**TABLE OF CONTENTS**

**Title**                                                                                   **Page No**

## 1.0. Introduction

This report covers the visualisation techniques adopted for the analysis of a publicly available Superstore dataset. The anonymous Superstore is a perceived market leader in the retail distribution industry in the United States (US). This Superstore seeks to enhance its sustainable competitive advantage by understanding business areas that need to be developed for increased sales, demand, and profit. Also, the business is interested in identifying operations, products, and locations that contribute significantly to financial loss and decreased sales. The developed visualisations aim to reflect the business performance, enabling key drivers draw insights, and make informed business decisions.

### 1.1. Dataset Information

The dataset used for this visualisation contains 21 columns and 9994 rows, carrying information of sales, profit, product-category etc., for an anonymous superstore over a 3-year period (2014-2017). Details on the dataset can be found in the appendix of this report.

## 2.0. Data Pre-Processing

Microsoft Power Bi was used as the visualization tool in this report. This tool was chosen because of its simplicity and wide range of functionalities. The following steps were further taken to prepare the dataset.

1) The raw data was imported into Microsoft Power Bi using the *'Import data from Excel'* button shown in figure 1. This feature was used because the dataset was stored in an excel spreadsheet.
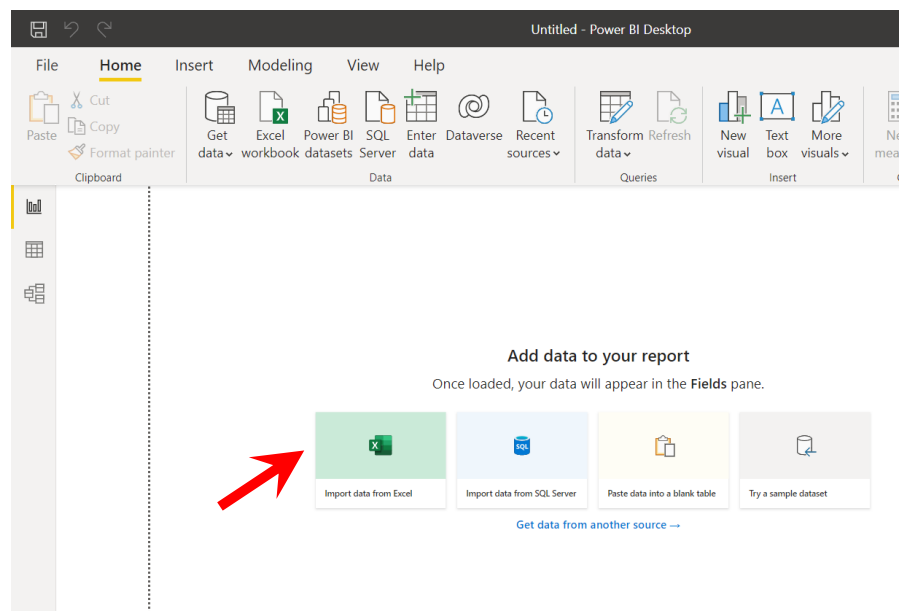


**Figure 1**

2) Upon analysis of the imported data, it was realized that certain formats needed to be adjusted, such as discount, profit etc. The transform button in figure 2 was used to organize the rows and columns in the dataset and discard irrelevant information.
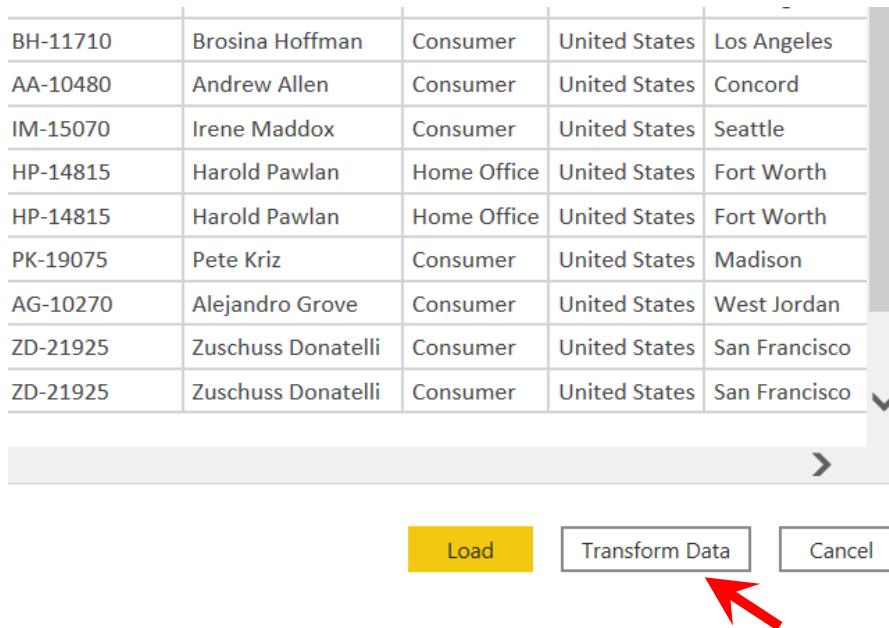
**Figure 2**

