1. **Project Overview**

***Superstore Sales Performance***

* **Objective**: The primary goal of this project is to analyze sales trends, customer segments, and product performance based on a historical dataset of store transactions.
* **Scope of Analysis**:
  + Analyze sales performance across various product categories, regions, and customer segments.
  + Understand delivery performance by analyzing shipping dates and shipping modes.
  + Provide actionable insights for business decision-making based on the dataset.

1. **Data Model Overview**

* **Data Source**: The dataset contains a single table with transactional data from a superstore.
* **Data Model**: A **flat file** model with each row representing a single transaction or order. All relevant details about the order, customer, and product are recorded in this table.

1. **Data Dictionary & Data Types**

The dataset has several columns, and each column’s data type plays a role in how it is used for analysis. Below is a detailed list of each column, its description, and its data type.

| **Column Name** | **Description** | **Data Type** |
| --- | --- | --- |
| Order ID | Unique identifier for each transaction or order. | **String** |
| Order Date | Date when the order was placed. | **Date** |
| Ship Date | Date when the order was shipped. | **Date** |
| Ship Mode | Mode of shipment used for the order (e.g., Standard, Second Class). | **String** |
| Customer ID | Unique identifier for the customer placing the order. | **String** |
| Customer Name | Full name of the customer. | **String** |
| Segment | Customer segment classification (e.g., Consumer, Corporate, Home Office). | **String** |
| Country | Country where the order was placed. Typically only one value (e.g., USA). | **String** |
| City | City associated with the customer’s shipping address. | **String** |
| State | State associated with the customer’s shipping address. | **String** |
| Postal Code | Postal or ZIP code of the shipping address. | **String** |
| Region | Geographic region (e.g., East, West, Central) where the customer is located. | **String** |
| Product ID | Unique identifier for the product purchased in the order. | **String** |
| Category | Product category (e.g., Furniture, Office Supplies). | **String** |
| Sub-Category | Sub-category within the product category (e.g., Chairs, Binders). | **String** |
| Product Name | Name of the product purchased in the order. | **String** |
| Sales | Total sales amount for the product in the order. | **Decimal** |

1. **Data Cleaning and Preparation**
2. **Handling Missing Data**

* **Issue:** The *postal code* (ZIP code) for Burlington, Vermont was missing.
* **Action Taken:** We filled in the missing postal code with 05401, which is the correct ZIP code for Burlington.

1. **Data Type Correction**

* **Issue:** The *Order Date* and *Ship Date* columns were not in the correct format.
* **Action Taken:** We converted both *Order Date* and *Ship Date* columns to Date format in Power Query to enable accurate date-based analysis.

1. **Duplicate Data**

* **Action Taken:** We checked for duplicates by examining all rows and the *Row ID* column. No duplicates were found in the dataset.

1. **Cleaning the Product Name Column**

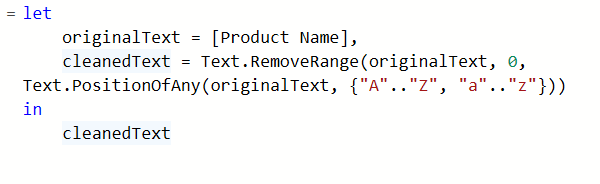
The product names contained unnecessary details, such as descriptions and numbers. We cleaned the column in the following steps:

**Step 1: Remove Unnecessary Descriptions**

* **Action:** After Trimming and Cleaning we split the Product Name column by the **comma** (,) delimiter, keeping only the main product name and removing the descriptive parts (e.g., size or dimensions). This was done by splitting at the **right-most** delimiter twice, which removed excess information after the product name.

**Step 2: Remove Numbers at the Beginning of Product Names**

* **Action:** We added a custom column to remove numbers and unnecessary characters from the beginning of the *product name*. The following M-code was used:



**Explanation:** This code removes any characters (such as numbers or symbols) before the first letter of the product name.

