Company X Data Department Report

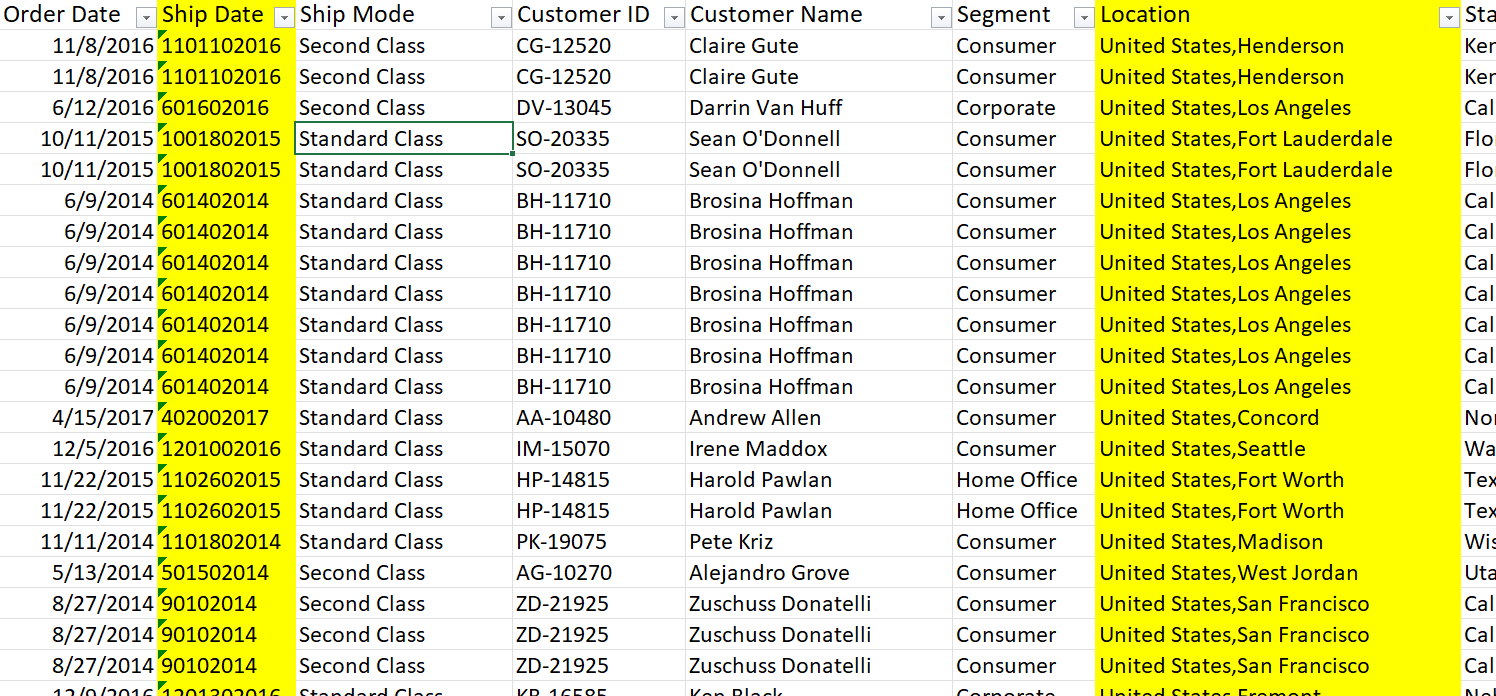
# Executive Summary

I exported the sales data from our database, and I conducted analysis ono it to gain insight into some of our sales operations and study ways of enhancing our services based on certain scopes.