3) Before the row and column formats were adjusted, the contents of each column were viewed by ticking the column quality and column distribution boxes under the view panel in figure 3. This was done to ensure all cell components were valid, error free and not null.
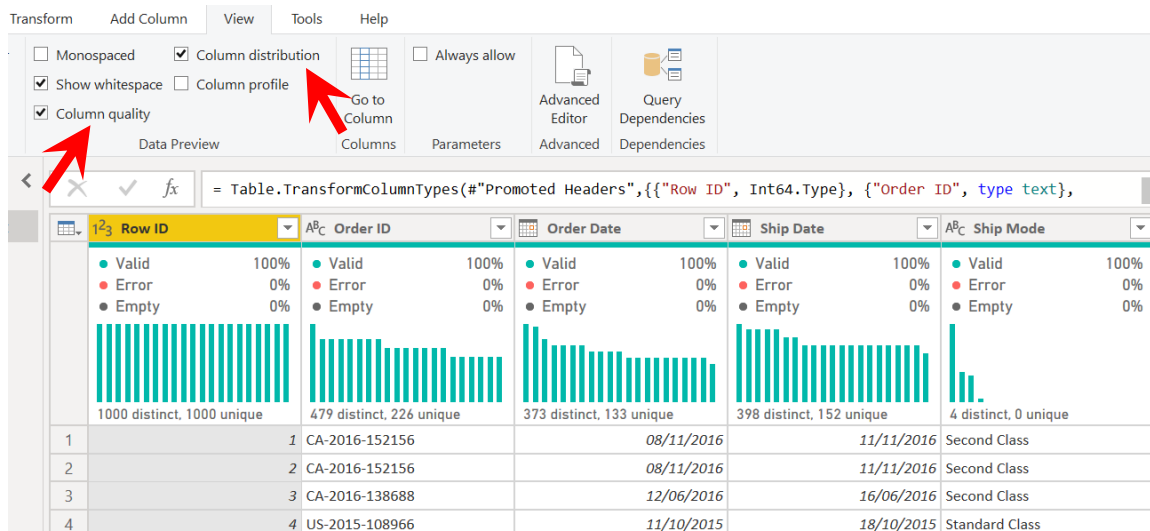


**Figure 3**

4) Upon confirmation of the validity of all cell contents, the format for the sales and profit columns were changed from decimal to fixed decimal for consistent rounding and reporting of all figures. The discount column was also changed from decimal to percentage for proper percentile representation as seen in figure 4.

**Figure 4**

The default format for other columns were acceptable and therefore remained unchanged. All transformation operations were applied, and the data was ready for visualization.

## 3.0. Data Visualisation

The first step in the visualisation process was to understand the variables in the dataset and create a relationship between these variables for better insights. Seven dashboards were created to achieve the business objective.

