**1. Superstore Sales Analysis**

**1.1 Overview**

This sales analysis project focuses on analyzing the overall sales performance of a U.S.-based Superstore. It involves the development of an interactive dashboard that enables users to explore sales data across various dimensions, providing in-depth analysis of order trends, product performance, and customer segmentation.

**1.2 Objectives**

* Evaluate overall sales performance to identify key trends and patterns over time.
* Determine best-selling products and categories by analyzing sales volume.
* Analyze customer segments to understand their distribution and highlight the most frequent buyers.
* Examine order trends across different time periods, regions, and channels to uncover seasonality and growth opportunities.
* Provide a user-friendly, interactive dashboard that enables stakeholders to explore and interpret sales data with ease.
* Support data-driven decision-making in marketing, inventory management, and sales strategy.

**2. Data Source and Description**

Dataset is in a csv format, including the following fields:

1. Order ID (String) – A unique identifier for each order.
2. Order Date (Date) – The date when the order was placed.
3. Ship Date (Date) – The date when the order was shipped.
4. Ship Mode (String) – The shipping method used (e.g., Standard Class, Second Class).
5. Customer ID (String) – A unique identifier for each customer.
6. Customer Name (String) – The full name of the customer.
7. Segment (String) – The type of customer (Consumer, Corporate, Home Office).
8. City (String) – The city where the order was placed.
9. State (String) – The state where the order was placed.
10. Region (String) – The region of the country (West, East, Central, South).
11. Product ID (Whole number) – A unique identifier for each product.
12. Category (String) – The broad category of the product (Furniture, Office Supplies, Technology).
13. Sub-Category (String) – A more specific classification of the product (e.g., Chairs, Labels, Storage).
14. Product Name (String) – The name of the product.
15. Sales (Float) – The total revenue generated from the sale.

**3. Analysis Methodology**

**3.1 Tools**

* Power BI (Power Query – Dax)

**3.4 Analysis Questions**

* Total sales by year/quarter/month
* ⁠Total sales by state/city
* Total sales by product category
* Total sales by customer segment
* Sales growth over years
* Number of orders by year/quarter/month
* Top 5 products by sales
* Top 5 products by Number of Orders
* Distribution of customers across segments
* Number of orders by customer
* Key influencers to reduce Average shipment duration

**4. Data Cleaning**

1. Load Data into Power Query
2. Convert Data Types (Ensure that each column has the correct data type):  
   - Order Date & Ship Date → Convert to Date format "Using Locale" to specify the correct format.  
   - Sales → Convert to Decimal Number.  
   - Other categorical columns (e.g., Customer ID, Product ID, Ship Mode) should be Text.
3. Remove duplicates and errors  
   - Select all columns  
   - Click Remove Duplicates under the Home tab.  
   - Remove errors
4. Add Custom Column "Shipping Duration"

- Click Add Column > Custom Column.

- Use the formula: Shipping Duration (Days)=Duration.Days([Ship Date] - [Order Date])

1. Replaced empty cells in Postal Code Column

Each city is supposed to have a unique Postal Code for identification, While “Burlignton” had an empty cell in the city column.

- Select column “City”

- From Replace Values in Transform tab, value to find is an empty string “” replace with “00001”

1. Defined New Postal code for “Encinitas”

Two cities California and Encinitas had the same Postal Code, the solution was to define a new postal code for one of the two cities.

- Add new Postal Code column using formula : if ([City]="Encinitas" and [Postal Code]="92024") then "00000" else [Postal Code]

- Delete the old Postal code Column

1. Defined a new Product Id

A problem we face was that several products had the sames Product ID which may lead to inaccurate analysis reuslts.

* Removed the Product ID column
* Duplicated the table and removing all columns except for Product Name
* Removed all duplicates in product name table
* Populated a new Product ID using Index Column from Add Column tab
* Now this table will serve as lookup table for Product ID
* Then merged the main table and lookup table on Product Name column with join kind left join

**5. Data Modeling (Detailed)**

**Overview**

The data model was structured using a star schema design to ensure clarity, performance, and ease of use. The model is centered around a single fact table ("Sales") and multiple dimension tables for slicing and filtering.

**5.1 Tables Created**

**5.1.1 Sales Table (Fact Table)**

* Contains the transactional sales data.
* Key columns include:
  + Order ID, Order Date, Customer ID, Product ID, Ship Date, Sales, Ship Mode, Shipment Duration, and Postal Code.

**5.1.2 Date Table**

* Created using DAX to support time-based analysis.
* Includes derived columns such as:

Day, Day of Week, Month, Month Name, Quarter, Quarter Name, and Year.

DAX used for creating the base date range:

DAX

Date = CALENDAR(MIN(Sales[Order Date]), MAX(Sales[Order Date]))

**5.1.3 Product Table**

* Contains product-related descriptive data:
  + Product ID, Product Name, Category, Sub-Category.

**5.1.4 Customer Table**

* Contains customer attributes:
  + Customer ID, Customer Name, and Segment**.**

**5.1.5 Address Table**

* Stores geographic data for analysis and mapping:
  + City, State, Region, Country, Postal Code.