## Data Exploring:

Upon reviewing the data, I observed the following:

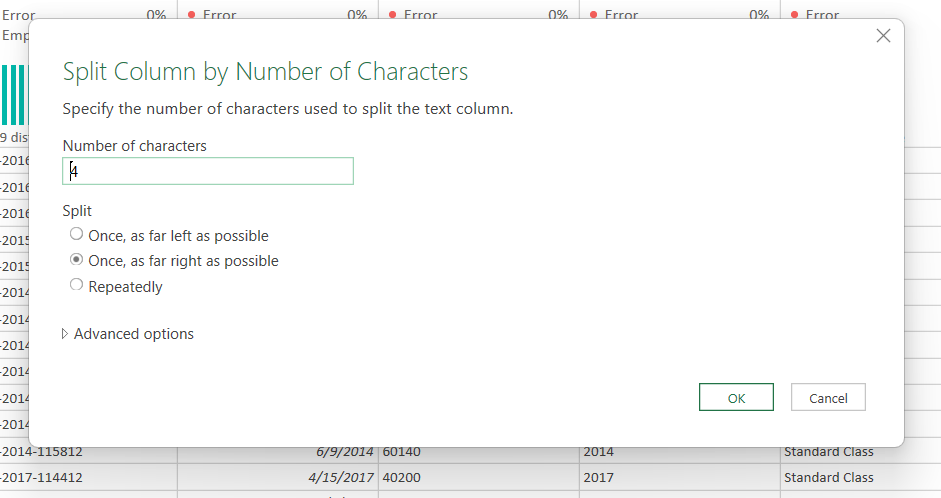
1. The **Location** column contains both the country and city values in the same field. This column needs to be split into two separate columns: **Country** and **City**.
2. The **Ship Date** column requires correction to ensure accurate data representation.



Data Cleaning Process:  
**1. Correcting the 'Ship Date' Column**

To correct the 'Ship Date' column, I used a character-based splitting approach:

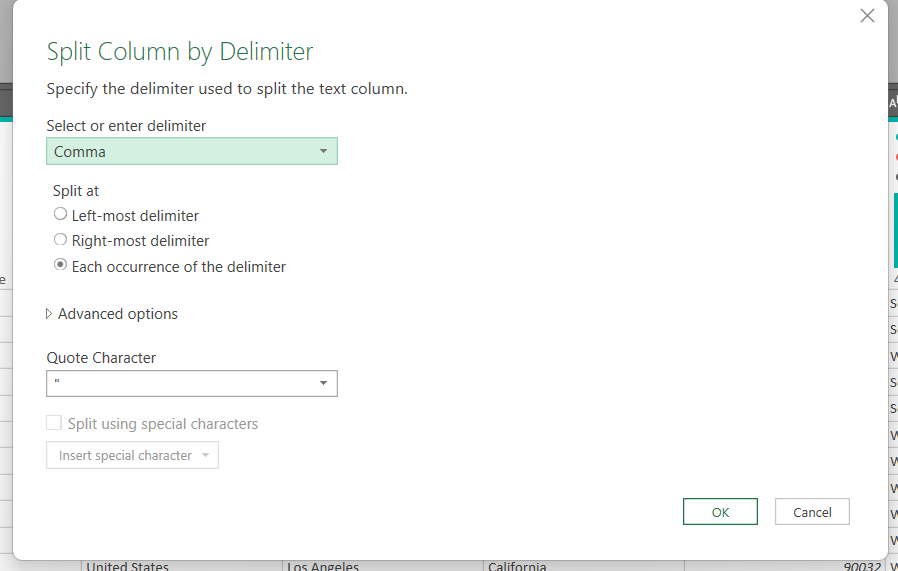
* **Year Extraction:** Split the column from the right into four characters, creating an independent column named **'Year'**.
* **Day Extraction:** Split the remaining values from the right into three characters, then further split one character from the right to isolate the day values in a separate column named **'Day'**. Columns containing only zeros were removed.
* **Month Extraction:** Split the column by one character from the right. Upon review, I found both zeros and numbers in the column. To resolve this, I separated the zeros into a different column, removed it, and merged the remaining values with the month column, renaming it **'Month'**.
* **Reconstructing Ship Date:** Finally, I combined the **'Year'**, **'Month'**, and **'Day'** columns to reconstruct the corrected **'Ship Date'** column.



