Superstore Sales Analysis

Overview

This report presents a comprehensive analysis of the Superstore Sales Dataset, a popular dataset used for retail sales analytics. The dataset contains sales data of a US-based superstore, including order details, customer information, product categories, and geographical data.

Key Objectives:

- Perform Exploratory Data Analysis (EDA) to uncover trends and patterns.
- Identify primary and secondary KPIs affecting sales and profitability.
- Analyze customer behavior, regional performance, and product category trends.
- Conduct time series forecasting (SARIMA) to predict future sales.

Dataset Source:

- Dataset Link: Superstore Dataset (Kaggle)
- GitHub Repository: GitHub - Superstore Analysis

Data Analysis Process (EDA)

- Missing Values: Identified and addressed missing entries, notably in the 'Postal Code' column.
- Data Types: Converted 'Order Date' and 'Ship Date' to datetime objects for temporal analysis.
- Duplicates: Removed duplicate records to ensure data integrity.
- Outliers: Detected and handled outliers in 'Sales' and 'Profit' using interquartile range (IQR) method.

```
print(df.isnull().sum())
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
City            0
State           0
Postal Code     11
Region          0
Product ID      0
Category        0
Sub-Category    0
Product Name    0
Sales           0
dtype: int64
```
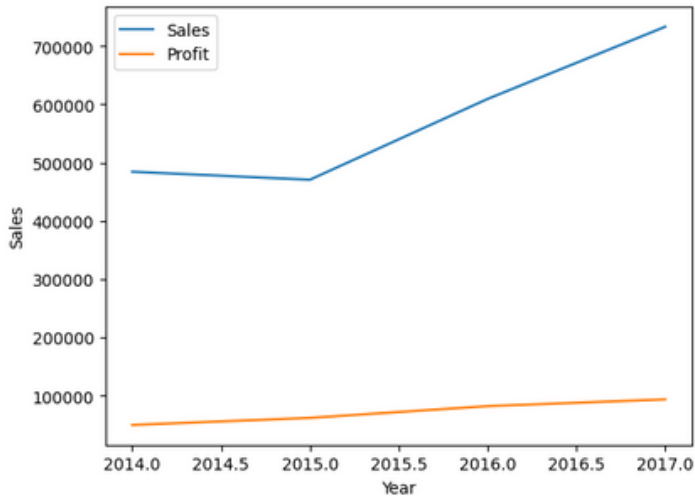
We can see that the postal code is not mentioned only for Burlington city in Vermont state. So, we need to fill the postal code of that city.

```
df['Postal Code'] = df['Postal Code'].fillna(5401)
```

```
df['Order Date'] = pd.to_datetime(df['Order Date'], dayfirst=True)
df['Ship Date'] = pd.to_datetime(df['Ship Date'], dayfirst=True)
```
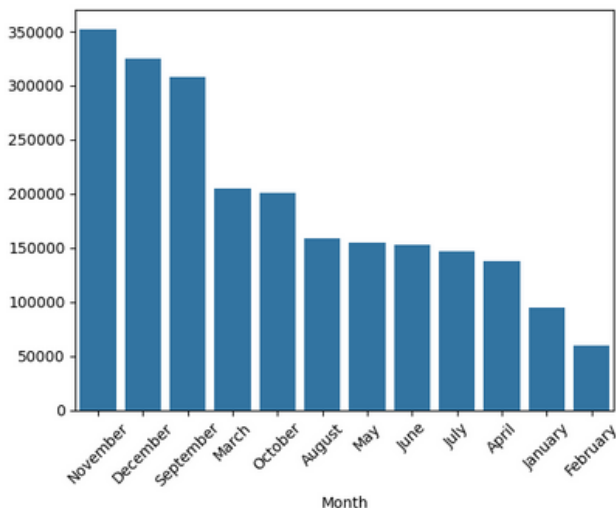
Analysis

