**FURNITURE SALES ANALYSIS**

Introduction

The dataset was a furniture inventory sales CSV file and was downloaded from [Kaggle](https://www.kaggle.com/datasets/zahraaalaatageldein/sales-for-furniture-store). I wanted to understand the breakdown of sales both by value and quantity. There were multiple columns, and this presented an opportunity to analyze the data from several angles and identify patterns in the data.

Objectives

1. **Breakdown Sales**. By understanding how their sales are broken down, the organization can understand their market better. This is key to the marketing department understanding where they might want to focus their efforts on.
2. **Identify Sales Patterns.** Understanding the sales trends over the years by state, segmentation etc. to understand historical data and plan.
3. **Strategic Expansion.** By understanding sales trends by state, category and region, management can make informed decisions on where to expand based on historical revenue.
4. **Production Resource Allocation**. To be efficient in production, the organization should understand the most in demand products by customer segmentation to understand where to invest more resources.
5. **Risk management.** Analyzing which periods of the year demand is low for which state/ region/ category, management can plan to slow down production to avoid over-supply.
6. **Market Segmentation.** Analyzing profitability of regions/ categories/ state/ cities, the organization can segment their market to know where the high & low margin customer categories lie.

Process

1. Data Profiling

The dataset I used had **2121** rows and **21** columns. The dataset didn’t have any missing values therefore, it was easy to work with. The RowID column was not sequential.

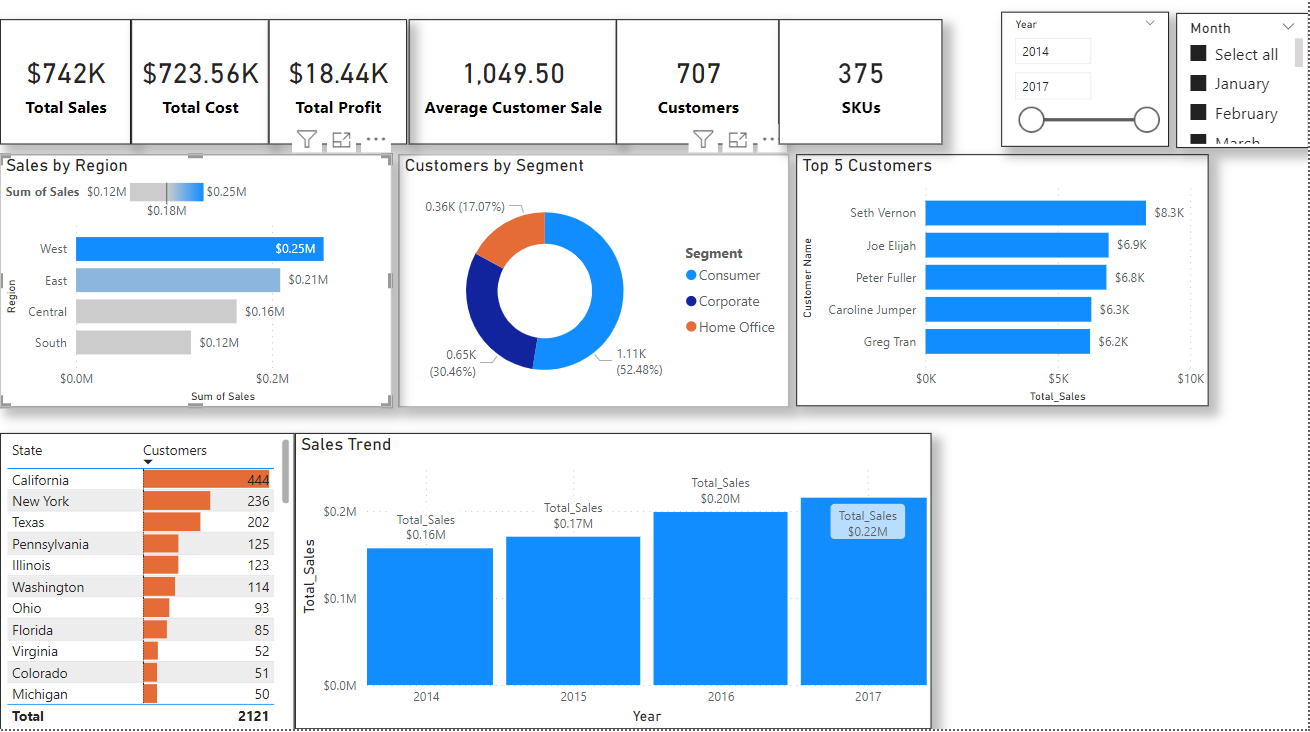
1. Data Cleaning
2. Row-ID column had no sequential numbering – formatted the column to add sequential unique numbers & formatted as whole number.
3. Sales & Profit columns formatted as decimal type for consistent data.
4. Order & Ship date columns formatted as **dd-mm-yyyy** date format.
5. Discount columns formatted as a percentage.
6. Postal column formatted as whole number.
7. Quantity column formatted as whole number.
8. Data Transformation
9. Cost column addition. Sales less Profit.
10. Unit Price addition. Sales divided by Quantity.
11. Unit Cost addition. Total Cost divided by Quantity.
12. Unit Profit. Total Profit divided by quantity.
13. Has\_Discount to identify orders with discounts.
14. Data Analysis