**2. Splitting the 'Location' Column**

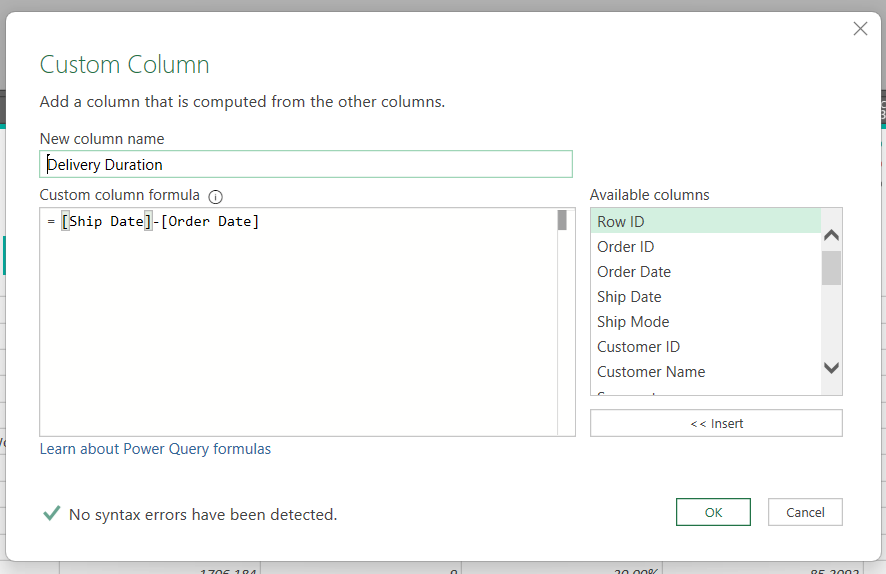
I split the **'Location'** column by the comma delimiter, creating two separate columns:

* **Country**
* **City**



**3. Calculating Delivery Duration**

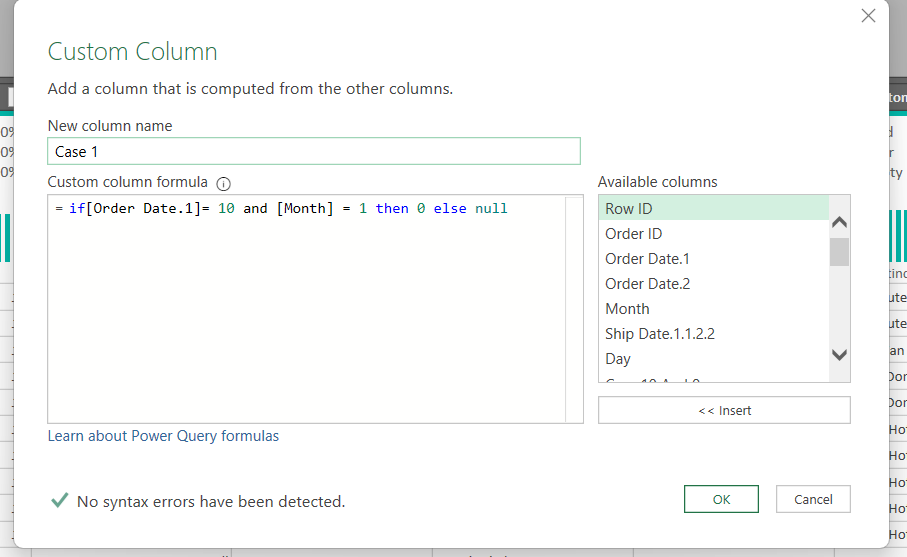
A custom column labeled **'Delivery Duration'** was created by calculating the difference between **'Ship Date'** and **'Order Date'**.