1. **What are the overall sales and profit trends over time?**

- Sales grow YoY, peaking in 2017 ($484k).
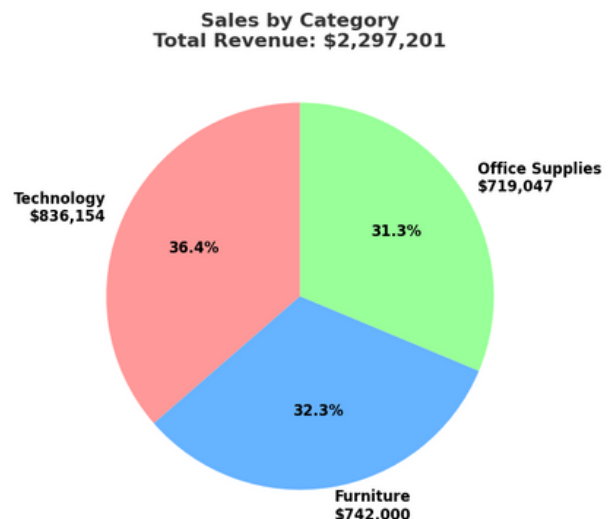- Profit fluctuates, highest in 2016 ($61k).

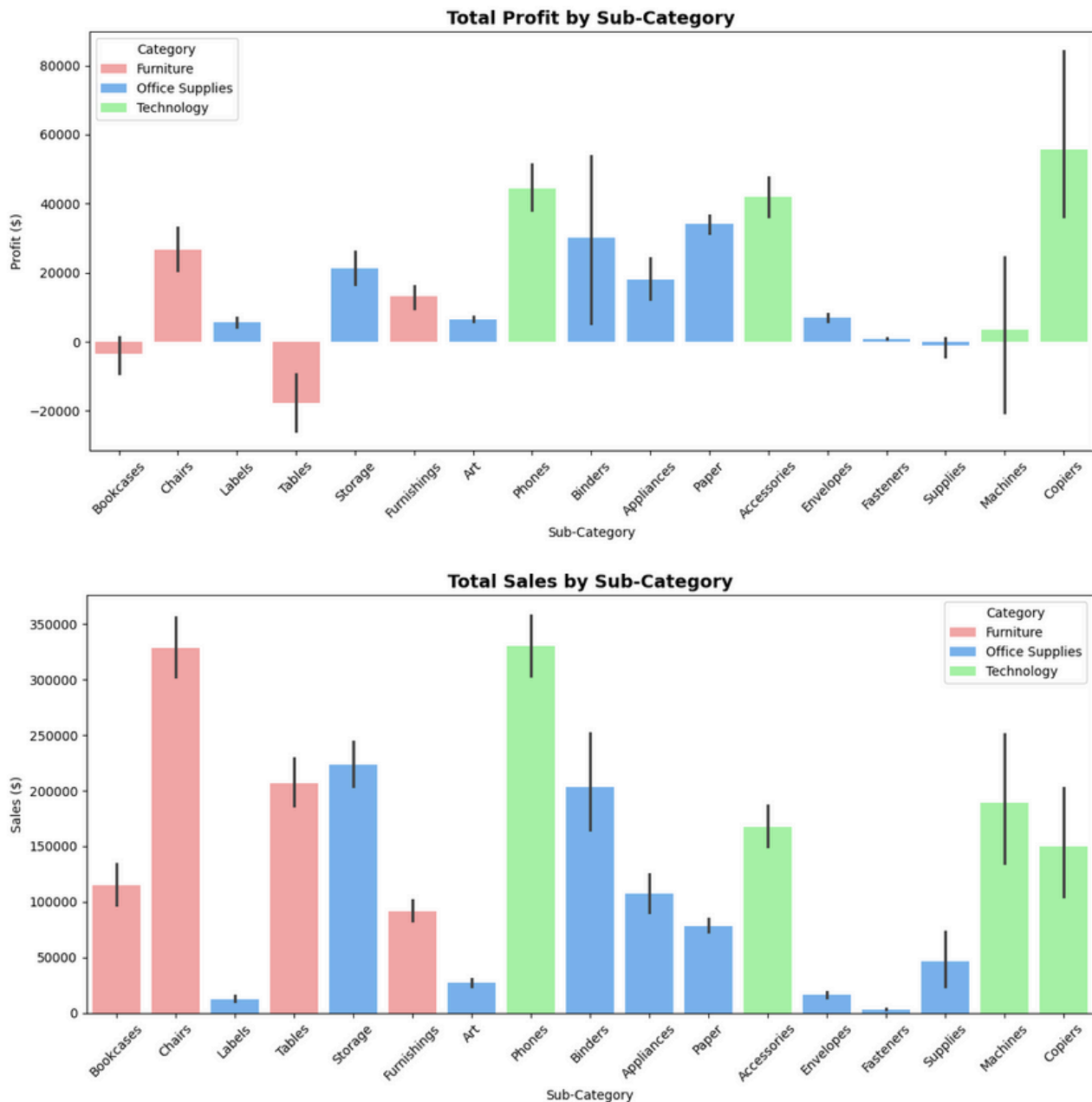2. **Which months have the highest sales?**

- We can see that November has the highest Sales, most probably due to holiday season.

