

DATA ANALYSIS REPORT

SUPER STORE



Data Exploration

Business Problem

Research Questions

Analysis & Findings

Conclusion

DATA EXPLORATION

To conduct a comprehensive data exploration for the superstore analysis, the dataset is collected from Kaggle. Kaggle often has datasets related to retail and e-commerce businesses, such as the "Sample - Superstore" dataset. This dataset typically includes information on orders, sales, profits, customer segments, shipping modes, and regions.

Link of the Data Set used in this analysis -

<https://www.kaggle.com/datasets/amrboghdady74/superstoreus/data>

BUSINESS PROBLEM

The super store has experienced declining sales in certain states. To address this issue, the objective is to analyse factors contributing to the decline comprehensively. This includes evaluating shipping modes to enhance efficiency and customer satisfaction, identifying high-value customer segments based on demographics and purchasing behaviour, and pinpointing product categories with low sales for strategic adjustments. Additionally, conducting a region-wise profitability analysis will help understand cost-effectiveness and revenue generation in each state. By focusing on these areas, the goal is to reverse the sales decline, improve operational efficiency, and develop tailored strategies to maximize profitability across regions, ensuring sustained growth and market competitiveness.

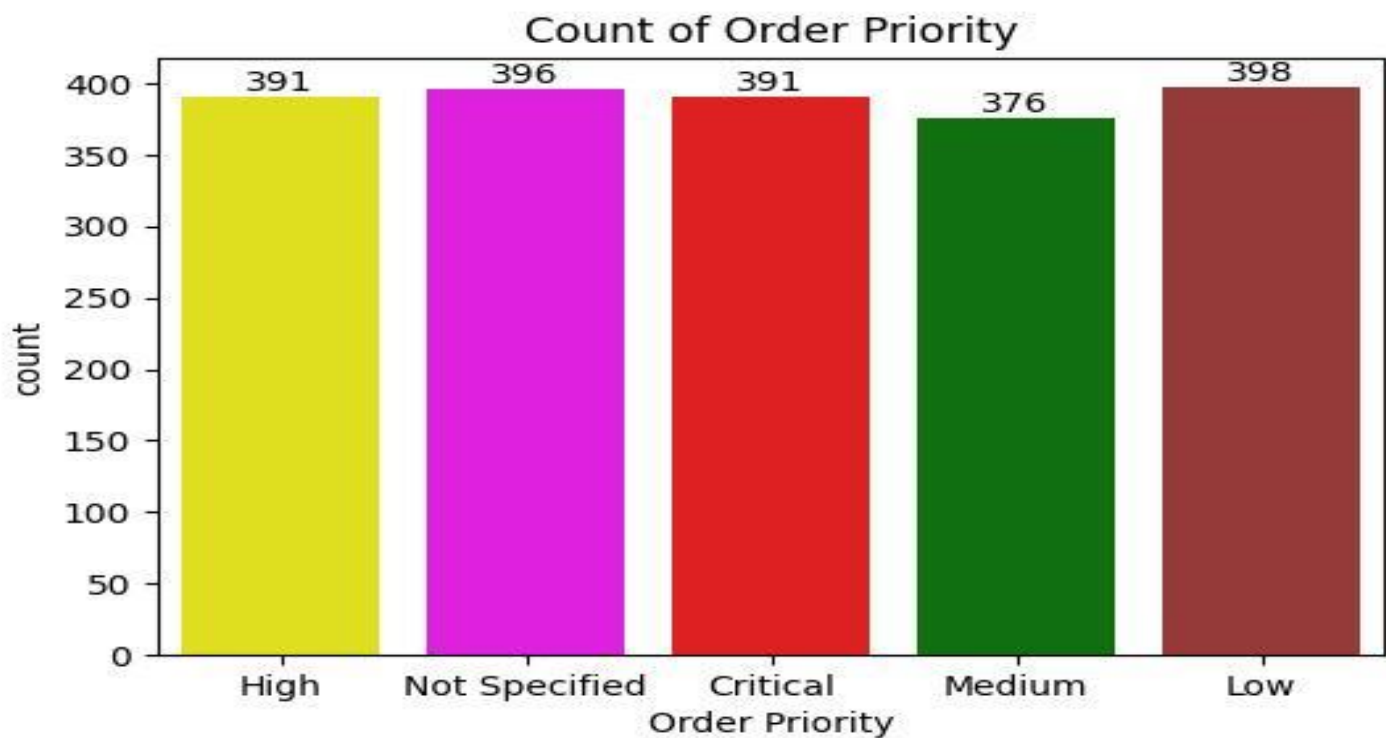
Research Questions

1. How many orders are there for each order priority?
2. Which delivery method is the most effective?
3. How many shipping modes are used for each product category?
4. Count of customer segments ?
5. How many product categories are there?
 - Product sub-category (Office Supplies)
 - Product sub-category (Technology)
 - Product sub-category (Furniture)
6. What is the profit for each product category?
7. Which state has the highest number of sales?
8. Which state has the lowest number of sales?
9. Which region has the highest number of sales?
10. What is the profit for each region?

Analysis & Findings

Analysing Superstore data involves leveraging powerful Python libraries such as NumPy, Pandas, Matplotlib, and Seaborn to derive meaningful insights. NumPy provides essential functions for numerical computations, allowing for efficient manipulation of data arrays and matrices. Pandas, on the other hand, excels in handling structured data, enabling tasks like data cleaning, merging, and aggregation. Matplotlib is used for creating static, animated, and interactive visualizations, while Seaborn enhances these plots with attractive statistical graphics. Together, these tools enable comprehensive analysis of Superstore data, from understanding sales trends and customer behaviour to optimizing inventory management and identifying market opportunities.