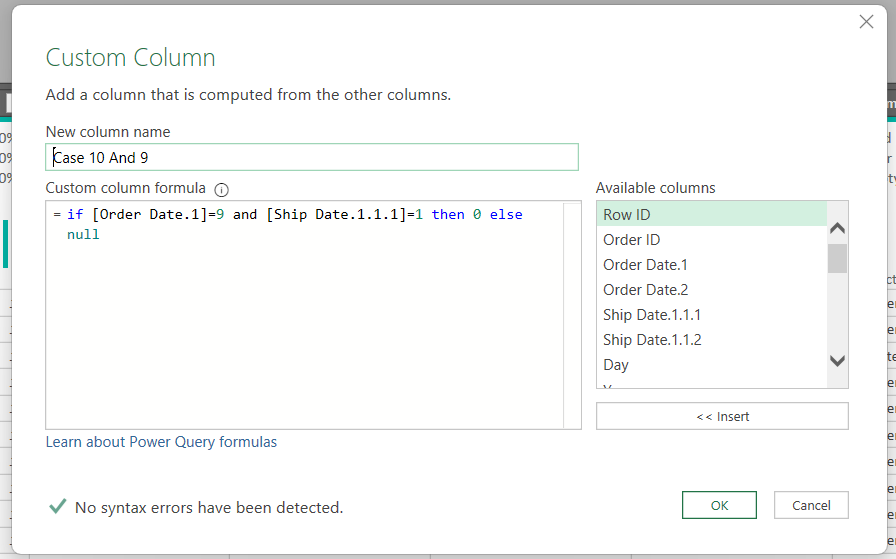


**4. Fixing the 'Ship Date' Calculation Issue**

After adding the **'Delivery Duration'** column, I noticed incorrect and negative values. Upon analysis, the issue was traced back to errors in the months **9 (September)** and **10 (October)**.  
To resolve this, I added two custom columns named:

* **Case1**
* **Case10and9**

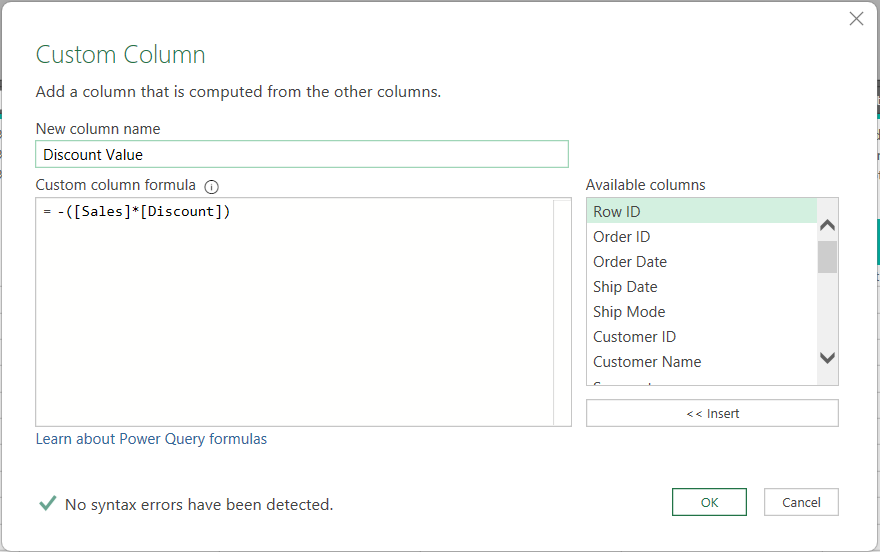




These columns were used to correct the date inconsistencies.

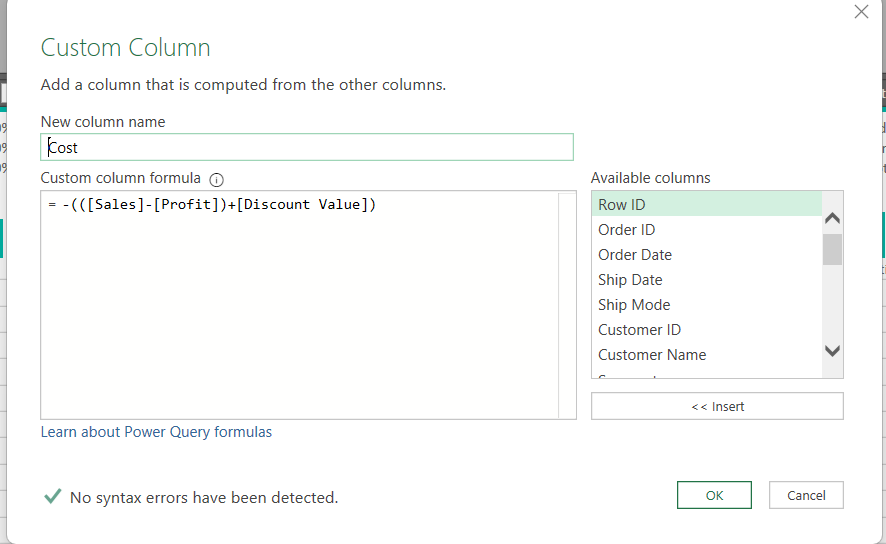
**5. Calculating Discount Value**

A custom column labeled **'Discount Value'** was created using the formula:  
**Discount Value** = – (Sales \* Discount)