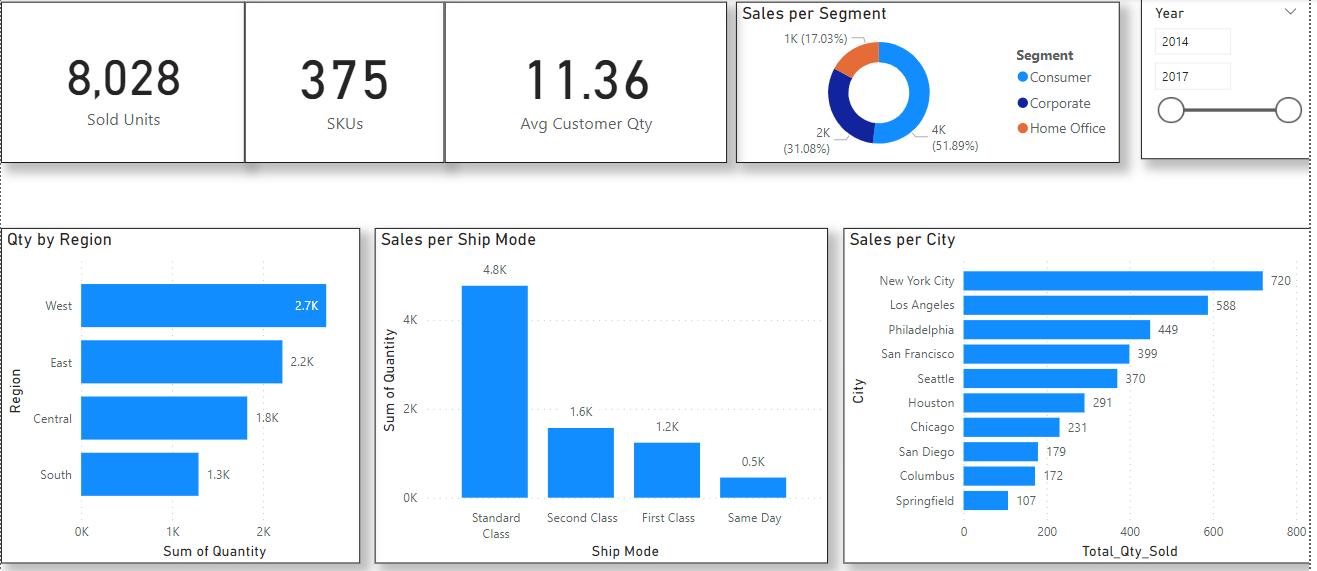
The following were the analysis points:

* ~~Total Sales Value~~
* ~~Average Sales Value~~
* ~~Average Customer Sales Value~~
* ~~Total Sales Qty~~
* ~~Total SKUs~~
* ~~Total Customers~~
* Average Sales per State
* Average Profit per Segment
* ~~Customer Concentration per City~~
* ~~Sales Distribution per Region~~
* ~~Sales per State~~
* ~~Sales per Sub-Category~~
* ~~Sales vs. Profit per Region~~
* ~~Distribution of Shipping Mode per State~~
* Sub-category distribution per State – to understand product demand per state.

1. Data Visualization

The following were the first draft visualizations.





Insights

1. Overall, in all years, the number of orders cumulatively increases as the year progresses. Most orders made increase starting to August.
2. Increase marketing campaigns to drive up sales in the South Region as it was the lowest.