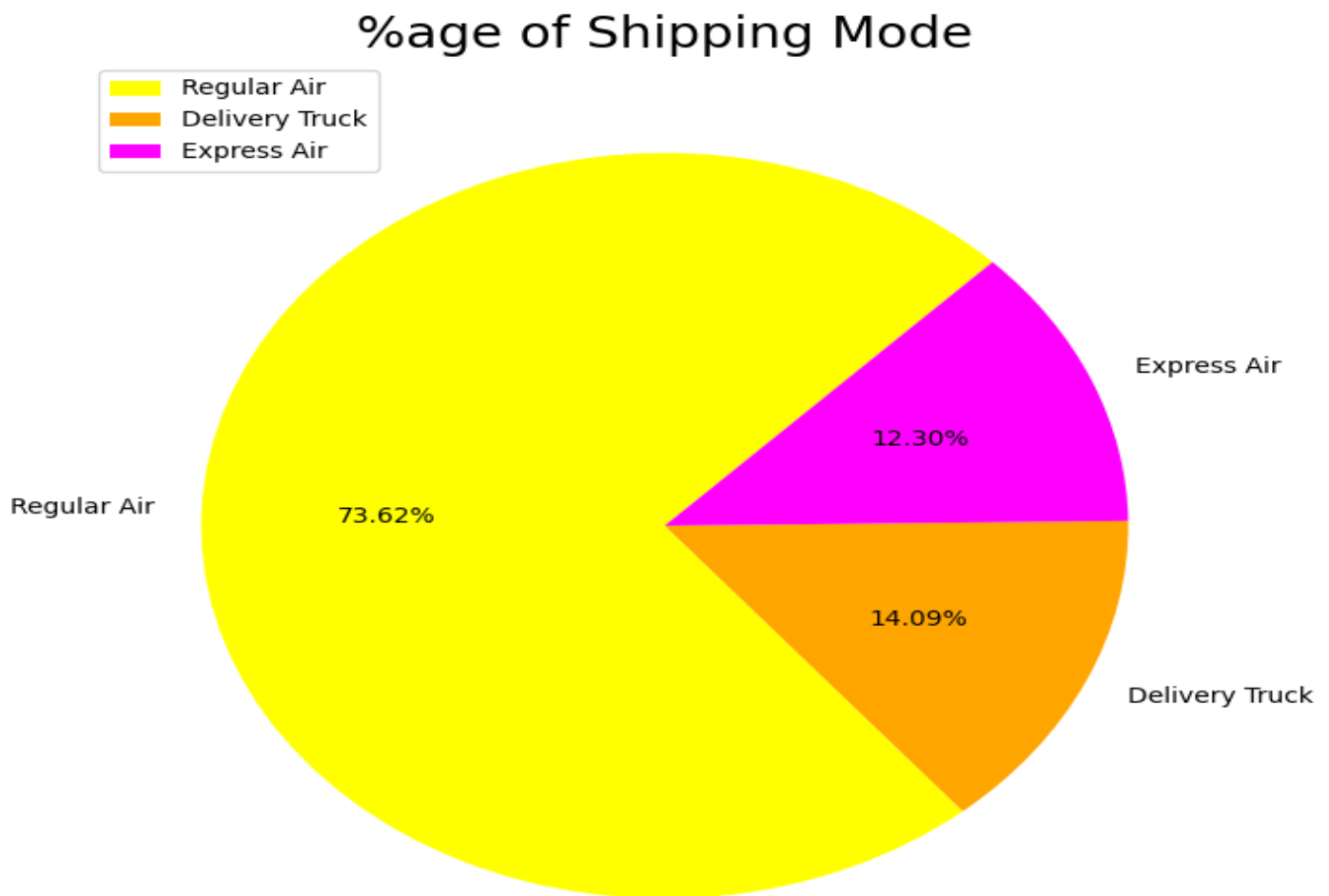
1. How many orders are there for each order priority?



The bar chart provides a clear depiction of the count of orders for each order priority category. Upon examining the data, it is evident that the total number of orders across all priority categories sums up to 1952. The distribution of these orders is relatively balanced, with the "Low" priority category having the highest count of 398 orders, and the "Medium" category having the lowest count of 376 orders. Interestingly, both the "High" and "Critical" categories have an identical count of 391 orders, while the "Not Specified" category is slightly higher at 396 orders.

This even distribution suggests that there is no overwhelming preference for any particular priority category, and thus, resources and attention should be evenly distributed across all categories to ensure efficient processing and handling. Additionally, the substantial number of orders in the "Not Specified" category indicates a potential area for improvement in data entry or categorization processes.

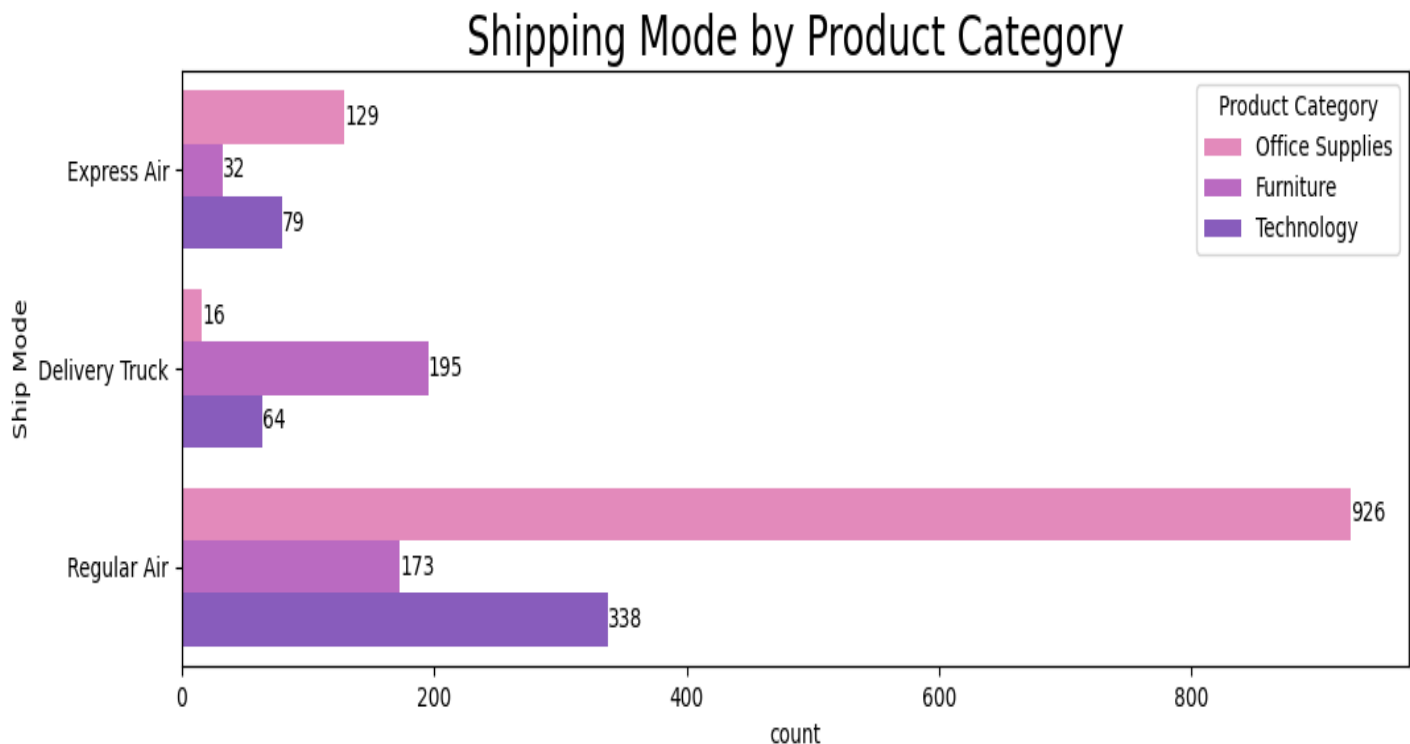
2. Which delivery method is the most effective?



- **Regular Air** is the most used shipping method, accounting for **73.62%** of the total shipping.
- **Delivery Truck** is the second most used method with **14.09%**.
- **Express Air** is the least used method with **12.30%**.