a) The first dashboard was designed to visualize the overall sales performance of the Superstore by seasons, across all products and regions. Four charts were created in this dashboard and drilled down for expansion of the operational period. A slicer was also created to visualize the performance of specific dates in the business year.
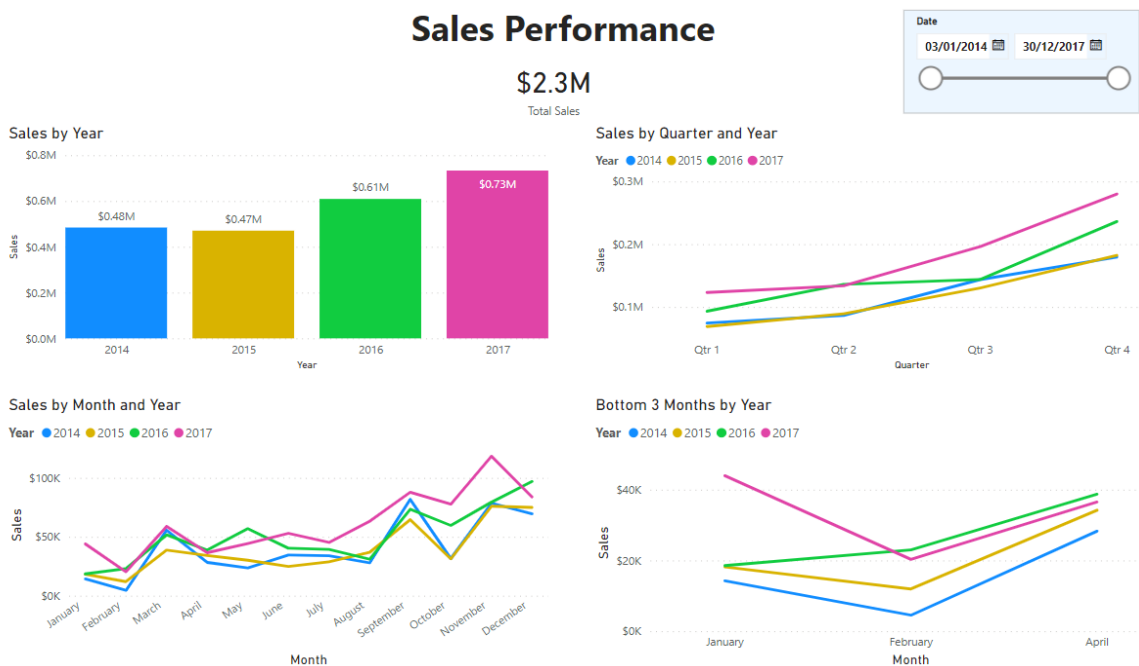


**Figure 5**

The first chart was created using a stacked column chart to show the yearly sales performance. This visualization technique was used to capture changes over time and is more suitable because the periods being compared are few (2014 – 2017). However, the chart showed a clear progression of sales year on year. This visual was created by selecting the stacked column chart in figure 6 and then the sales and order date columns were imported into the empty chart.
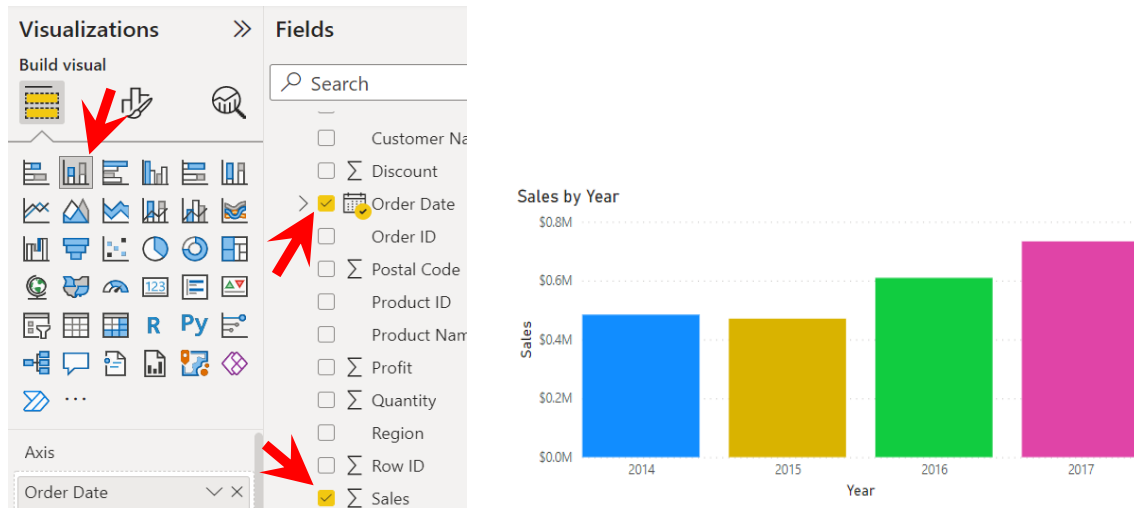