**5.2 Relationships Between Tables**

The following one-to-many relationships were defined (direction: from dimension to fact):

* Date[Date] → Sales[Order Date]
* Product[Product ID] → Sales[Product ID]
* Customer[Customer ID] → Sales[Customer ID]
* Address[Postal Code] → Sales[Postal Code]

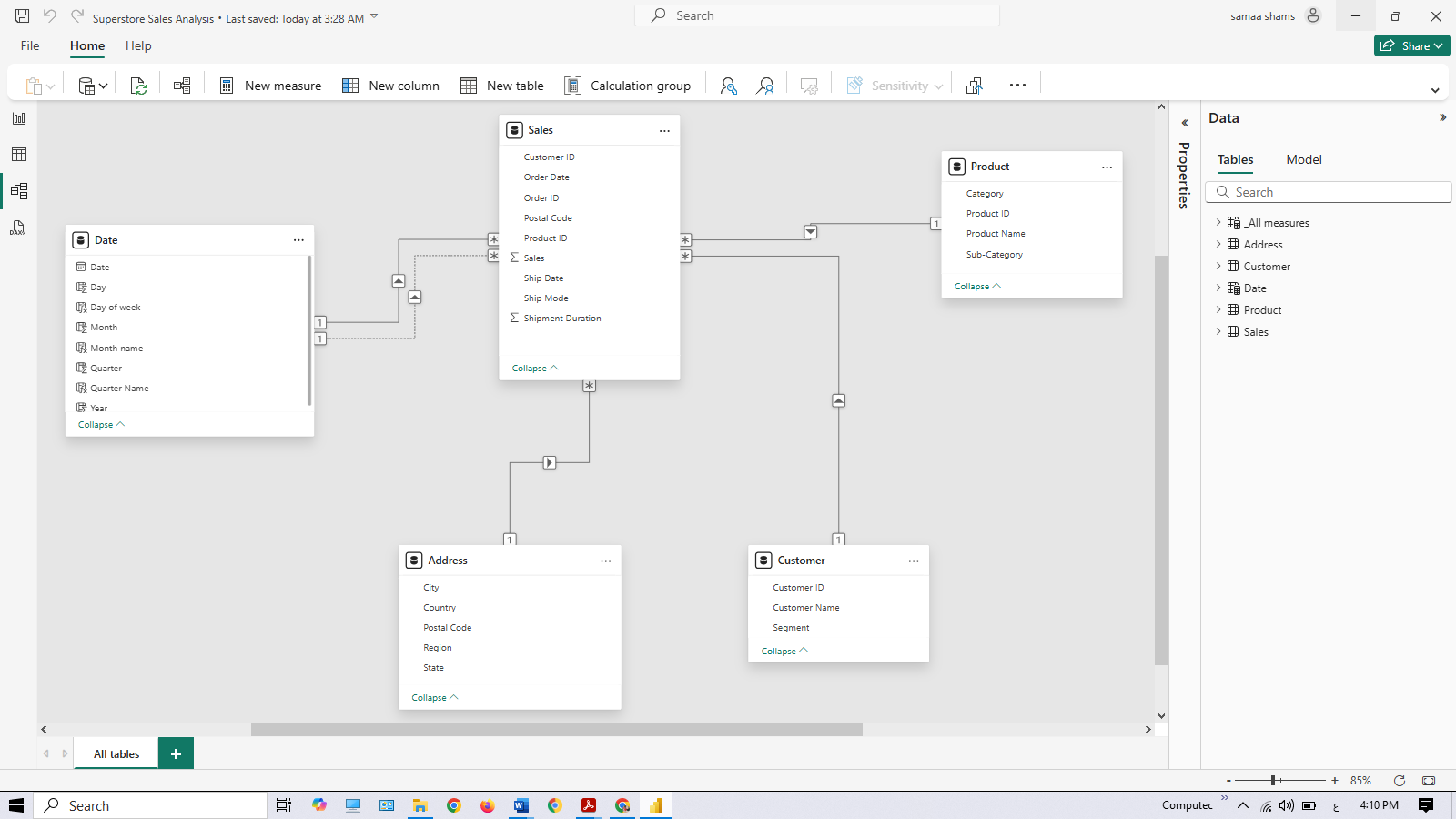
These relationships enable filtering and grouping of sales data by time, product, customer, and geography.

**5.2.1 Cardinality & Direction:**

* All relationships are one-to-many (1:\*) with single-directional filtering, which avoids ambiguity and performance issues.