Although Regular Air is the most utilized, determining the most effective delivery method requires further context such as cost, speed, reliability, and customer satisfaction. Regular Air's high usage suggests it may offer a favourable balance between cost and delivery speed, while Delivery Truck and Express Air are likely chosen for specific needs, such as ground transportation or expedited delivery.

3. How many shipping modes are used for each product category?



1. Regular Air:

- **Office Supplies:** This category has the highest count with 926 shipments.
- **Technology:** This category has the second-highest count with 338 shipments.
- **Furniture:** This category has a moderate count with 173 shipments.

2. Express Air:

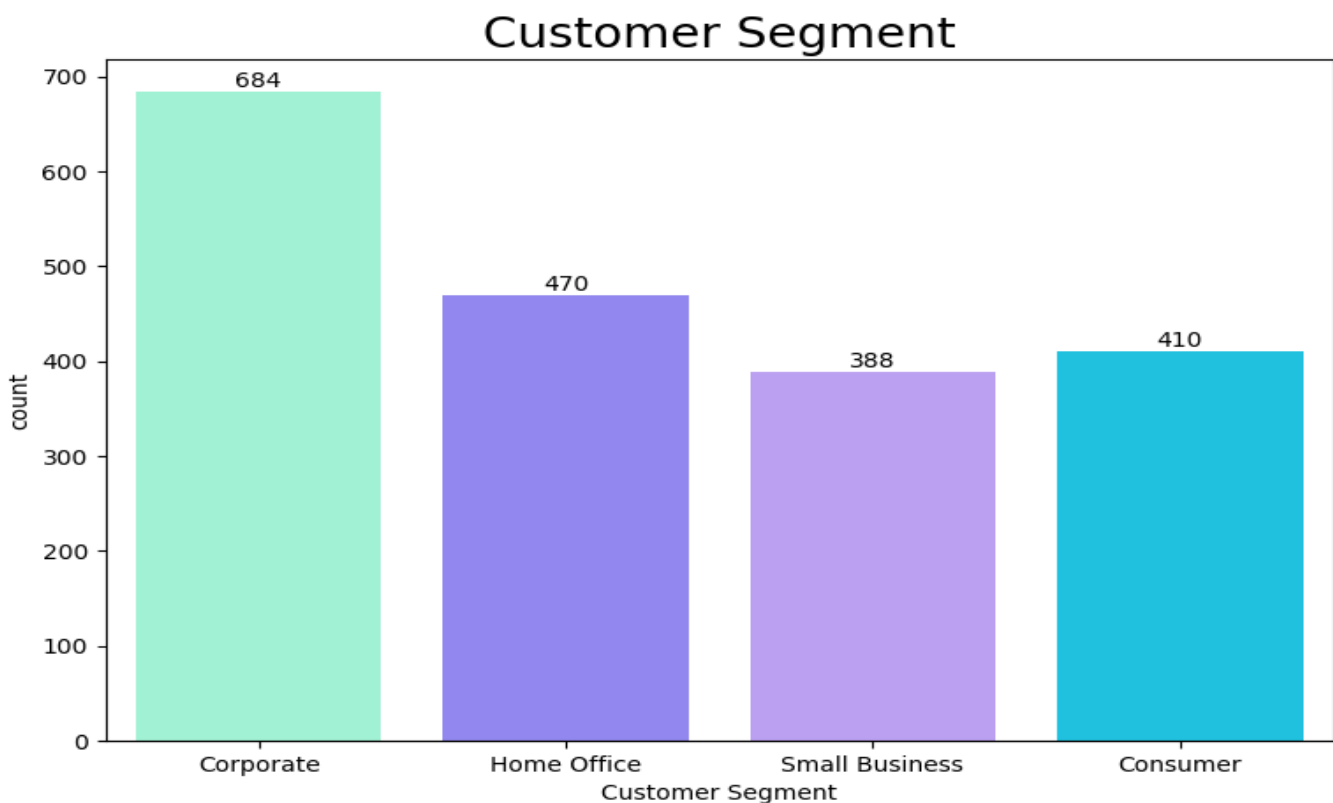
- **Office Supplies:** This category has 129 shipments.
- **Technology:** This category has 79 shipments.
- **Furniture:** This category has the lowest count with 32 shipments.

3. Delivery Truck:

- **Office Supplies:** This category has 195 shipments.
- **Technology:** This category has the highest count for Delivery Truck with 64 shipments.
- **Furniture:** This category has the lowest count with 16 shipments.

These findings highlight that Regular Air is the most common shipping method for all product categories, particularly for Office Supplies. The insights can help optimize logistics and supply chain strategies by aligning shipping methods with product category preferences.

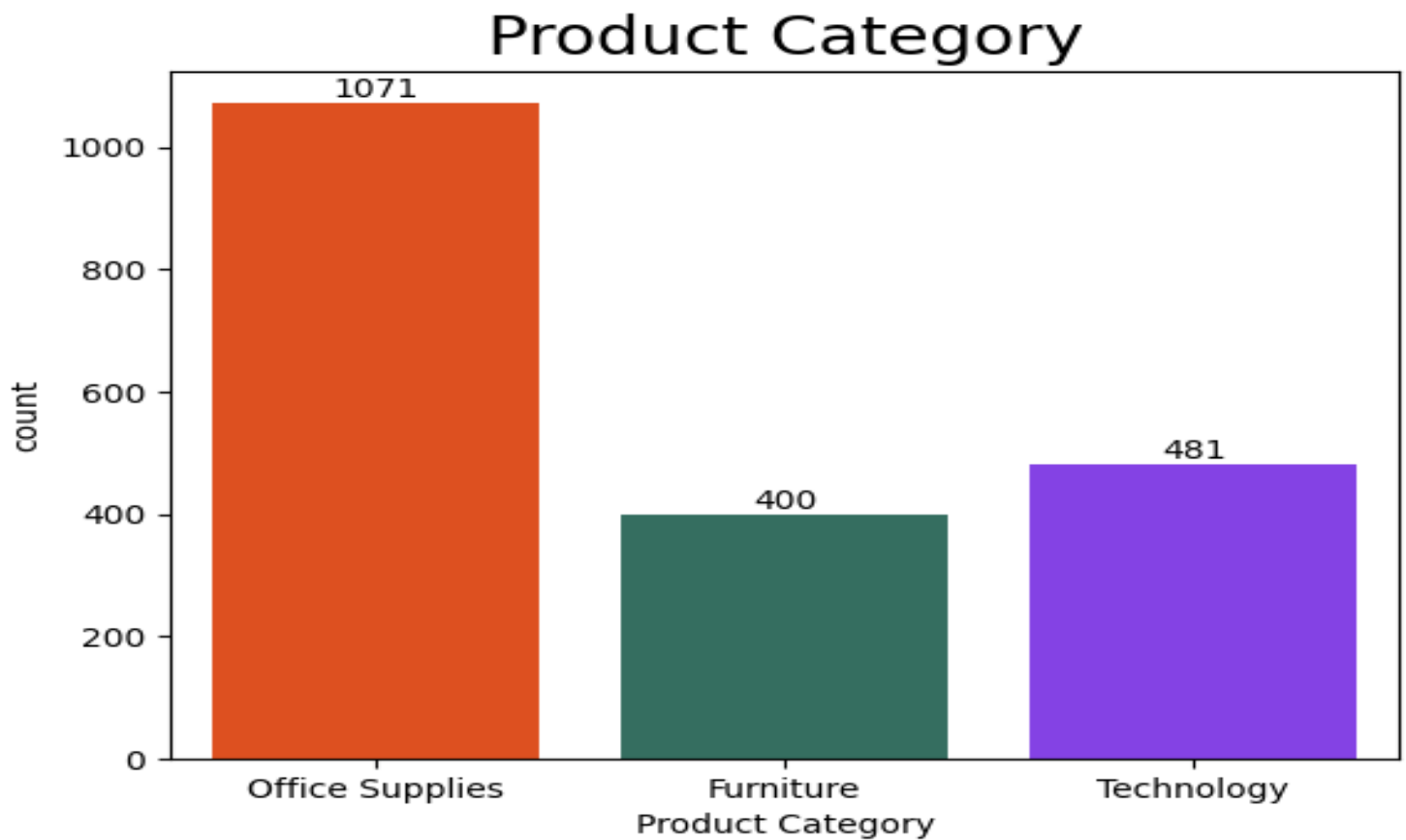
4. Count of customer segments?