**Step 3: Remove Numbers at the End of Product Names**

* **Action:** We added another custom column to remove any numbers or unnecessary characters from the **end** of the product name. The following M-code was used:

A computer code with text

Description automatically generated with medium confidence

**Explanation:** We reversed the text, removed characters from the beginning (which were originally at the end), and then reversed it back to restore the original order without trailing numbers or symbols.

### ****Creating New Columns****

To enhance the dataset for further analysis, we added new calculated columns.

#### **Step 1: Shipping Time**

* **Action:** We calculated the shipping time by subtracting the Order Date from the Ship Date.



* **Explanation:** This calculates the number of days taken to ship the product, which can be used to analyze delivery performance.

**Step 2: Product Brand**

* **Action:** We extracted the first word by Splitting the cleaned Product Name to create a new column representing the product brand.

Product Brand = Text.BeforeDelimiter([Product Name], " ")

1. **DAX Measures**
2. **Total Sales**  
   TotalSales = SUM('Superstore Sales Dataset'[Sales])  
   **Brief**: This measure calculates the total sales by summing up all values in the Sales column of the dataset. It provides the overall revenue generated.
3. **Total Orders**  
   TotalOrders = COUNT('Superstore Sales Dataset'[Order ID])  
   **Brief**: This measure counts the total number of orders by counting the number of Order IDs in the dataset. It helps track the volume of orders placed.
4. **Average Order Value**  
   AverageOrderValue = AVERAGEX('Superstore Sales Dataset','Superstore Sales Dataset'[Sales])  
   **Brief**: This measure calculates the average value of each order by taking the average of the Sales column. It gives insight into how much customers typically spend per order.
5. **Unique Customers**  
   UniqueCustomers = DISTINCTCOUNT('Superstore Sales Dataset'[Customer ID])  
   **Brief**: This measure counts the number of distinct customers by counting the unique Customer IDs in the dataset. It helps determine the customer base size.
6. **Customer Lifetime Value (CLV)**  
   CustomerLifetimeValue = DIVIDE([TotalSales], [UniqueCustomers])  
   **Brief**: This measure divides total sales by the number of unique customers to determine the average lifetime value of each customer. It helps measure the long-term value of each customer.
7. **Top Customer**  
   TopCustomer = TOPN(1, VALUES('Superstore Sales Dataset'[Customer Name]), CALCULATE(SUM('Superstore Sales Dataset'[Sales])), DESC)  
   **Brief**: This measure identifies the top customer by sales. It retrieves the customer’s name that generated the highest revenue, helping to identify key accounts.
8. **Total Distinct Orders**  
   TotalDistinctOrders = DISTINCTCOUNT('Superstore Sales Dataset'[Order ID])  
   **Brief**: This measure counts the number of distinct order IDs, ensuring that only unique orders are considered. It can be useful in cases where duplicate orders might skew results.
9. **Cities**  
   Cities = DISTINCTCOUNT('Superstore Sales Dataset'[City])  
   **Brief**: This measure counts the distinct number of cities in the dataset, providing a way to track the geographical distribution of orders.
10. **Average Shipping Time**  
    AverageShippingTime = AVERAGEX('Superstore Sales Dataset', 'Superstore Sales Dataset'[Shipping Time])  
    **Brief**: This measure calculates the average shipping time by averaging the values in the Shipping Time column. It helps monitor the efficiency of the shipping process.
11. **Most Used Ship Mode**  
    MostUsedShipMode = TOPN(1, VALUES('Superstore Sales Dataset'[Ship Mode]), CALCULATE(COUNT('Superstore Sales Dataset'[Order ID])), DESC)  
    **Brief**: This measure identifies the most frequently used shipping method by counting the number of orders per shipping mode. It’s useful for understanding which delivery methods are most popular.
12. **Data Visualization**

The dashboard presents a range of visualizations to highlight key metrics and trends across sales, customer insights, and operational efficiency.