**Figure 6**

The yearly sales performance was further drilled down into a quarterly representation. The line chart visualisation technique was used to concisely show the time intervals and sale changes in each quarter. Sales was observed to have gradually increased overtime, shown in figure 7.
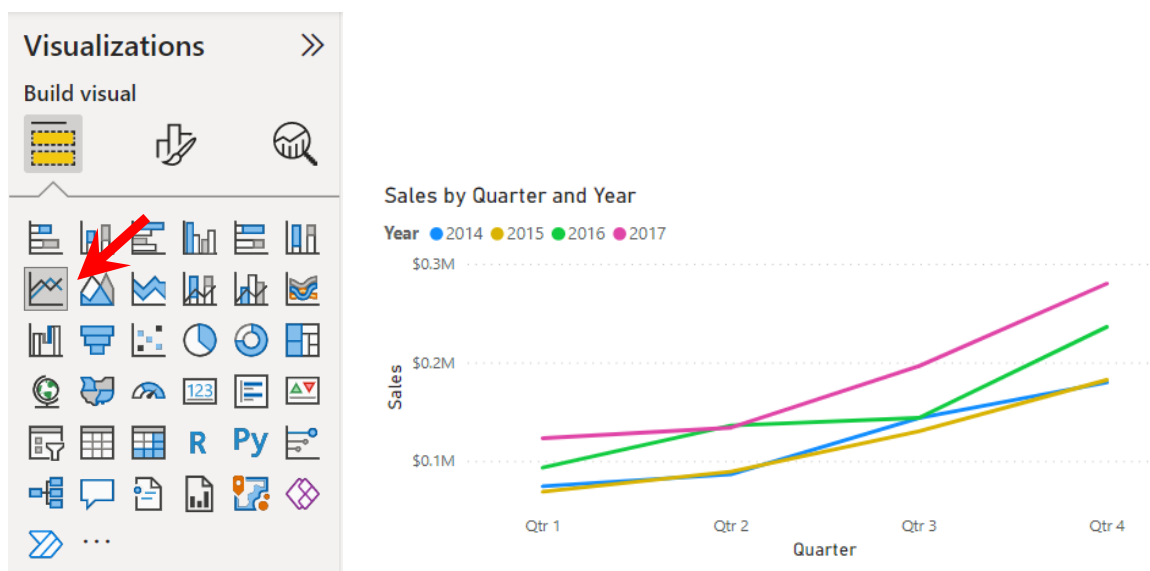


**Figure 7**

The quarterly chart was further drilled down to show the monthly sales performance. An additional line chart was then created to show months with the least sales performance across four years.
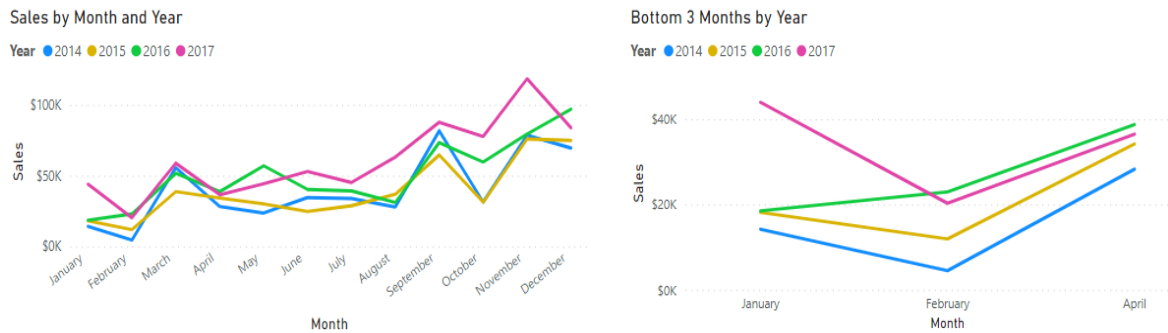