**6. Calculating Cost**

A custom column labeled **'Cost'** was created to represent the cost of each order using the formula:  
**Cost** = -([Sales]-[Profit])+[Discount Value])



## Scope

* Customer Loyalty
* Strengths and Weak Points
* Performance
* Customer Experience

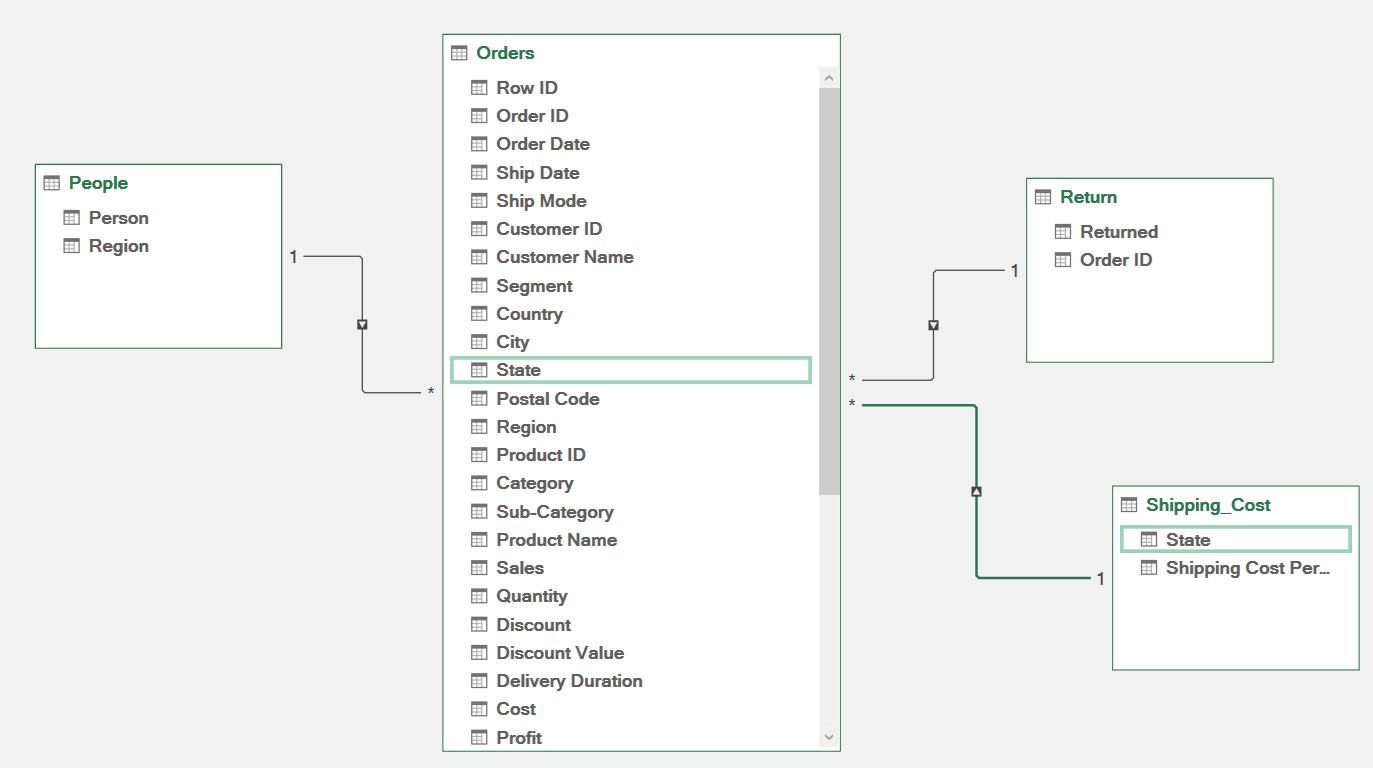
Data Specs:

Here are the tables we used to conduct our analysis.