- The Corporate segment has the highest count of customers, with 684.
- The Home Office segment has the second highest count, with 470 customers.
- The Consumer segment has 410 customers, making it the third highest.
- The Small Business segment has the lowest count among the four segments, with 388 customers.

These insights indicate that the corporate segment is the most significant, potentially warranting more focused business efforts. Meanwhile, the Small Business segment, despite having the lowest count, still represents a substantial customer base, suggesting opportunities for targeted growth strategies. Understanding these distributions can aid in resource allocation, marketing, and identifying areas for potential improvement.

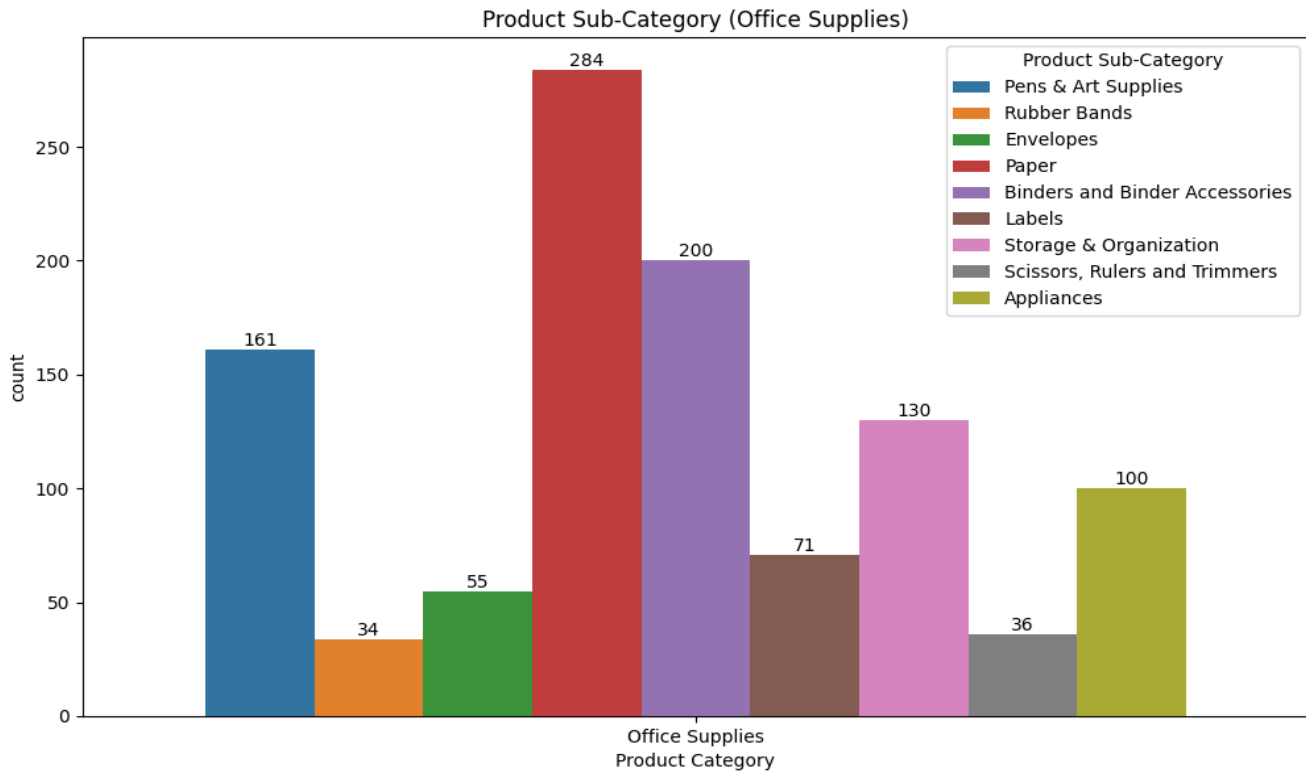
5. How many product categories are there?



- Office Supplies is the largest product category, with a total count of 1,071 items.
- Furniture is the smallest category, with 400 items.
- Technology falls in between, with a count of 481 items.

These insights suggest that Office Supplies is a primary focus for the business, likely due to its substantial inventory count. Despite having the fewest items, the Furniture category still represents a significant portion of the overall inventory, indicating room for strategic growth or optimization. The Technology category, while not as large as Office Supplies, also holds a considerable number of items, highlighting its importance and potential for increased emphasis. These distributions can inform decisions regarding inventory management, marketing strategies, and resource allocation to enhance business operations and growth.