**Figure 8**

As shown in figure 9, a relationship is created connecting all charts. For instance, by selecting 2017 in the diagram, all related information is displayed.
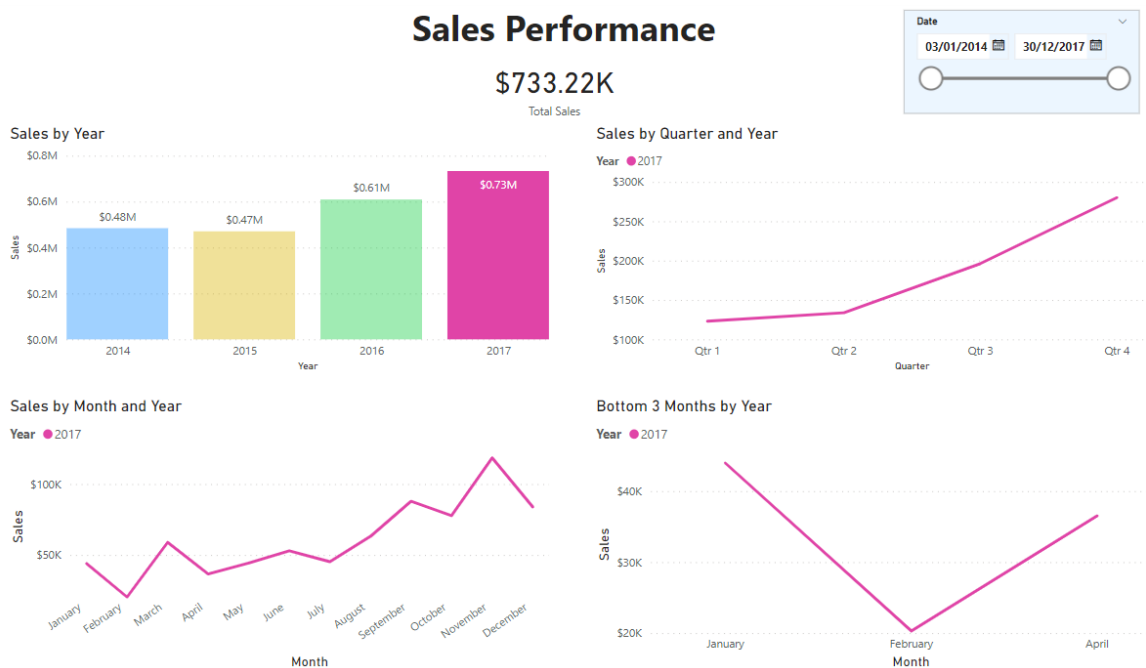


**Figure 9**

b) The second dashboard was designed to show the product categories that contributed mostly to sales and profit. A new measure was first created using cards, to show the total number of products sold and total sales and profit made. The stacked bar chart was used because it clearly shows the connection between different product sub-categories and business performance. It was observed that phones recorded the highest sales compared to other products. Likewise, financial losses were observed for table, bookcases, and supplies. It was also evident that technology products were the most profitable category while furniture products were the least profitable. A slicer was included to filter the performance by specific product names.
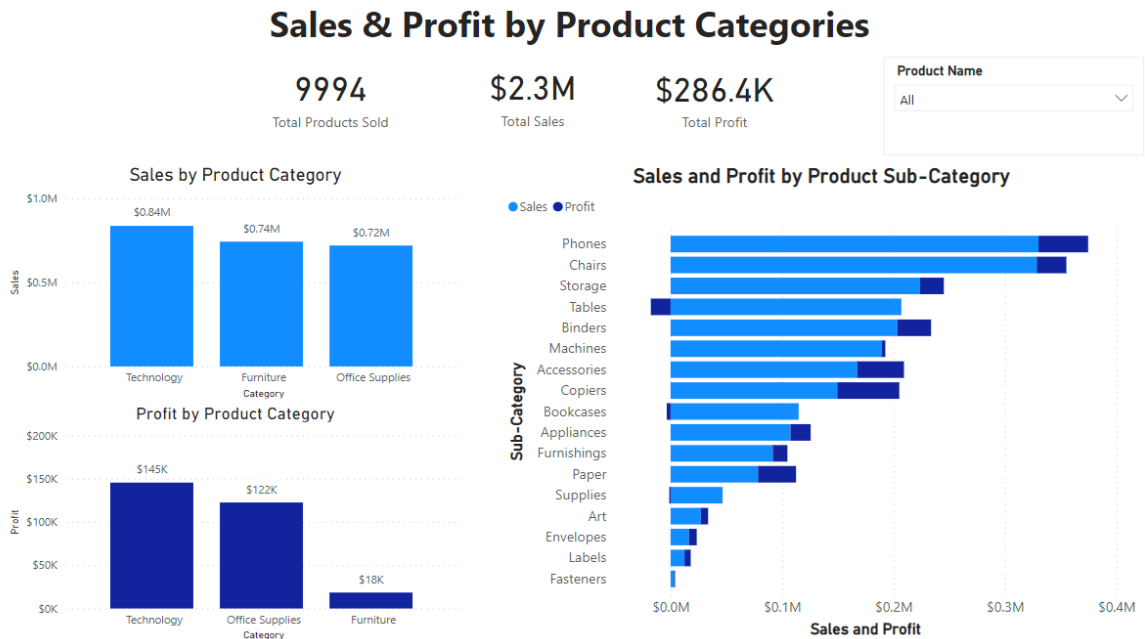
**Sales & Profit by Product Categories**