* Orders
* Return
* Shipping Cost
* People

## Data Modeling:

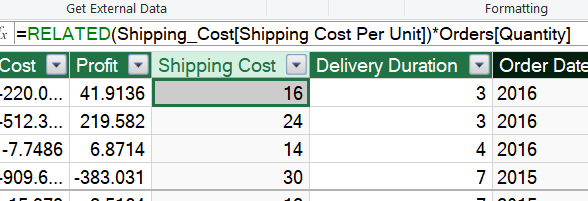
Using the Power Pivot tool, I performed data modeling and established relationships between the tables to enhance data analysis.



## Creating New Calculations and Measures:

**Shipping Cost Column**

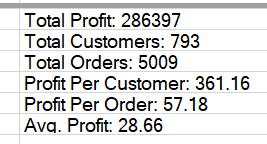
After building the data model, I used DAX formulas to create a column that calculates the shipping cost for each state, based on the quantity for that state.



**Measures Created Using DAX**

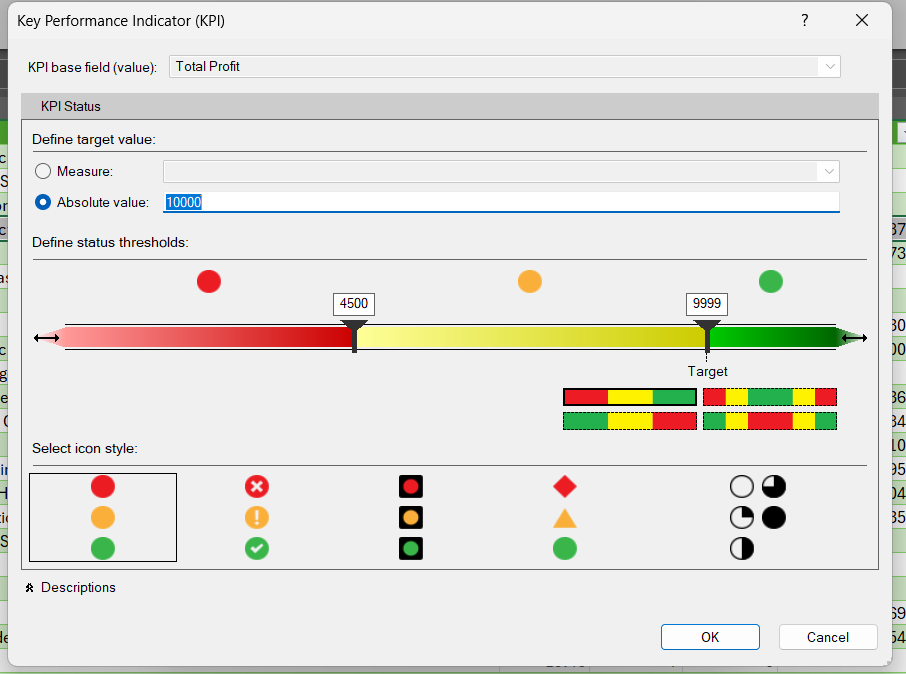
I created the following measures:

* **Total Profit:** SUM(Orders[Profit]) to calculate the total profit for all orders.
* **Total Customers:** DISTINCTCOUNT(Orders[Customer ID]) to calculate the total number of unique customers.
* **Total Orders:** DISTINCTCOUNT(Orders[Order ID]) to calculate the total number of unique orders.
* **Profit Per Customer:** Total Profit ÷ Total Customers to calculate the profit generated per customer.
* **Profit Per Order:** Total Profit ÷ Total Orders to calculate the profit generated per order.
* **Average Profit:** AVERAGE(Orders[Profit]) to calculate the average profit per order.



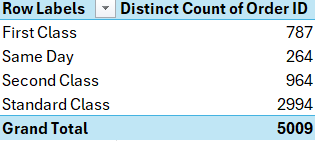
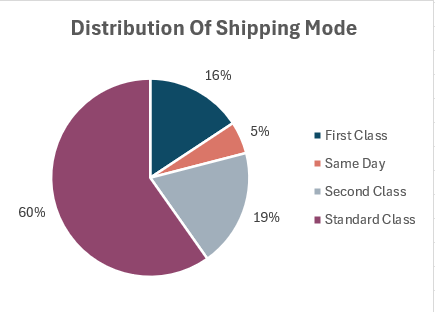
**Creating KPIs (Key Performance Indicators)**

I created a KPI for **Total Profit**, setting a target value of **10,000** to monitor performance.