▪ Product sub-category (Office Supplies)



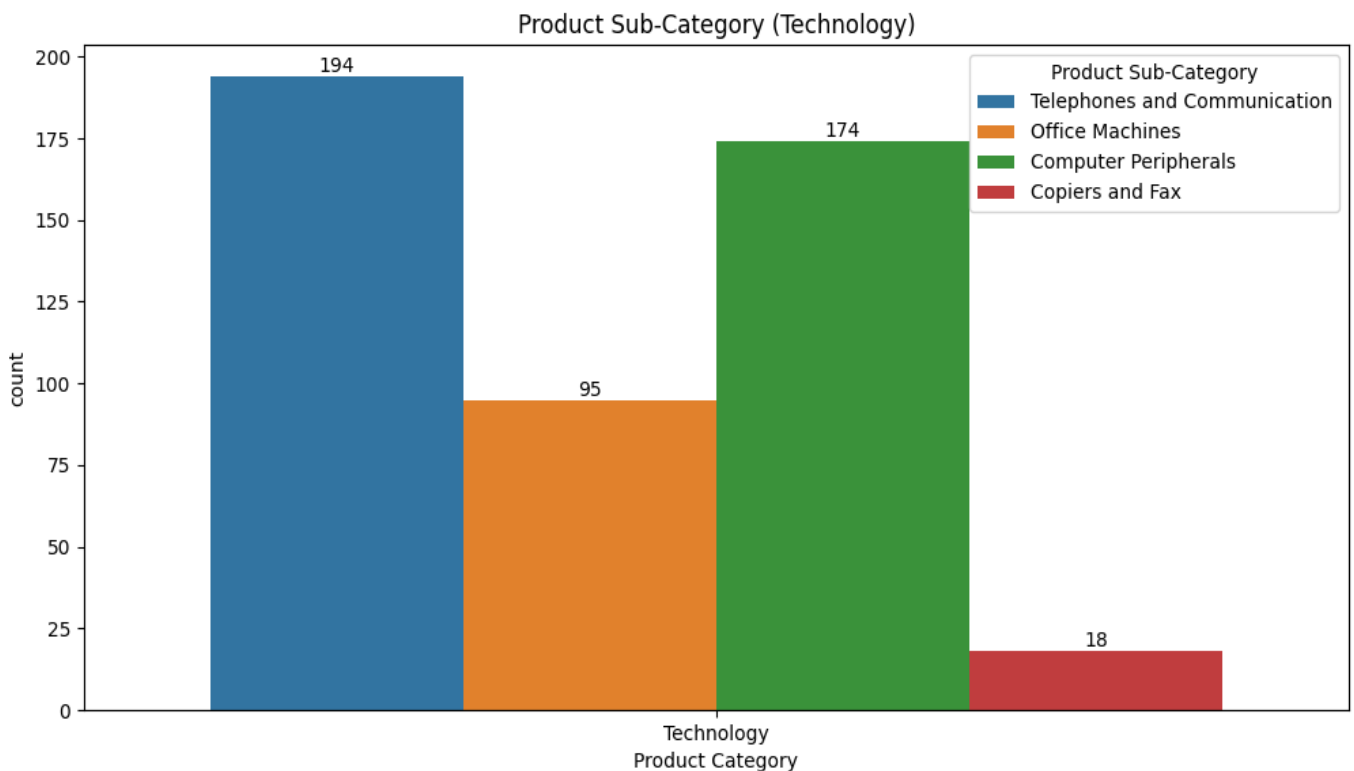
The bar chart presents the count of items within various sub-categories of the Office Supplies category.

- Paper is the largest sub-category, with 284 items.
- The Binders and Binder Accessories sub-category has 200 items, making it the second largest.
- Pens & Art Supplies has a count of 161 items, placing it third in size.
- This sub-category has 130 items, indicating a significant portion of the Office Supplies inventory.
- Appliances account for 100 items.
- This sub-category has 71 items.
- Envelopes have a count of 55 items.
- Labels have 36 items.
- Rubber Bands is the smallest sub-category, with 34 items.

Overall Insights:

- Paper is the most significant sub-category within Office Supplies, suggesting a high demand or usage rate for these products.
- Binders and Binder Accessories also represent a large portion, highlighting their importance.
- The smaller sub-categories like Rubber Bands, Labels, and Envelopes, while not as substantial in count, still play a role in the overall inventory.
- Understanding the distribution of these sub-categories can help in inventory management, ensuring that popular items are well-stocked and identifying opportunities to optimize or promote lesser-stocked items.

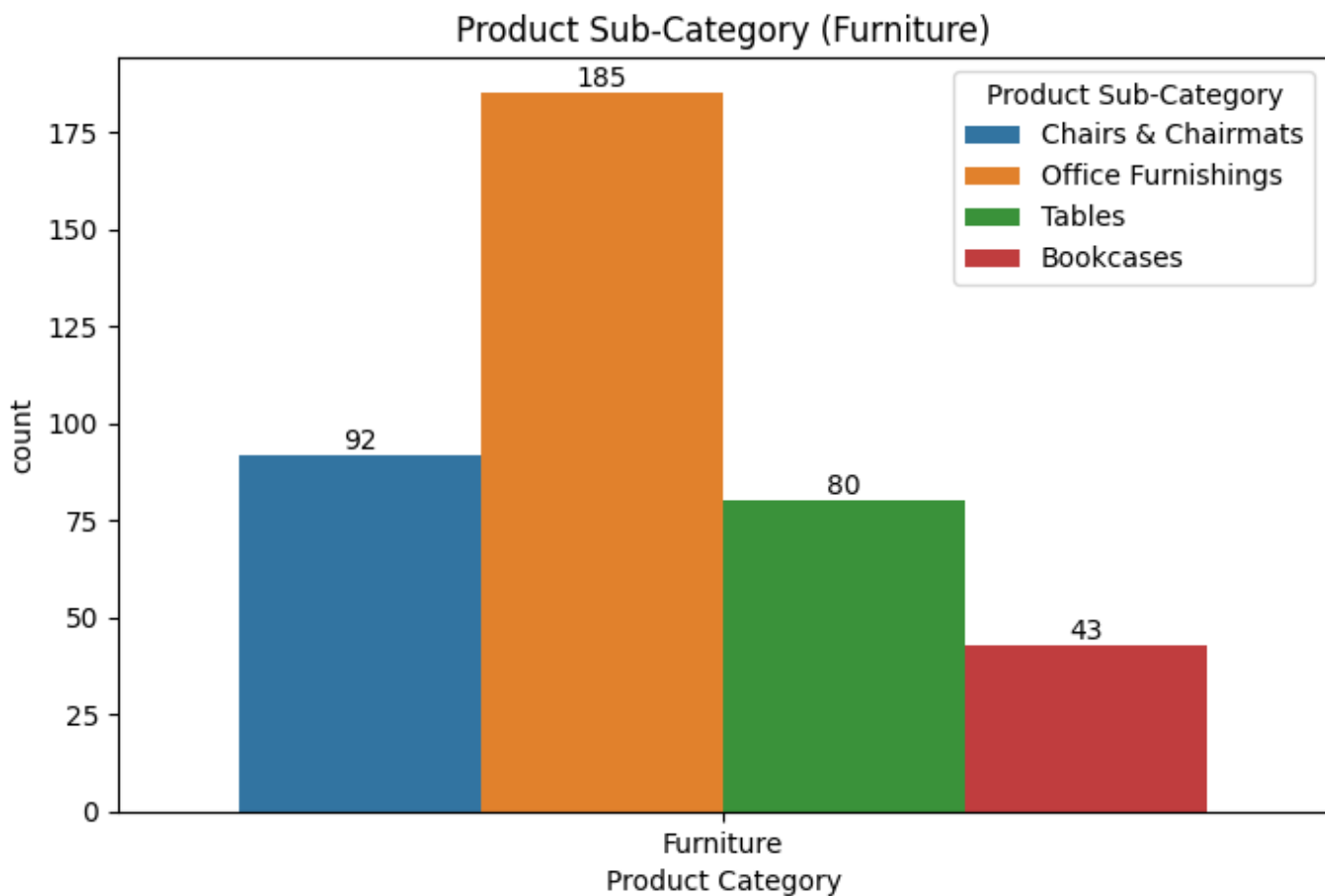
▪ **Product sub-category (Technology)**



- The **Telephones and Communication** sub-category has the highest count, with **194 units**. It indicates that products in this sub-category are the most prevalent or in highest demand within the technology category.
- The second highest count is for **computer peripherals**, with **174 units**. This suggests a strong presence or demand for peripherals like keyboards, mice, monitors, etc.
- The **Office Machines** sub-category has **95 units**, which is significantly lower than the top two but still substantial. It may include items like printers, shredders, and other office equipment.
- The least represented sub-category is **Copiers and Fax**, with only 18 units.