**Figure 10**

c) The third dashboard concentrates on business performance by location. The clustered bar chart was used to visualize sales and profit by regions in the US. This technique was more suitable to simultaneously view sales and profit performance within the operational regions. The map chart was further used for geographic visualisation to view sales and profit across states within the US. It was observed that the western region yielded the highest sales and profit, most of which was attributed to California. A slicer was created to further filter this performance by cities within each state.
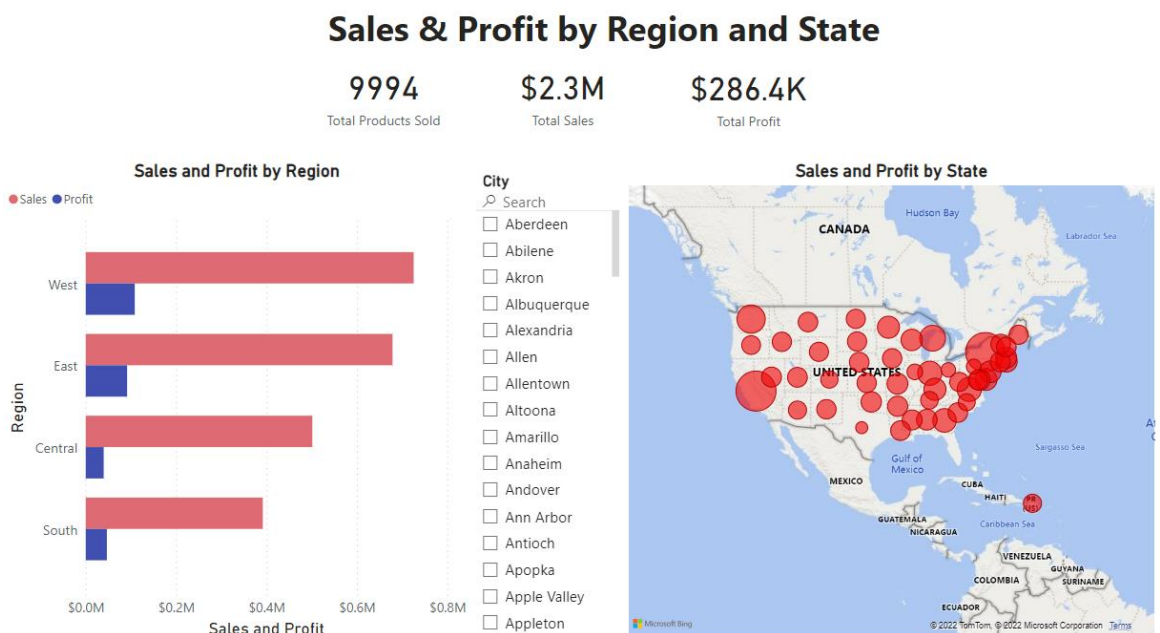
**Sales & Profit by Region and State**



**Figure 11**

d) To understand the pattern of sales, the fourth dashboard was designed using the tree map visualization technique to observe the Superstore's best-selling products.
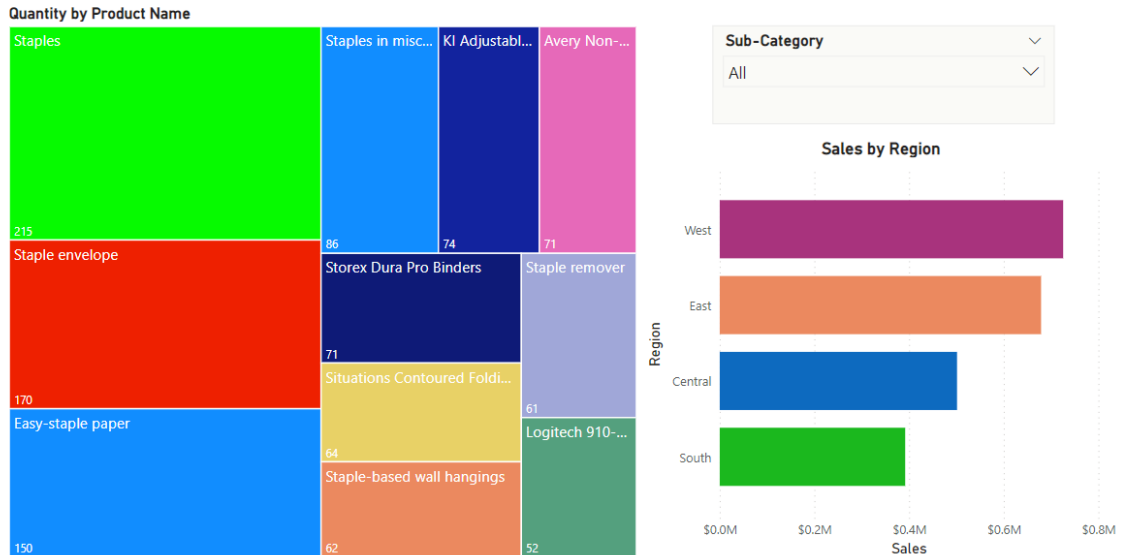