# The Analysis

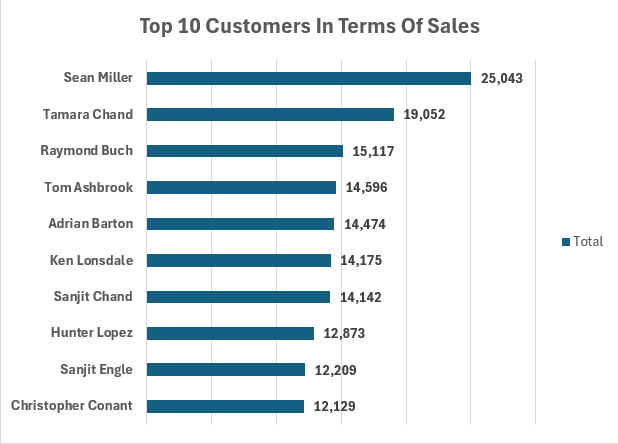
## Customer Loyalty

*  What is the most ship mode used by our customers?

Based on the analysis, the most used shipping mode is **Standard Class**, with 2,994 orders (60% of total orders). The least used mode is **Same Day**, with 264 orders (5% of total orders).

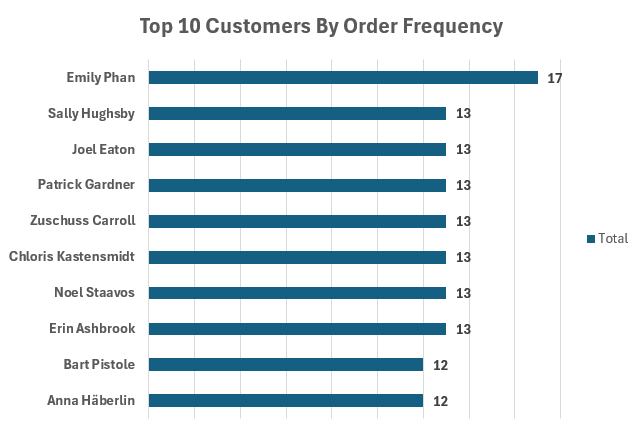
* Who are our top 10 Customers In Terms Of Sales?

A chart highlights the top 10 customers by sales value.



* Who are our top 10 Customers in terms of Order Frequency?

The analysis shows that **Emily Phan** has the highest order frequency with 17 orders.



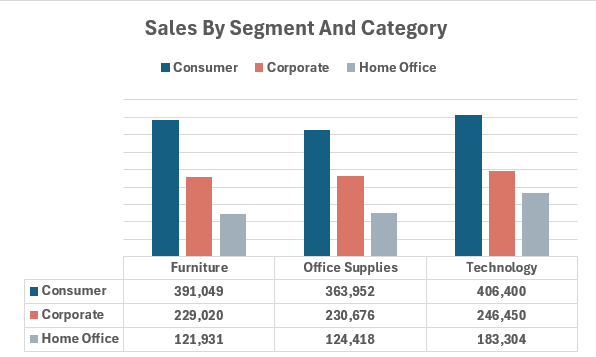
### Recommendations:

Based on our analysis above we suggest that:

* Review the pricing for **Same Day** and **First-Class** shipping modes to make them more competitive.
* Offer special incentives, such as free shipping, to the top 10 customers to enhance loyalty.

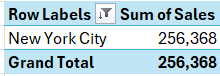
## To monitor our strengths and weak points:

* Which Segment of Clients Generates the Most Sales by Category?



The **Technology** category and **Consumer** segment generated the highest sales, totaling **406,400**.

* Which City Has The Most Sales Value?



**New York City** has the highest sales value.

* Which State Generates the Most Sales Value contribution By Region?

**California**, in the West region, contributes the most to sales, accounting for **23.3%**.

### Recommendations:

* The **Home Office** segment has the lowest sales. Consider introducing targeted offers to increase sales in this segment.

## Performance Measurement:

* What is the most profitable product that we sell?

The most profitable product is **Copiers**, which falls under the **Technology** category.

## Customer Experience:

* On Average how long does it take the order to reach to our customers?

## Sales Dashboard:

Finally, I created a sales dashboard and made it dynamic by using slicers.

