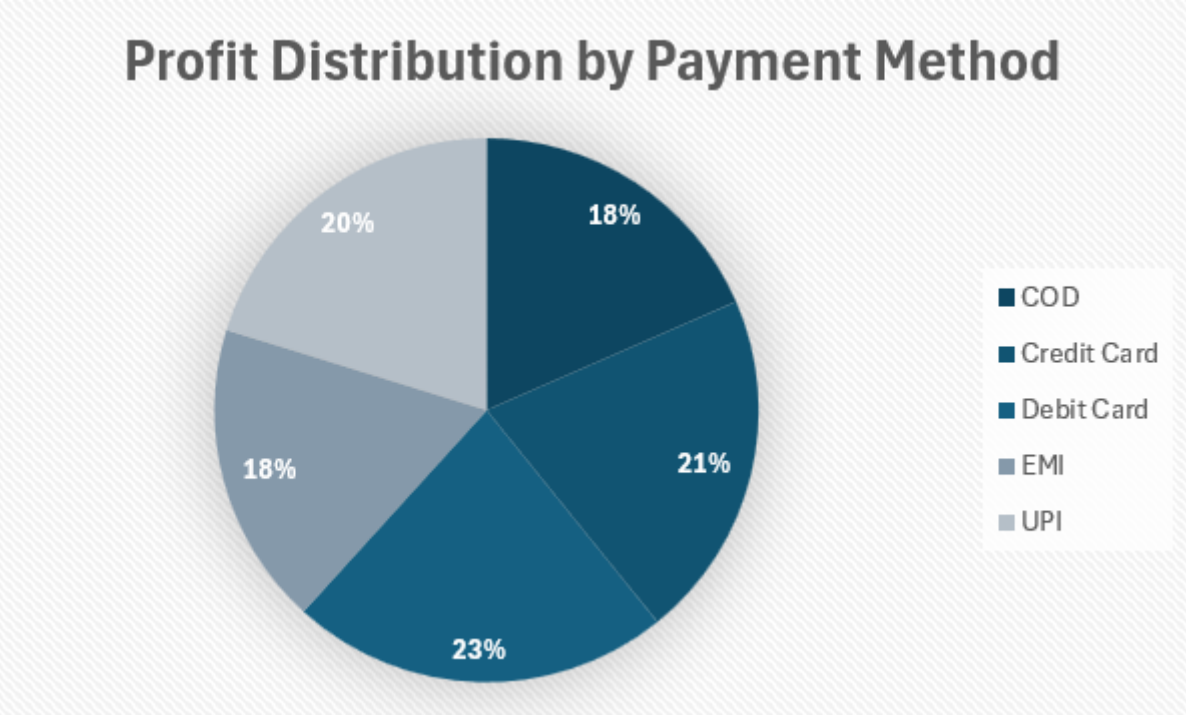
After creating pivot table, displaying the payment methods percentage in a pie chart

Observation:

Nearly a quarter of customers use Debit Card, while 18 % of people made their shopping through COD (Cash on delivery) and EMI (Equated Monthly Instalment). Meanwhile, roughly a fifth of customers preferred Credit Card and UPI (Unified Payments Interface).

Recommendations:

Since Debit Cards are the most commonly used payment method, maintaining smooth and secure debit transactions should remain a priority. However, the relatively lower usage of Credit Card and UPI payments suggests an opportunity to promote these methods through discounts, cashback offers, or loyalty points. Encouraging digital payments can not only improve customer convenience but also reduce cash-handling costs associated with COD orders.



Title: Sales Amount by City

Purpose:

To determine the purchase distribution among the given cities

Method:

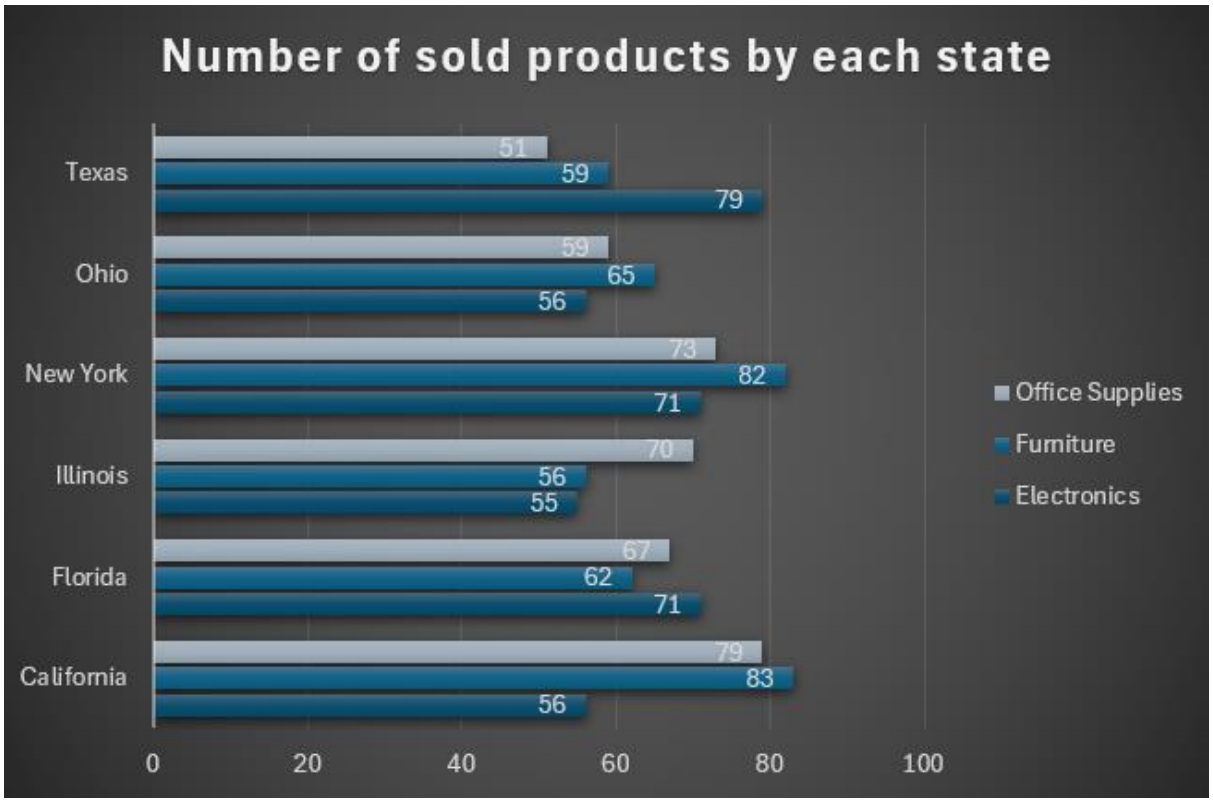
Bar chart based on City and Amount column

Observation:

From the chart, it is evident that *California* and *New York* record the highest number of products sold across all three categories. In particular, *Electronics* perform exceptionally well in these two states, reaching around 83 and 82 units respectively. *Texas* also shows strong demand for Electronics, while *Illinois* demonstrates relatively lower sales across all categories. Overall, *Electronics* consistently outperform *Furniture* and *Office Supplies* in most states, suggesting a strong customer preference toward this category.

Recommendations:

The strongest markets are California and New York, so keeping enough inventory and focusing marketing efforts there should continue to be top priorities. Given the strong demand for electronics, the business should think about implementing loyalty programs or package deals to increase sales even more. Conducting regional promotions or examining consumer preferences could assist discover problems restricting sales success and increase total market reach in states with smaller numbers, such as Illinois.