**5.3.1 Why This Model Works**

* Improves report performance and reduces redundancy.
* Keeps analysis logic clean and makes DAX calculations simpler.
* Ensures a scalable structure for future enhancements (e.g., adding cost/profit tables later)



**6. Analysis Methodology**

**6.1 DAX Measures:**

* Total Sales = SUM(Sales[Sales])
* Order Count = DISTINCTCOUNT(Sales[Order ID])
* Number of customers = COUNTROWS('Customer')
* Number of products = COUNTROWS('Product')
* Average Ship days = AVERAGE(Sales[Shipment Duration])

**6.2 Dashboard Creation**

The Superstore Sales Analysis project features a set of interactive dashboards designed to provide a comprehensive overview of sales performance, customer behavior, product trends, and shipping efficiency. The dashboards were developed using Power BI and are organized into the following main sections:

* **Sales Dashboard:** Presents high-level KPIs such as total sales, total orders, number of customers, number of products, and average shipment days. It also includes trend analysis by year, sales by region, state, segment, and product category.
* **Orders Dashboard:** Focuses on order distribution by state, city, region, year, segment, and shipping mode. It highlights top-performing states and cities, and visualizes order trends over time.
* **Products & Customers Dashboard:** Provides insights into sales by product sub-category, top products by sales and order count, top customers by order volume, and customer segmentation.
* **Key Influencers Dashboard:** Utilizes AI-driven analysis to identify the main factors influencing key metrics such as average shipping days and total sales.
* **Decomposition Tree:** Allows users to drill down into sales data by region, state, segment, year, category, and shipping mode to uncover detailed patterns and drivers of performance.

Each dashboard is designed to enable users to interactively explore the data, filter by different dimensions, and gain actionable insights for business decision-making.

**6.3 Visualization**

Visuals of Sales Dashboard:

* Card: Total Sales.
* Card: Total Orders.
* Card: Number of Customers.
* Card: Number of Products.
* Card: Average Shipment Days.
* Line chart: Total sales by year.
* Bar chart: Sales year-over-year growth % by year.
* Bar chart: Total sales by region.
* Pie chart: Total sales by product category.
* Pie chart: Total sales by segment.
* Map: Total sales by state.

Visuals of Orders Dashboard:

* Bubble chart: Total orders by state.
* Treemap: Top 5 cities by total orders.
* Donut chart: Total orders by ship mode.
* Line chart: Total orders by year.
* Bar chart: Total orders by region.
* Donut chart: Total orders by segment.

Visuals of Products & Customers Dashboard:

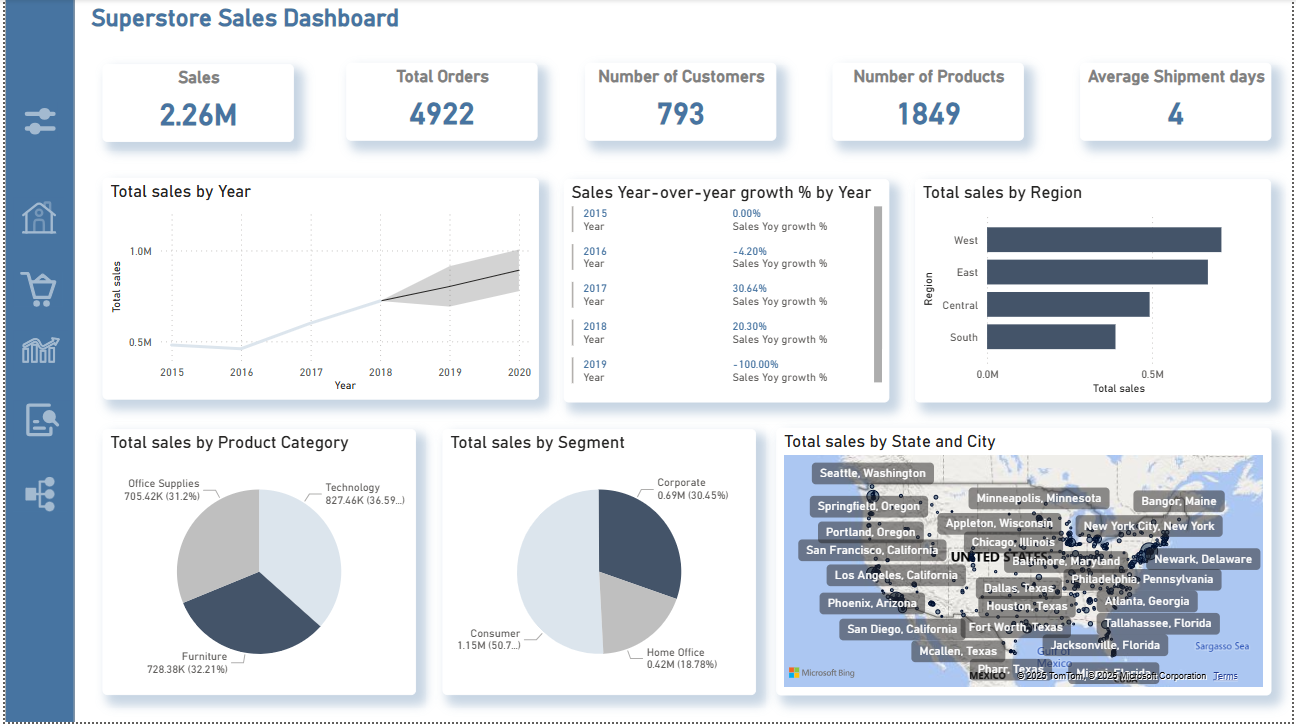
* Bar chart: Total sales by sub-category.
* Scatter plot: Top 5 products by total sales.
* Bar chart: Top 5 products by orders.
* Bar chart: Count of customers by segment.
* Table: Total orders by customer.