**Figure 12**

This technique was chosen because it is best for visualising values in a structured hierarchical format, making the result more interpretable. The quantity and product names were inserted into the group and value fields to return products with higher purchase frequency. Staples were observed to be the most purchased product, which is a valuable information for the business. Figure 13 shows a drill down on the top 10 products sold by region and by customer segments, using clustered bar charts. Consumers recorded a higher purchase of these products in western US.
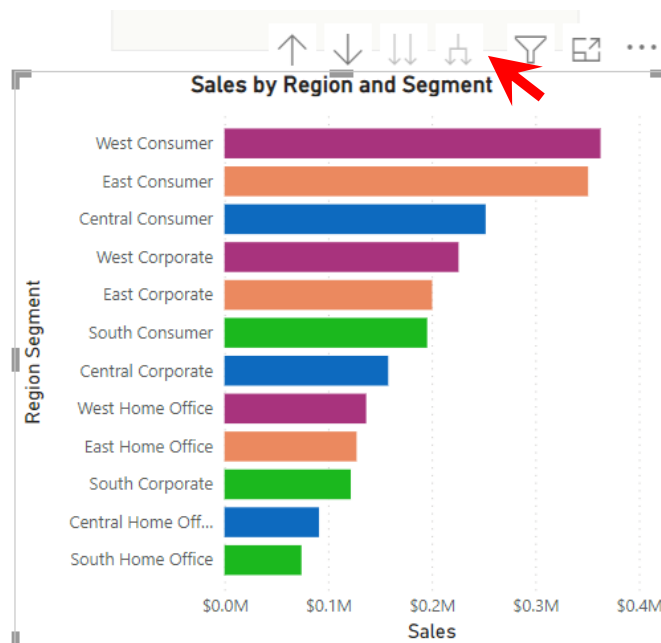


**Figure 13**

e) Customer segment is an important variable, reflecting groups of customers contributing to overall sales and profit. The pie chart was used to represent profit while the donut chart represented sales. These charts were chosen to reflect proportionality of customers in business performance and suitable because time factor is not visualized.
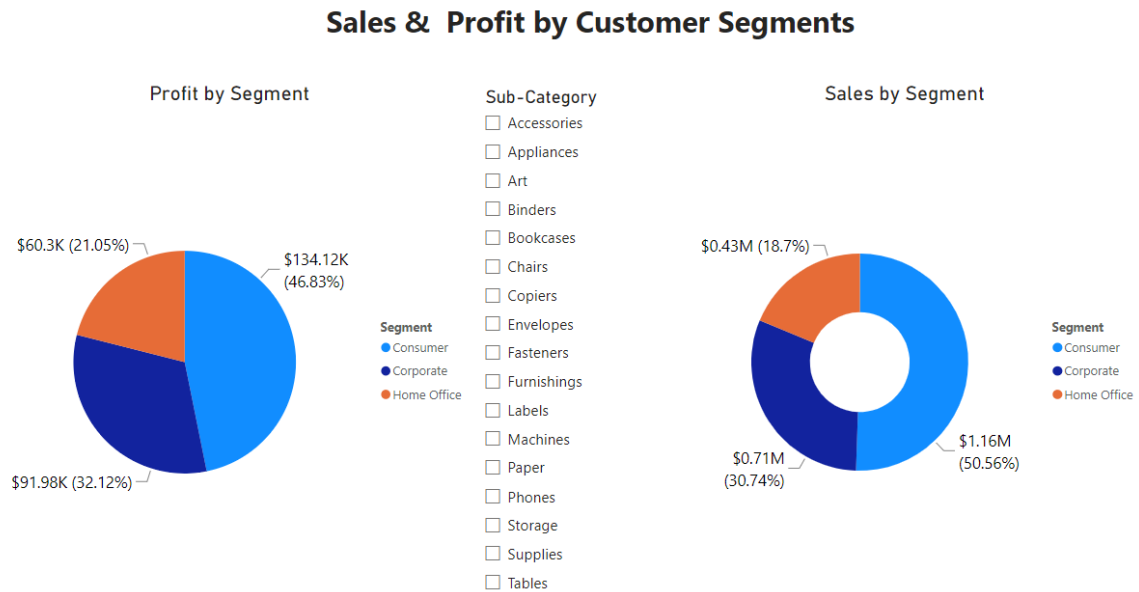