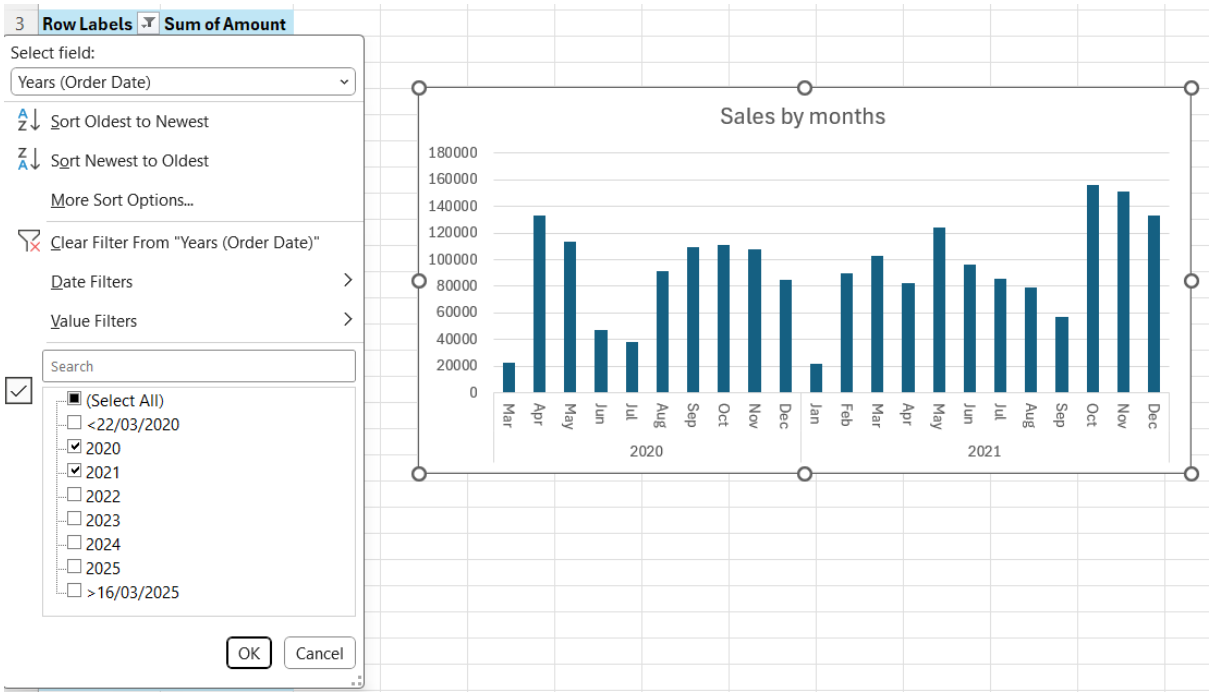
Title: Sales by Months

Purpose:

To analyse the monthly and seasonal sales performance in order to identify periods of high and low demand and gain a better understanding of customer purchasing behaviour.

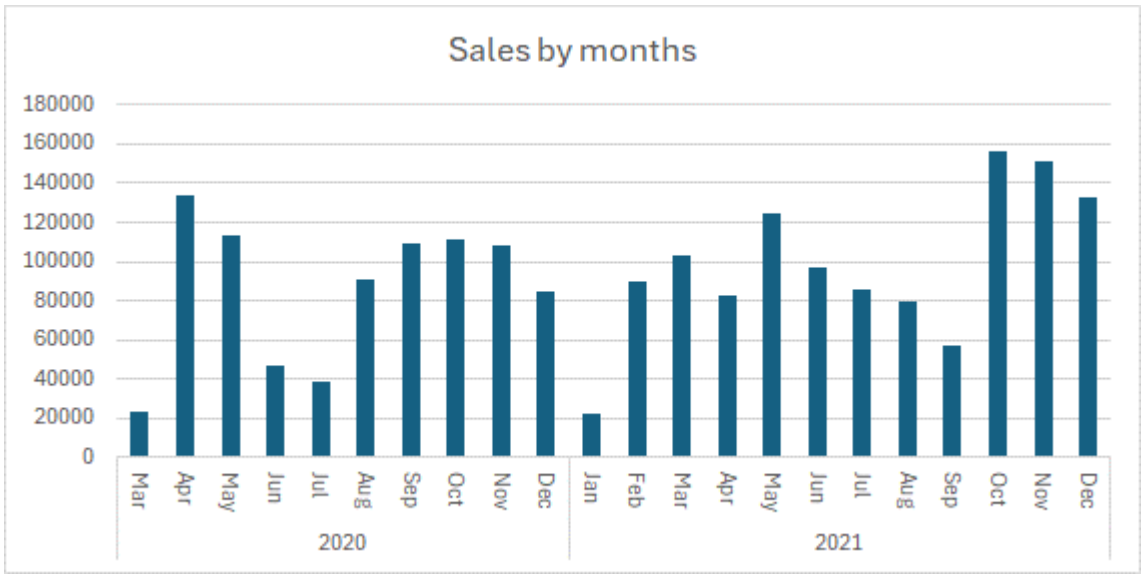
Method:

To ensure clearer visualization, I analysed each two years' sales data separately using a Pivot Table. This allowed all twelve months of each two years to be displayed accurately in a bar chart, providing a complete view of the sales trends within individual years:

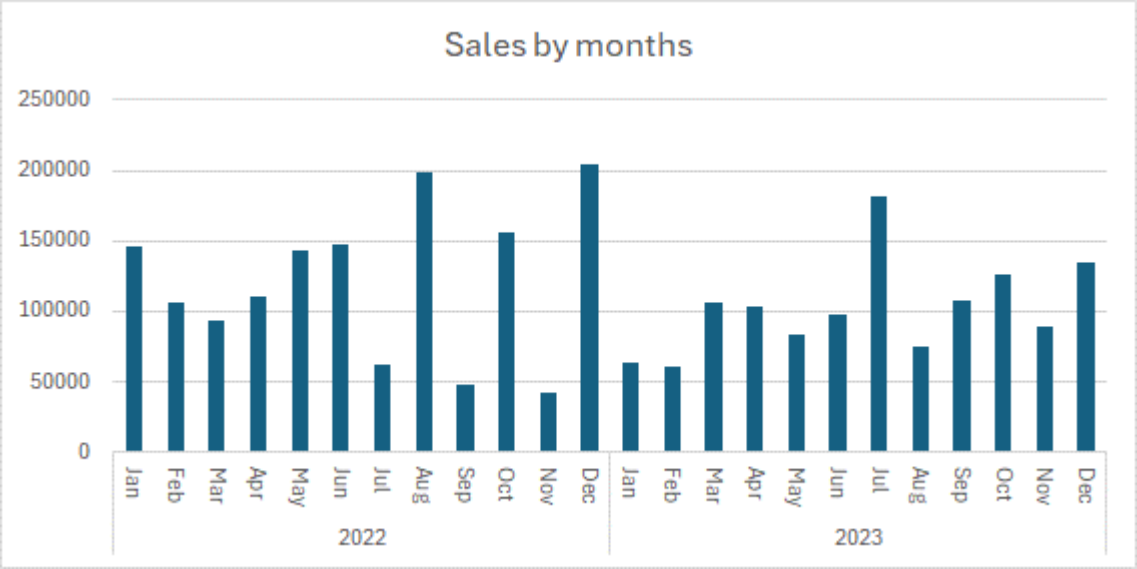


Observation:

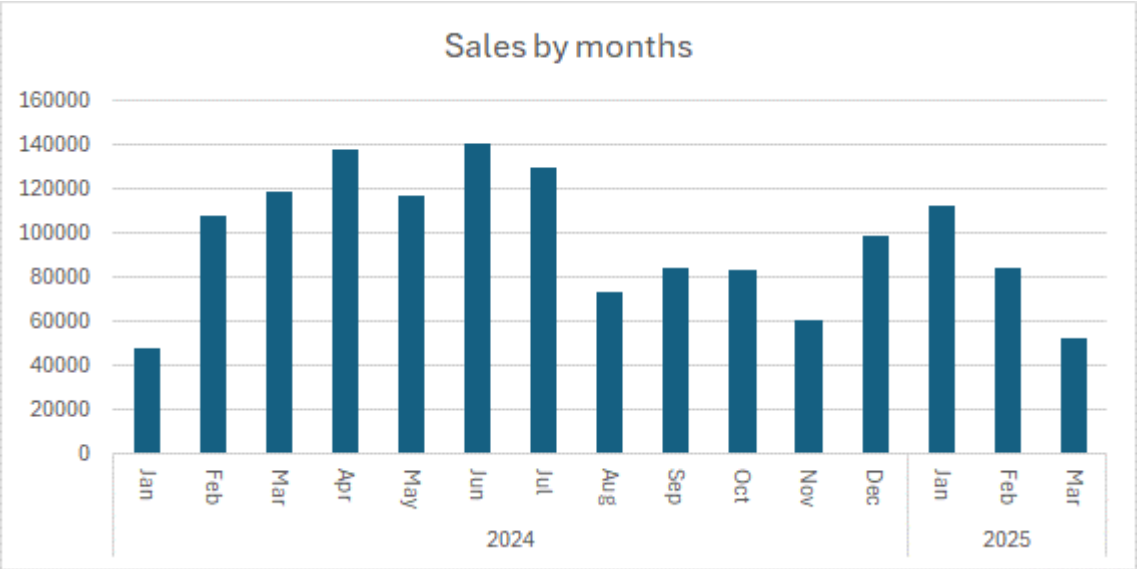
The sales trends for 2020 and 2021 indicate an early period of inconsistency followed by gradual stabilization. In 2020, sales increased sharply from March to May, showing signs of recovery after a weak start, while the final quarter maintained moderate stability. In 2021, the monthly pattern became more balanced, with notable peaks in September and October, suggesting improved planning and stronger market response compared to the previous year:



During 2022 and 2023, overall sales performance strengthened. In 2022, October and December recorded the highest sales, reflecting seasonal demand. The trend continued in 2023, where consistent mid-year results and high sales in July and November demonstrated improved control over sales cycles. These two years collectively highlight a more predictable and efficient sales structure:



In the final period, sales remained steady with some month-to-month variations. The year 2024 showed consistent performance, reaching a peak in June, while 2025 started with moderate figures in January and February but declined slightly in March. Overall, this period reflects continued business stability with minor seasonal fluctuations:



Recommendations:

Based on the observed sales trends from 2020 to 2025, it is recommended that the company focus on strengthening its sales forecasting and seasonal planning strategies. Since the data reveals consistent peaks during mid-year and year-end periods, the marketing and inventory teams should prioritize these months for targeted campaigns and sufficient stock allocation.

Additionally, months showing moderate or declining sales—particularly in early quarters—should be used to introduce promotional offers or loyalty programs to maintain customer engagement throughout the year. The relatively stable performance across later years also suggests that continuing to optimize data-driven decision-making can help sustain long-term growth and operational efficiency.

Finally, developing a predictive model or dashboard to visualize monthly sales patterns could help the organization respond proactively to market changes and plan more effectively for upcoming periods.

Findings

After completing the exploratory data analysis, I identified several key insights that reveal both performance strengths and potential areas for improvement within the company's sales operations:

1. Sales Performance by Product Category.

   The analysis shows that *Office Supplies* represent the largest share of total products sold, indicating strong and frequent customer demand for low-cost, everyday items. However, *Electronics* and *Furniture*—though sold in smaller quantities—generate considerably higher sales value per unit, suggesting that these categories are the primary drivers of total revenue. Balancing promotional efforts between high-volume (Office Supplies) and high-value (Electronics, Furniture) products could optimize overall profitability.

2. Payment Method Profitability

   Debit Card transactions accounted for the highest proportion of profit, followed by *COD* and *EMI*. Meanwhile, *Credit Card* and *UPI* payments are less frequently used, signalling an opportunity to encourage digital transactions. Implementing targeted incentives such as cashback or reward programs could shift more customers toward digital payment options, enhancing transaction efficiency and reducing operational costs linked to cash handling.

3. Regional and Category-Based Insights

   Geographical analysis highlighted *California* and *New York* as the company's strongest markets, both demonstrating consistently high sales across all categories—especially Electronics. This regional concentration suggests that marketing efforts and inventory management should remain focused on these states. Conversely, states such as *Illinois* and *Texas* showed lower sales volumes, revealing potential areas for market development through localized advertising or pricing strategies.

4. Monthly and Seasonal Sales Trends

   The trend analysis across 2020–2025 revealed that sales tend to peak during *mid-year (June–July)* and *end-of-year (October–December)* periods, reflecting clear seasonal patterns. This cyclical behaviour suggests the company has developed more stable and predictable sales cycles in recent years. Aligning marketing campaigns and inventory levels with these high-demand months will further strengthen overall sales performance.

Meanwhile, slower months—typically early in the year—can be targeted with discount campaigns or promotional bundles to maintain steady engagement.

5. Operational Stability and Growth Opportunity

   Across the five-year period, the data indicates a steady improvement in both sales' consistency and profit generation. The company appears to have moved from volatile performance in early years to a more stable and predictable pattern, showing effective adaptation to market trends. Going forward, deeper integration of data analytics for demand forecasting, sales optimization, and regional performance tracking could enhance decision-making and drive sustainable growth.

Overall, the EDA reveals that while the company performs strongly in high-demand regions and product lines, there is still room to expand market reach and improve payment diversity. By leveraging these insights—focusing on category balance, regional targeting, and seasonal sales planning—the company can enhance profitability, improve operational efficiency, and strengthen its competitive position in the market.