This graph indicates that declining trend for traditional copiers and fax machines in favour of more modern, multi-functional devices. Overall, the data illustrates a clear preference and demand for communication devices and computer peripherals over traditional office machinery within the technology category.

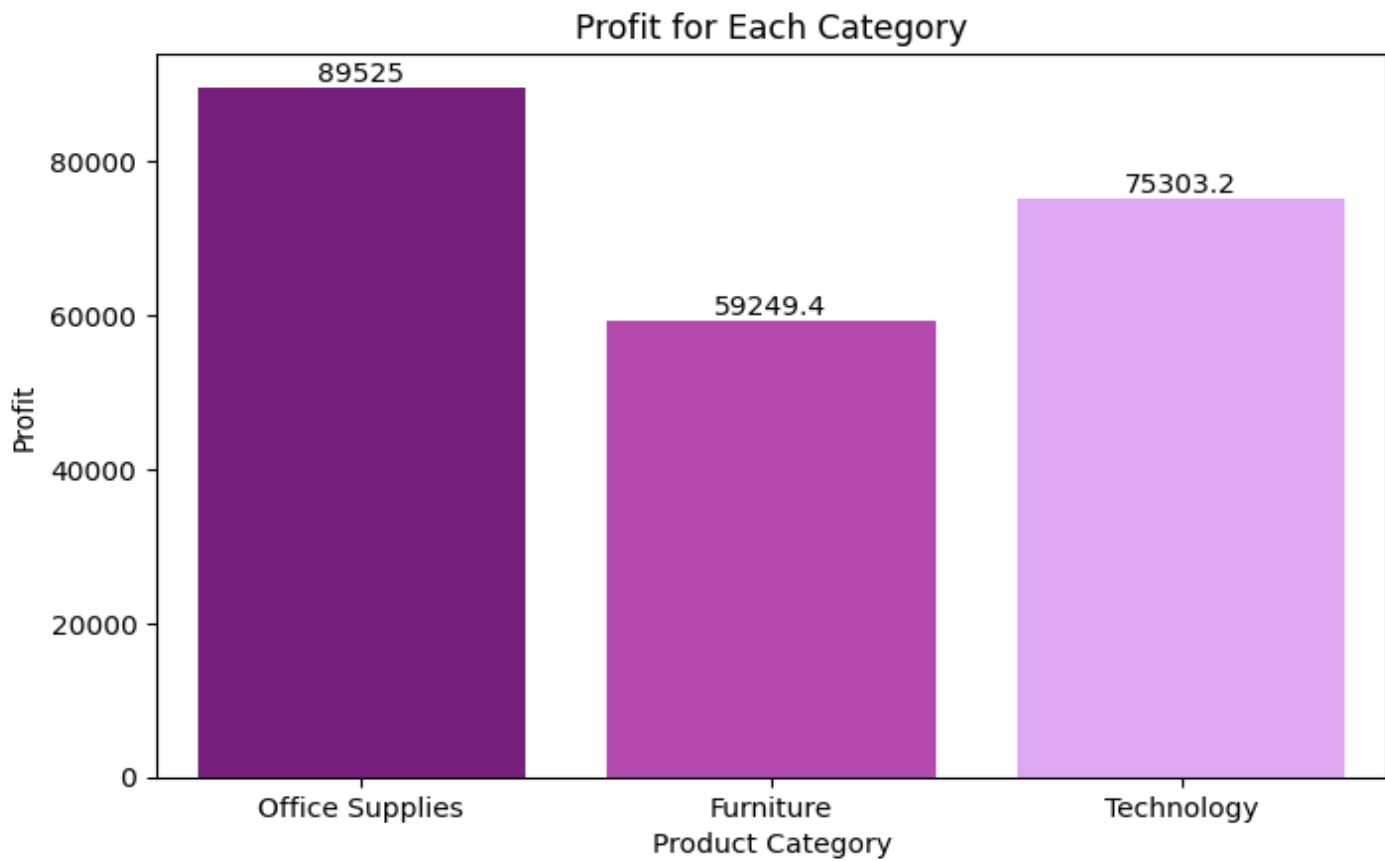
▪ Product sub-category (Furniture)



- The **Office Furnishings** sub-category has the highest count, with **185 units**, indicating that products in this category are the most prevalent or in highest demand within the furniture category.
- The second highest count is for **chairs and chair mats**, with **92 units**.
- The **Tables** sub-category has **80 units**.
- The least represented sub-category is **Bookcases**, with **43 units**.

This reflecting a lower demand or inventory level for bookcases compared to other furniture items. Overall, the data illustrates a clear preference and demand for office furnishings and seating solutions within the furniture category.

6. What is the profit for each product category?



- **Office Supplies:**

- This category has the highest profit, with **89,525**. This indicates that office supplies are the most profitable product category.