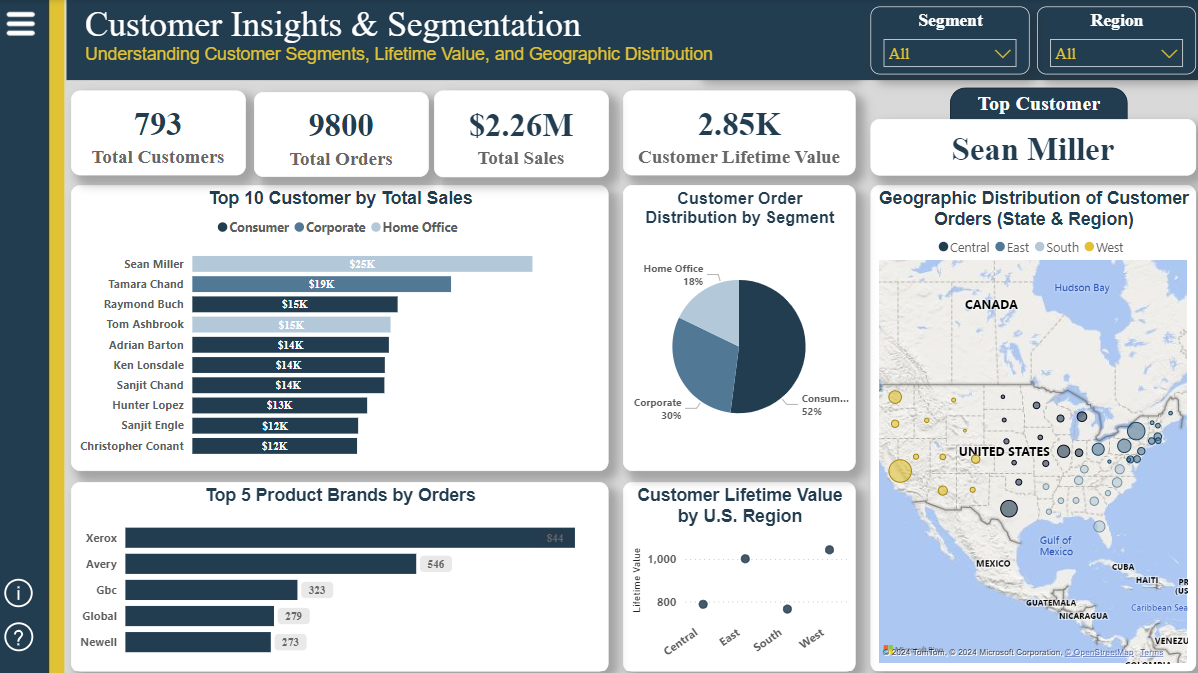
* 1. **Sales Insights & Trends (Page 1)**
* **KPIs**: Total Sales, Total Orders, Average Order Value
* **Sales Trend Over Time**: A line chart illustrating yearly sales trends, showing notable growth in the last two years.
* **Technology Product Sales**: A bar chart displaying sales performance by subcategory within the Technology category, with Phones being the highest-selling subcategory.
* **Furniture Product Sales**: A bar chart showcasing sales performance by subcategory within the Furniture category, with Chairs leading in sales.
* **Office Supplies Product Sales**: A bar chart highlighting subcategory performance in the Office Supplies category, with Storage and Binders showing the highest sales.
* **Sales by State (Map)**: A map illustrating the geographic distribution of sales across various states.
* **Sales by Customer Segment**: A pie chart representing the proportion of sales by customer segment, with Consumers contributing the largest share.

A screenshot of a computer screen

Description automatically generated

**5.2 Customer Insights & Segmentation (Page 2)**

* **KPIs**: Total Customers, Total Orders, Total Sales, Customer Lifetime Value
* **Top Customers by Sales**: A bar chart showcasing the top revenue-generating customers, segmented by category, with Sean Miller as the highest contributor.
* **Orders per Customer by Segment**: A pie chart showing the average number of orders per customer in each segment, where Consumers place the most orders.
* **Top 5 Brands by Orders**: A Clustered bar chart highlighting the number of orders per product brand, with Xerox leading the list.
* **Customer Lifetime Value by Region**: A dot plot comparing average customer lifetime value across regions, with the East and West regions showing the highest values.
* **Customer Orders by State and Region (Map)**: A map offering insights into customer orders, segmented by state and region.