Visuals of Key Influencers Dashboard:

* Key Influencers visual: What influences average ship days to decrease (analyzing by state, city, region, date, category, product name, segment, customer name, ship mode).
* Key Influencers visual: What influences total sales to increase (analyzing by state, city, region, segment, customer name, date, category, sub-category, product name).

Visuals of Decomposition Tree Dashboard:

* Decomposition tree: Drill-down of total sales by region, state, segment, year, category, and ship mode.

****

**Key Findings**

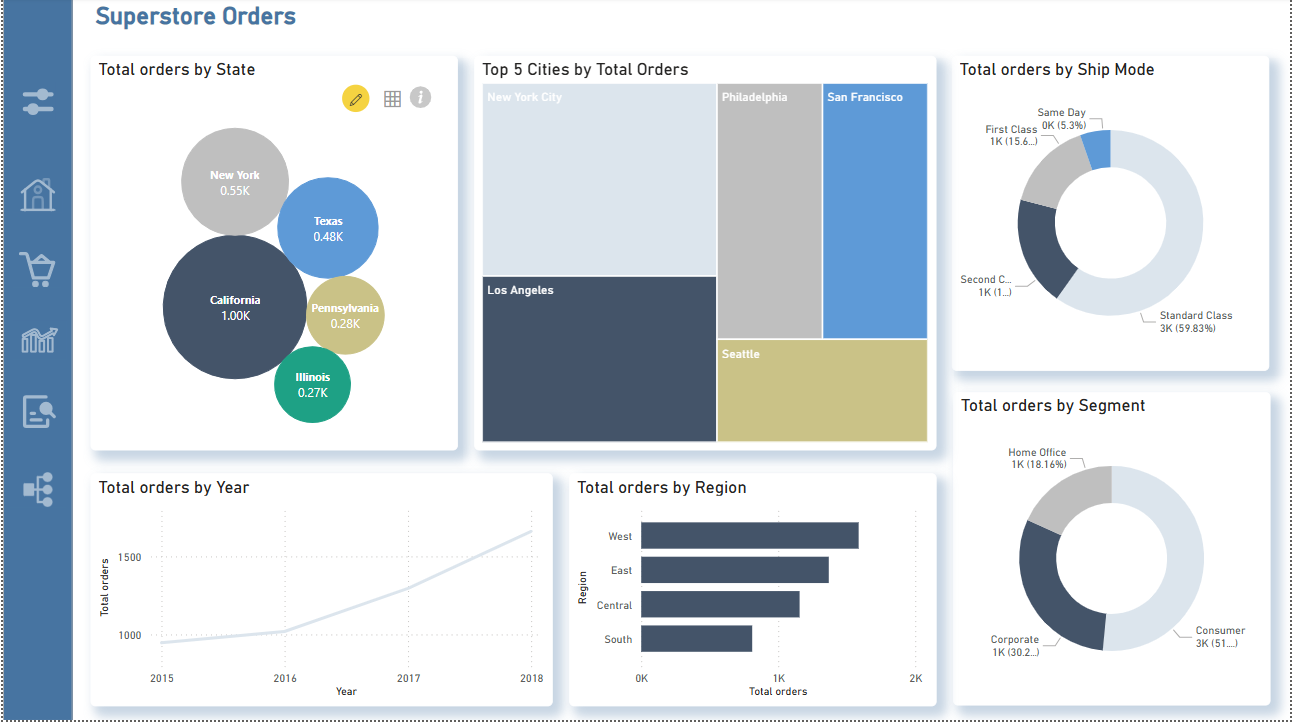
* Sales Growth: The dashboard shows a clear upward trend in total sales from 2015 to 2018, with the most significant year-over-year growth occurring in 2017 (30.64%) and continued strong growth in 2018 (20.3%).
* Regional Performance: The West region leads in total sales, followed by the East, Central, and South regions. This highlights the importance of focusing marketing and logistics efforts in the West and East regions.
* State-Level Insights: California stands out as the top-performing state in terms of sales, with other high-performing states including New York, Texas, and Pennsylvania.
* Product Category Distribution: Technology products account for the largest share of sales (36.5%), followed by Furniture (32.2%) and Office Supplies (31.2%). This indicates a balanced demand across major product categories, with a slight edge for technology.
* Customer Segmentation: The Consumer segment is responsible for over half of total sales (50.7%), while Corporate and Home Office segments contribute 30.45% and 18.78% respectively.
* Customer and Product Base: The business serves 793 unique customers and offers 1,849 distinct products, reflecting a broad and diverse market reach.
* Shipping Efficiency: The average shipment time is 4 days, indicating efficient order fulfillment processes.