- **Technology:**

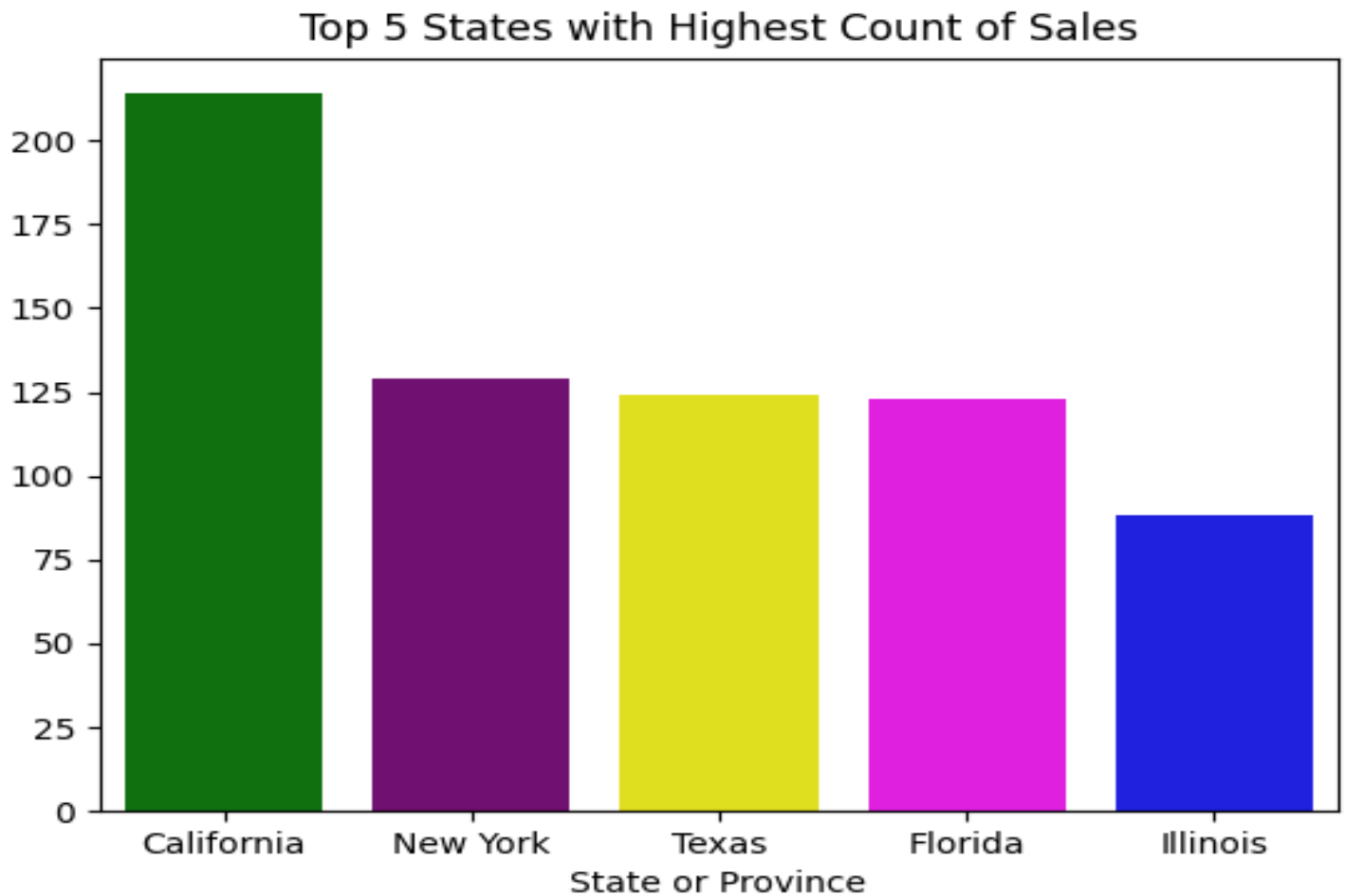
- The second highest profit is from the **technology** category, with **75,303.2**. This suggests that technology products are also highly profitable, though slightly less so than office supplies.

- **Furniture:**

- The **furniture** category has the lowest profit, with **59,249.4**. While still profitable, it lags behind the other two categories.

This data indicates a clear disparity in profit distribution, highlighting that office supplies and technology products are the main profit drivers, whereas furniture generates comparatively lower profit. Overall, the chart underscores the higher profitability of office supplies and technology products within the market.

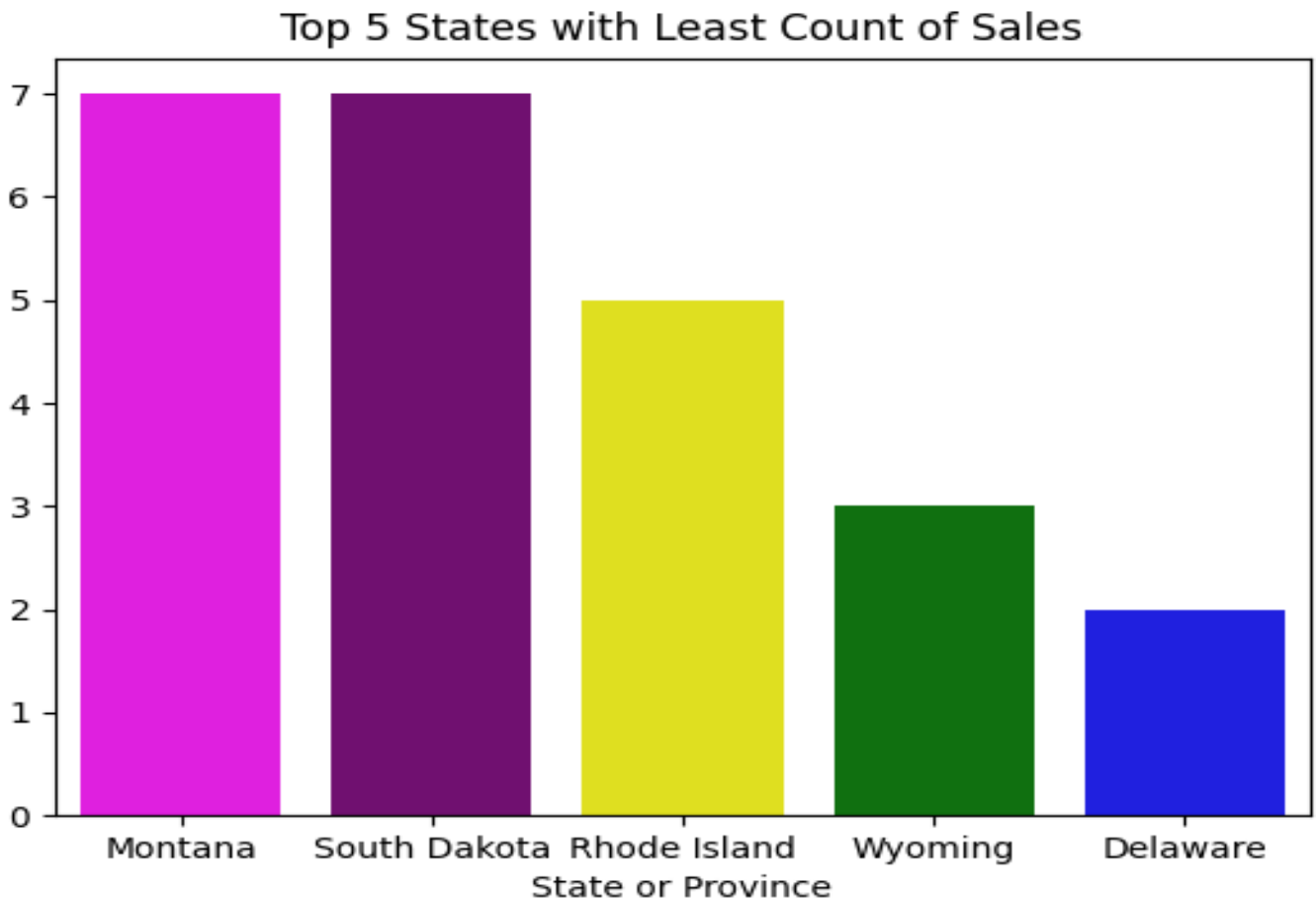
7. Which state has the highest number of sales?