**5.3 Shipping & Operational Efficiency (Page 3)**

* **KPIs**: Total Orders, Cities Served, Average Shipping Time
* **Average Delivery Time by Shipping Method**: A column chart illustrating average shipping times across shipping methods, highlighting Standard Class as the most time-consuming.
* **Shipment Distribution by Delivery Method**: A pie chart showing the frequency of shipping modes, with Standard Class being the most used and Same Day the least.
* **Delivery Time by Product Category and Shipping Method**: A clustered bar chart comparing average shipping times by product category and shipping method.
* **Regional Delivery Time Comparison**: A column chart comparing delivery times across regions, with Central having the longest average delivery time (4.07 Days) and East the shortest (3.91 Days).

A screenshot of a computer

Description automatically generated

**5.4 Dashboard Info (Page 4)**  
This page provides a detailed overview of the key insights from the Sales, Customer Insights, and Shipping Efficiency dashboards. It consolidates the information from the previous three pages, offering a summary of performance across all domains for a quick, high-level understanding.

**5.5 Q&A (Page 5)**  
This page includes a Q&A visual, allowing users to query the dataset and receive immediate visual answers based on the data.

In addition, slicers are available for filtering data by **segment, state, region, and city**, offering dynamic filtering across all dashboards to tailor insights based on specific locations or customer segments.

1. **Performance Analyzer**

To ensure optimal performance, the **Performance Analyzer** feature in Power BI was utilized to identify and address any performance bottlenecks across the dashboard. The tool was essential for:

* **Monitoring Query Performance**: By using Performance Analyzer, the load times are tracked for each visual, highlighting the most time-consuming queries. This showed which visuals or calculations needed optimization.
* **Optimizing DAX Formulas**: Several complex DAX formulas were streamlined to reduce their calculation time. For example, calculations involving customer lifetime value and shipping times were optimized to speed up report performance.
* **Reducing Report Render Time**: Performance Analyzer identified visuals that were slow to render. Based on this, data granularity and modified filters are adjusted to reduce load times, ensuring a smoother user experience across all pages.
* **Improving User Interactivity**: By identifying heavy visuals, interactivity was improved, allowing faster response times when using slicers like **segment, state, region, and city**. This ensured the dashboard remained responsive even when large datasets were filtered.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

1. **Challenges and Solutions**

* **Challenge 1: Data Inconsistencies Across Sources**  
  **Solution**: Power Query was used to cleanse and standardize data, ensuring consistency. Missing data was imputed where necessary to maintain accuracy.
* **Challenge 2: Sales Outliers**  
  **Solution**: Sales outliers were flagged and analyzed to ensure only valid records were included in the final analysis, improving overall data reliability.

1. **Conclusion**

**Key Insights:**

 **Sales Performance**: Total sales reached 2.26M $, with an average order value of 230.77$. Identifying top-performing product categories and brands provided a clear view of revenue drivers.

 **Customer Insights**: The analysis revealed 793 unique customers, with repeat buyers making significant contributions. The top customer accounted for 10% of total sales, highlighting opportunities for increasing loyalty among high-value customers.

 **Shipping Efficiency**: The average shipping time was analyzed across different shipping methods and regions. Standard Class showed the longest shipping time, while the same-day delivery was used the least. Central region had the highest average shipping time, while the East had the lowest. These insights offer opportunities to improve delivery speed in key regions and optimize shipping methods for efficiency.

The comprehensive dashboard offers valuable insights across sales, customer behavior, and operational efficiency, equipping decision-makers with the tools needed to drive performance, optimize processes, and enhance customer engagement strategies.

1. **Appendix**

* **Tools Used**: Power BI, Power Query, Excel
* **Data Sources**: CRM system, Order management system
* **Key Metrics**: Total Sales, Average Order Value, Customer Lifetime Value, Average Shipping Time