**Analysis Summary**

The Superstore Sales Dashboard provides a comprehensive overview of business performance across multiple dimensions. The analysis reveals robust sales growth over the years, with the West region and California emerging as the primary drivers of revenue. Technology, Furniture, and Office Supplies are the leading product categories, catering to a wide range of customer needs.

The Consumer segment dominates sales, suggesting that targeted marketing and customer retention strategies for this group could yield significant returns. The business maintains a healthy customer and product base, supporting sustained growth and market expansion.

Efficient shipping processes, as indicated by the average shipment days, further enhance customer satisfaction and operational effectiveness. Overall, the dashboard highlights key areas of strength and opportunities for further growth, providing actionable insights for strategic decision-making.

****

**Key Findings**

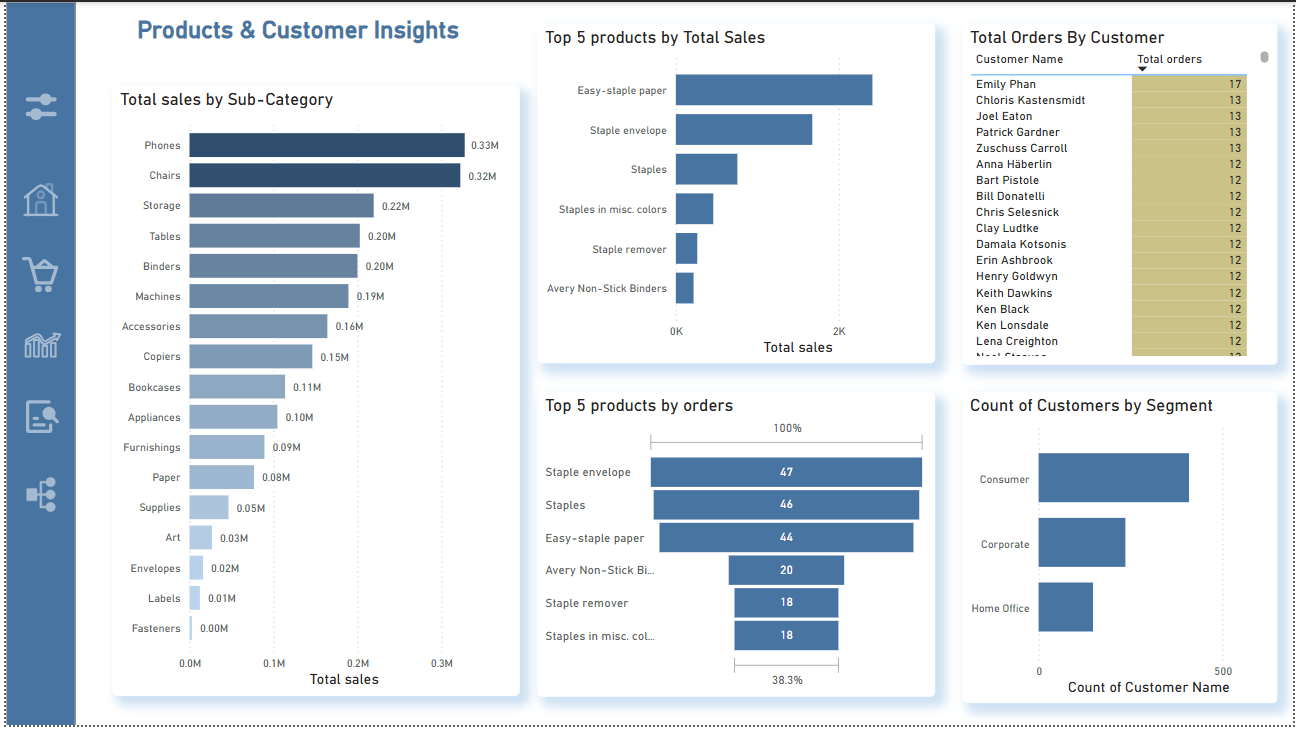
* State-Level Order Distribution: California leads all states with the highest number of orders (1,000), followed by New York (550), Texas (480), Pennsylvania (280), and Illinois (270). This highlights California as the primary market for order volume.
* Top Cities: The cities with the most orders are Los Angeles, New York City, Philadelphia, San Francisco, and Seattle, indicating strong urban demand in both coastal and central metropolitan areas.
* Shipping Mode Preferences: The majority of orders (59.83%) are shipped via Standard Class, with Second Class (19.1%) and First Class (15.6%) also being used. Same Day shipping accounts for a smaller share (5.3%), suggesting that most customers opt for standard delivery.
* Order Growth Over Time: There is a steady increase in total orders from 2015 to 2018, reflecting business growth and expanding customer engagement.
* Regional Order Trends: The West region has the highest order volume, followed by East, Central, and South, mirroring the sales distribution and reinforcing the importance of the West and East regions.
* Customer Segmentation: The Consumer segment is responsible for the majority of orders (51%), with Corporate (30.2%) and Home Office (18.16%) segments also contributing significantly.