- **California** has the highest number of sales, with a count of **over 200**.
- **New York** follow, with sales counts **around 125**.
- **Florida** and **Taxes** has a sales count slightly below **125 each**.
- **Illinois** has the lowest sales count among the top five states, with a count **around 75**.

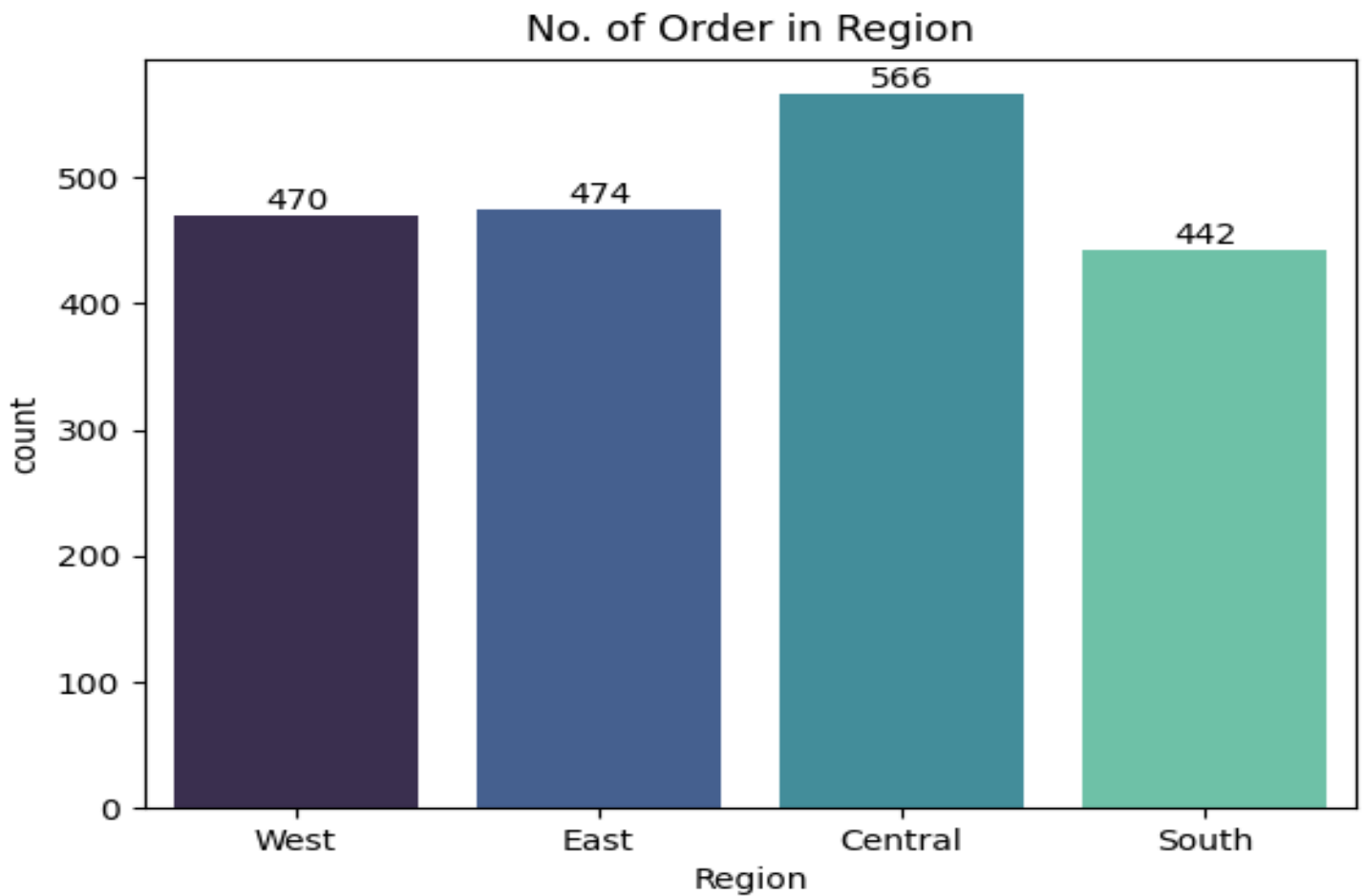
These insights highlight California as a significant market leader in sales. The substantial gap between California and the other states suggests that California is a key area for sales focus. New York and Texas also present strong sales opportunities, while Illinois, with its relatively lower sales, may require additional efforts to boost performance. Understanding the reasons behind California's high sales could provide valuable strategies that might be applied to improve sales in other states.

8. Which state has the lowest number of sales?



- **Delaware has the lowest number of sales:** The chart shows that Delaware has the smallest bar, indicating it has the fewest sales among the five states listed.
- **Wyoming has the second lowest number of sales:** The bar for Wyoming is taller than Delaware's but shorter than the bars for the other states.
- **Rhode Island, South Dakota, and Montana have relatively similar sales counts:** The bars for these three states are close in height, suggesting that they have comparable sales counts.
- **Montana and South Dakota have the highest sales among the five:** Both states have the tallest bars, indicating they have more sales compared to the other states listed.

9. Which region has the highest number of sales?



The region with the highest number of sales is the **Central** region, with a total of **566 orders**.

The regions with the next highest sales are:

East with **474** orders

West with **470** orders

South with **442** orders

This data highlights that while all regions are actively generating sales, the **Central** region is outperforming the others, whereas the **South** region shows a relatively lower sales volume.

10. What is the profit for each region?

Region	Profit
Central	77365.472669
East	85291.403446
South	-14424.054379
West	75844.790100

1. **East Region:** This region has the highest profit at **85,291.40**.
2. **Central Region:** This region has the second-highest profit at **77,365.47**.
3. **West Region:** This region has a profit of **75,844.79**, which is slightly lower than the Central region.
4. **South Region:** This region is operating at a loss, with a negative profit of - **14,424.05**.

This disparity suggests that while some regions are thriving, the South region faces challenges that require immediate attention to address the underlying issues. It may be beneficial to analyse the successful strategies in the East region and consider applying similar approaches to improve the performance of the South region. Additionally, maintaining the performance in the Central and West regions should remain a priority to ensure overall business stability.

Conclusion

The analysis of the superstore dataset reveals several key insights that are crucial for understanding the factors affecting sales and profitability. Regular Air emerges as the most commonly used shipping method, indicating a favourable balance between cost and delivery speed. Office Supplies is the largest and most profitable product category, suggesting high demand and effective inventory management in this segment. California leads in sales, making it a critical market, while states like Delaware show significantly lower sales, highlighting areas needing targeted marketing efforts. The Central region outperforms others in sales, indicating robust market presence. These insights underscore the need for a strategic approach focusing on high-demand regions and categories to reverse declining sales and drive sustained growth.