3. **Which product category has the highest sales?**

- We can see that Category Technology generated the highest revenue of about $836,154.
- The Total Revenue generated by all the categories: $2,297,201

Sales by Category
Total Revenue: $2,297,201

Technology $836,154 — 36.4%
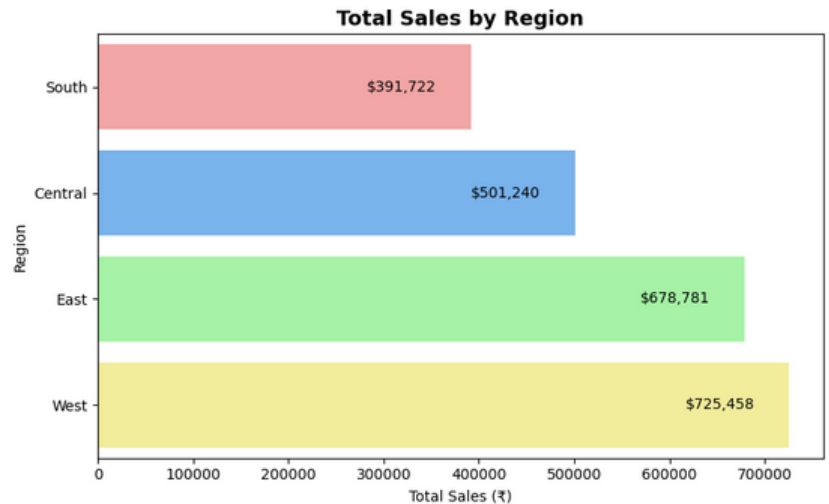Office Supplies $719,047 — 31.3%
Furniture $742,000 — 32.3%

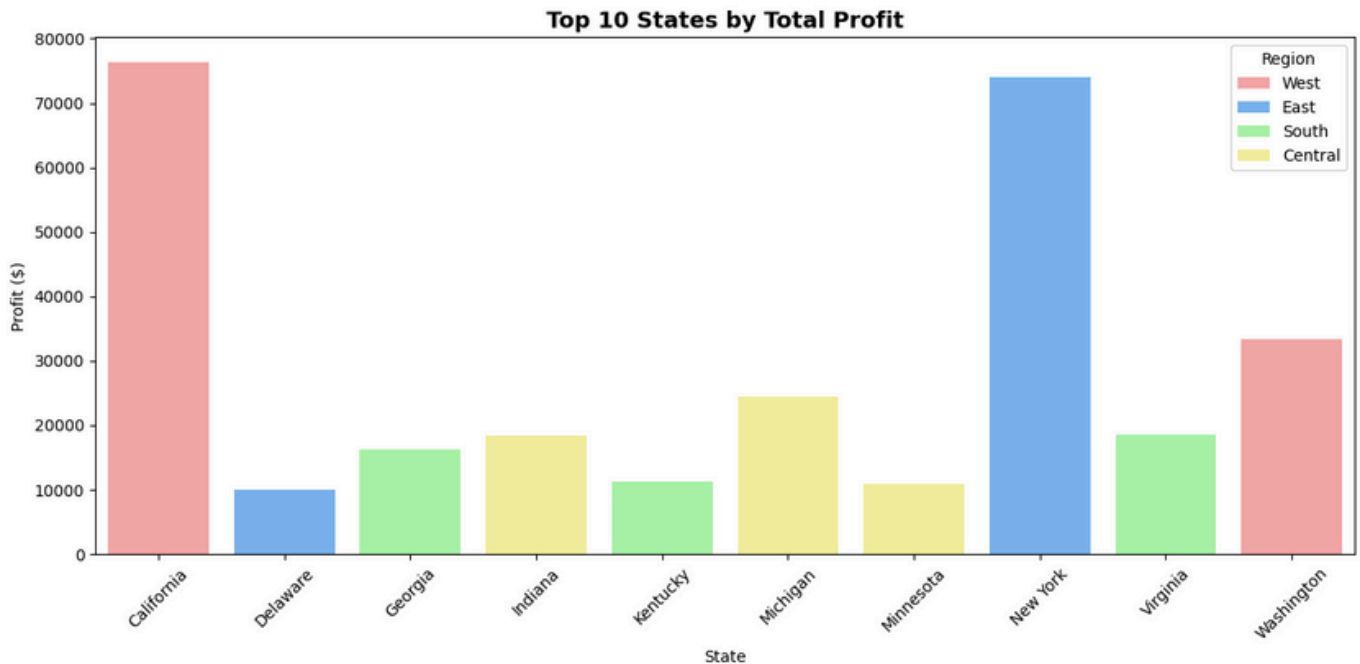## 4. Sales vs. Profit by Category and Sub-Category:

- The Technology category emerged as the leading performer in both sales and profit. Within this category, the Phones sub-category stood out with impressive sales and profit figures.
- On the other hand, the Furniture category showcased significant sales but had a notably lower profit margin. Within the Furniture category, the Tables displayed decent sales but operated at a loss.
- Additionally, both Bookcases and Supplies resulted in a negative profit for the company.

5. Which region has the highest sales?

- The West region boasted the highest revenue, with sales amounting to $725K, followed by the East region with sales of $678K.
- The Central region with sales of $501K, and the South with sales of $392K

Total Sales by Region

Region	Total Sales (₹)
South	$391,722
Central	$501,240
East	$678,781
West	$725,458

6. Which state contributes the most to profit?


Top 10 States by Total Profit

- California and Washington from West Region generated the highest profit.
- From the East Region, we can see New York leading.

## 7. How does shipping mode affect profit margins?

- Most customers preferred Standard Class shipping, indicating that they might prioritize cost savings over faster delivery.

8. Is there a correlation between discount and profit?

We can see that there is a negative correlation between Profit and Discount. Higher discounts lead to reduction in profit.
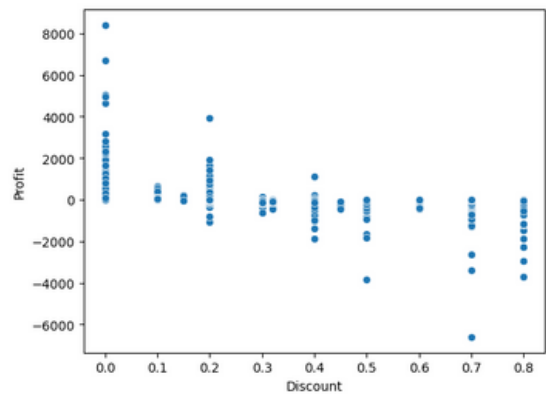
9. Which customer segment is most profitable?
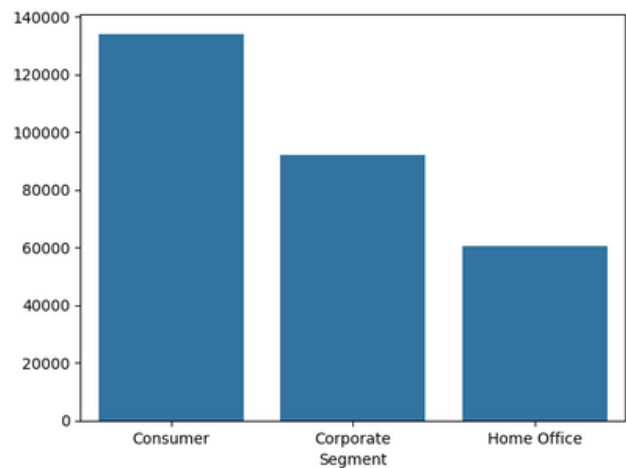
The Consumer segment dominates in profit. This suggests that consumer-oriented products might be the most profitable items.

10. Who are the top 5 customers by sales?

Raymond Buch, Sean Miler, and Tamara Chand were the most frequent shoppers with a max of 6 orders for the past 4 years. This pattern suggests a potential gap in customer loyalty.

Customer Name	
Sean Miller	25043.050
Tamara Chand	19052.218
Raymond Buch	15117.339
Tom Ashbrook	14595.620
Adrian Barton	14473.571