**Analysis Summary**

The Superstore Orders Dashboard provides a detailed view of order distribution across states, cities, regions, shipping modes, and customer segments. The analysis reveals that California and major metropolitan cities are the main drivers of order volume, emphasizing the importance of these markets for business growth.

Standard Class shipping is the preferred method for most customers, likely due to its balance of cost and delivery time. The steady year-over-year increase in orders demonstrates the company’s ability to attract and retain customers, while the dominance of the Consumer segment suggests that marketing and service strategies should continue to focus on this group.

Regional analysis confirms that the West and East regions are the most active, both in terms of sales and order count. The dashboard’s insights support data-driven decision-making for targeted marketing, logistics optimization, and customer engagement strategies.

****

**Key Findings**

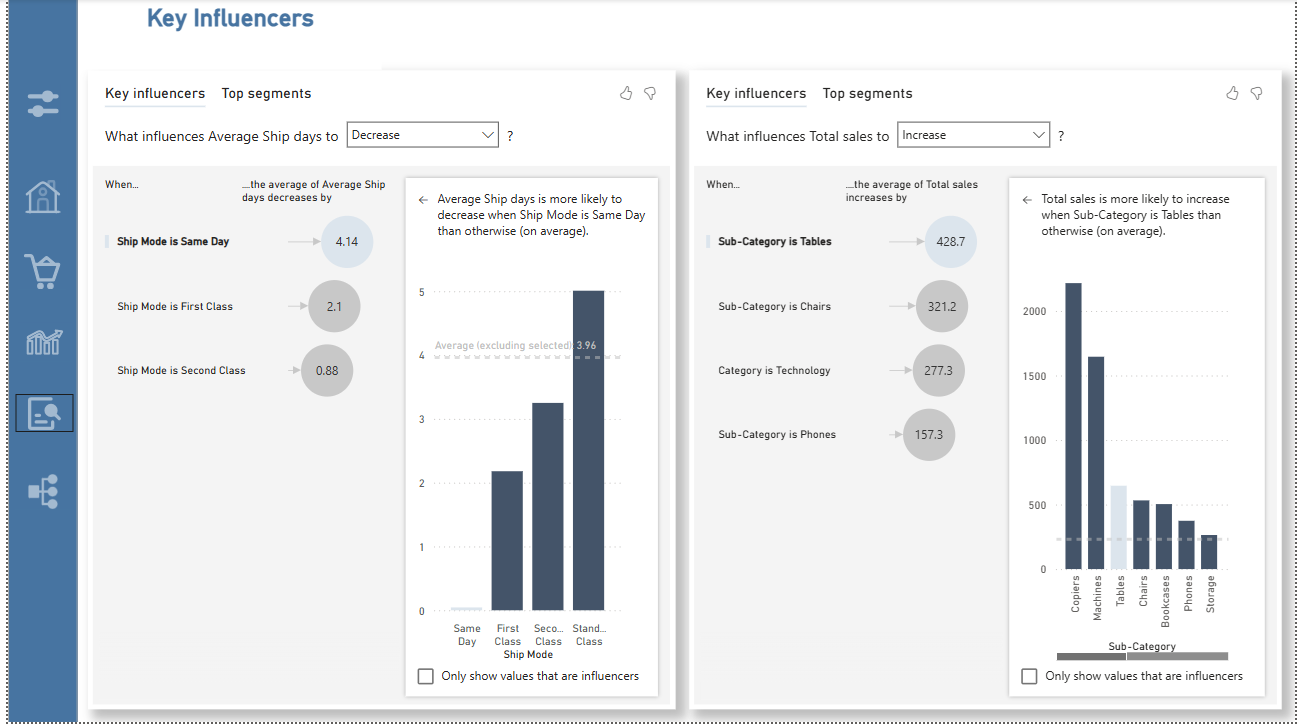
* Top Sub-Categories by Sales: Phones and Chairs are the leading sub-categories, generating $0.33M and $0.32M in sales respectively. Other high-performing sub-categories include Storage, Tables, Binders, and Machines.
* Best-Selling Products: The top products by total sales are Easy-staple paper, Staple envelope, and Staples, each contributing significantly to overall revenue.
* Most Ordered Products: Staple envelope, Staples, and Easy-staple paper are the top three products by order count, indicating strong and consistent demand for these items.
* Top Customers: Emily Phan is the most active customer with 17 orders, followed by several customers with 13 orders each, highlighting a group of highly engaged and loyal buyers.
* Customer Segmentation: The Consumer segment has the largest customer base, followed by Corporate and Home Office segments. This distribution suggests that most customers are individual consumers rather than business or home office clients.

**Analysis Summary**

The Products & Customer Insights Dashboard provides a detailed breakdown of sales and order patterns across product sub-categories, individual products, and customer segments. The analysis reveals that a small number of sub-categories and products account for a large share of total sales and orders, with Phones, Chairs, and essential office supplies leading the way.

Customer analysis shows a core group of highly engaged buyers who place frequent orders, offering opportunities for targeted loyalty programs and personalized marketing. The dominance of the Consumer segment in the customer base further emphasizes the importance of focusing on individual consumers for future growth.

Overall, the dashboard highlights the importance of maintaining strong inventory levels for top-selling products, nurturing relationships with high-value customers, and tailoring marketing strategies to the largest customer segment. These insights support data-driven decisions to maximize sales and enhance customer satisfaction.

****

**Key Findings**

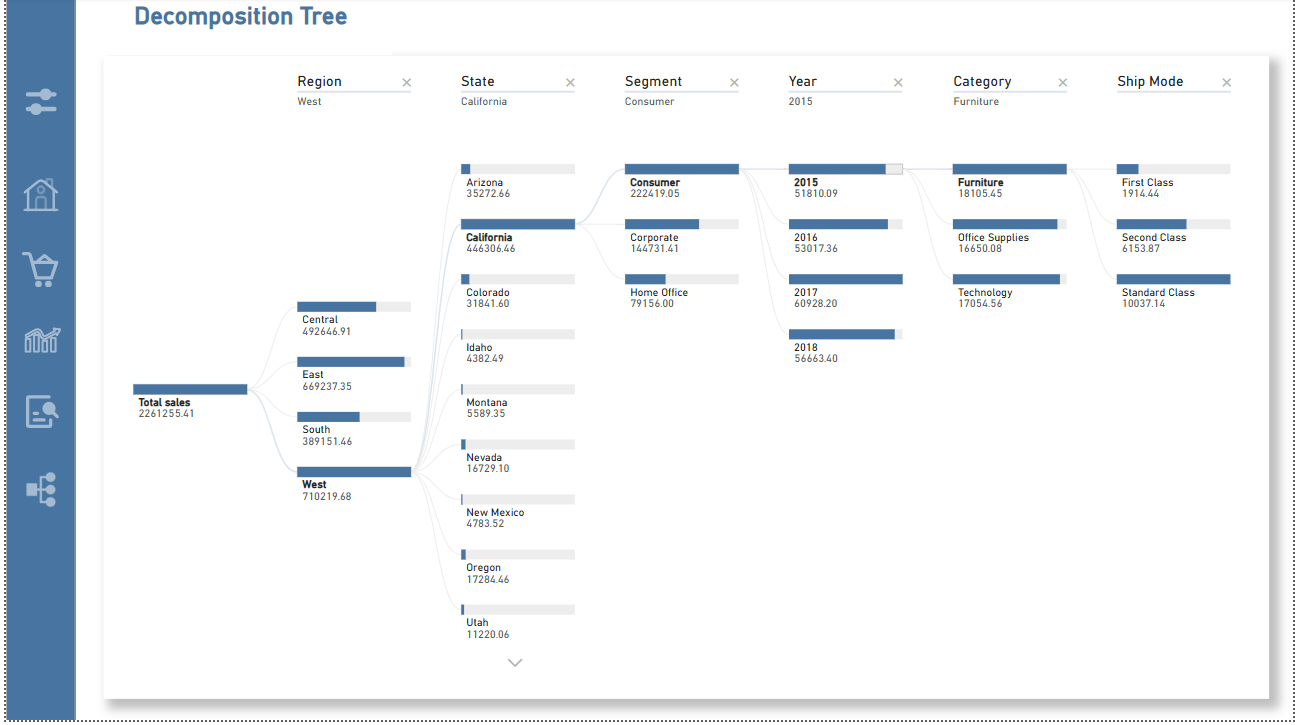
* **Average Ship Days Influencers:**The analysis identifies Ship Mode as the most significant factor influencing the reduction of average shipping days. Specifically:
  + When the ship mode is Same Day, the average shipping days decrease by 4.14 days compared to the overall average.
  + First Class shipping reduces average shipping days by 2.1 days.
  + Second Class shipping reduces average shipping days by 0.88 days.
  + The average shipping days for Standard Class is the highest among all modes.
  + Other variables analyzed (state, city, region, date, category, product name, segment, customer name) had less impact compared to ship mode.
* **Total Sales Influencers:**The analysis reveals that product sub-category and category are the most influential factors in increasing total sales. Specifically:
  + When the sub-category is Tables, average total sales increase by 428.7 compared to the overall average.
  + Chairs and Technology also significantly boost total sales, with increases of 321.2 and 277.3 respectively.
  + The sub-category Phones increases average total sales by 157.3.
  + Other variables analyzed (state, city, region, segment, customer name, date, product name) had a lesser effect compared to sub-category and category.

**Analysis Summary**

The Key Influencers Dashboard leverages AI-driven analysis to uncover the primary drivers behind two critical business metrics: average shipping days and total sales. The findings demonstrate that shipping mode is the dominant factor in reducing delivery times, with Same Day and First Class options providing the fastest fulfillment. This insight highlights the importance of offering and promoting expedited shipping options to enhance customer satisfaction.

For total sales, the product sub-category—particularly Tables, Chairs, and Technology—emerges as the strongest influencer. Focusing on these high-performing categories can drive revenue growth and inform inventory and marketing strategies.

The analysis, which considered a wide range of variables including geographic location, customer segment, and product details, confirms that operational choices (like shipping mode) and product mix (sub-category and category) have the greatest impact on key performance indicators. These insights empower the business to make targeted improvements in logistics and product offerings to maximize efficiency and profitability.

****

**Key Findings**

* **Sales Breakdown by Region**: The West region is the top contributor to total sales, generating $710,219.68, followed by the East, Central, and South regions.
* **State-Level Insights**: Within the West region, California stands out as the leading state with $446,306.46 in sales, far surpassing other states such as Arizona, Colorado, and Nevada.
* **Customer Segment Impact**: The Consumer segment is the primary driver of sales in California, accounting for $222,419.05, while Corporate and Home Office segments contribute less.
* **Yearly Trends**: Sales in the Consumer segment in California are distributed across the years, with 2017 being the peak year ($60,928.20), followed by 2018, 2016, and 2015.
* **Category Performance**: Within the Consumer segment, Furniture is the top-selling category in 2015, with $18,105.45 in sales, followed by Office Supplies and Technology.
* **Shipping Mode Influence**: For Furniture sales in 2015, Standard Class is the most used shipping mode ($10,037.14), followed by Second Class and First Class.

**Analysis Summary**

The Decomposition Tree Dashboard provides a powerful, step-by-step breakdown of total sales, allowing users to drill down from the overall business performance to granular details by region, state, customer segment, year, product category, and shipping mode. The analysis reveals that the West region—especially California—is the dominant market for Superstore sales, with the Consumer segment driving the majority of revenue.

Year-over-year analysis highlights 2017 as the strongest year for sales in the top-performing segment and state. Among product categories, Furniture leads in sales, and Standard Class is the preferred shipping method for these high-value orders.

This decomposition approach enables the business to pinpoint exactly where sales are strongest and which combinations of geography, customer type, product, and logistics contribute most to success. These insights support targeted strategies for regional growth, customer engagement, product focus, and operational efficiency.

**7. Conclusion**

The Superstore Sales Analysis dashboards deliver a holistic view of business performance across multiple dimensions. Key findings from the analysis include:

* **Geographic Insights:** California is the leading state in both sales and order volume, with major cities like Los Angeles and San Francisco driving significant business. The West region outperforms other regions in total sales.
* **Product Performance:** Technology and Furniture are the top-selling categories, with sub-categories like Phones and Chairs leading in sales. Products such as "Staple envelope" and "Staples" are among the most frequently ordered.
* **Customer Segmentation:** The Consumer segment accounts for the majority of sales and orders, highlighting the importance of targeting this group with tailored marketing strategies.
* **Shipping Efficiency:** Most orders are shipped via Standard Class, but Same Day shipping significantly reduces delivery time. Optimizing shipping methods can further enhance customer satisfaction.
* **Growth Trends:** There is a clear upward trend in both sales and order volume from 2015 to 2018, with notable year-over-year growth, especially in 2017.
* **Key Influencers:** The analysis reveals that shipping mode is the most significant factor in reducing shipping days, while product sub-category (especially Tables and Chairs) has the greatest impact on increasing sales.

**8. Recommendations**

* **Expand in High-Performing Regions**:  
  Focus marketing and logistics efforts on the West and East regions, especially California and New York, to capitalize on their high sales and order volumes.
* **Promote Fast Shipping Options**:  
  Encourage customers to use Same Day and First Class shipping by offering promotions or loyalty points, as these modes significantly reduce delivery times and can improve customer satisfaction.
* **Optimize Product Portfolio**:  
  Maintain strong inventory levels for top-selling sub-categories such as Phones, Chairs, and Tables, and regularly review product performance to phase out underperforming items.
* **Target the Consumer Segment**:  
  Develop targeted marketing campaigns and loyalty programs for the Consumer segment, which accounts for the majority of sales and orders.
* **Leverage Data-Driven Personalization**:  
  Use customer order history and segmentation to personalize offers, recommend products, and increase cross-selling and upselling opportunities.
* **Enhance Customer Retention**:  
  Identify and reward high-value and repeat customers (e.g., those with the most orders) with exclusive deals, early access to new products, or special services.
* **Improve Standard Shipping Efficiency**:  
  Analyze and streamline Standard Class shipping processes to reduce average shipment days and close the gap with faster shipping modes.
* **Monitor and Respond to Market Trends**:  
  Regularly analyze sales and order trends by year, region, and product to quickly adapt to changing customer preferences and market conditions.
* **Expand Product Diversity**:  
  Explore opportunities to introduce new products in high-growth categories and sub-categories, based on customer demand and market gaps.
* **Strengthen Customer Feedback Loops**:  
  Collect and analyze customer feedback on products and shipping experiences to identify areas for improvement and enhance overall satisfaction.
* **Invest in Technology and Automation**:  
  Implement advanced analytics, inventory management, and automated fulfillment solutions to support business growth and operational efficiency.
* **Foster Collaboration Across Teams**:  
  Encourage collaboration between sales, marketing, logistics, and customer service teams to ensure a seamless customer experience and unified business strategy.