**Figure 14**

These charts were preferred because, a few numbers of categories were involved. A clear indication of consumers contributing more to financial success compared to other customer segments, is shown in figure 14. Product sub-category was also added to the dashboard, to filter these customer groups for better understanding of product/customer relationships.

f) The sixth dashboard was created to understand the relationship between product shipping mode and sales. The line and stacked column chart were used to show a clear comparison between three measures. It is a more suitable technique because it combines two different charts to give a robust information. This chart shows that more customers opted for the standard class shipping (5-days delivery), and more products were delivered through this shipping mode. The clustered bar chart was further used to compare different shipping modes with customer segments. Consumers mostly adopted the standard shipping mode, compared to other customer types. A slicer was also created to view shipping mode by region.
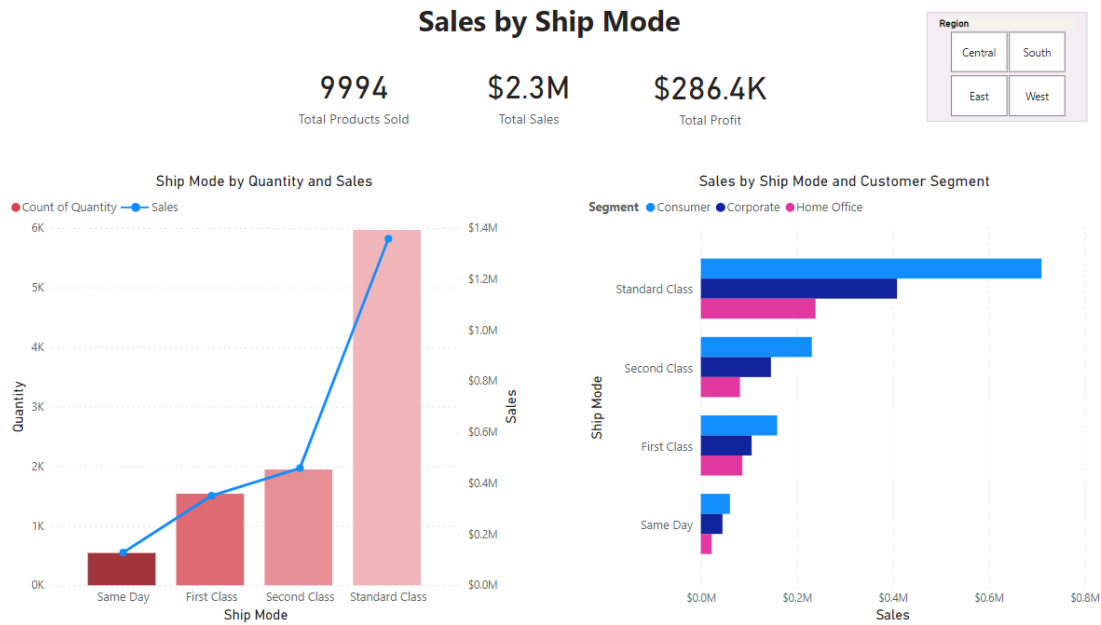
**Figure 15**

g) The last dashboard reflected the effect of discount on sales. The line and stack column charts were used to display the relationship between discount and sales on customer segment and product category, for efficient comparison of values. It was evident that discount had no impact on sales with respect to product category. Office supplies had the highest discount offers and recorded the lowest sales while technology products had the least discount offers and still recorded higher sales. Also, it was realized that consumers were offered more discounts compared to other customer segments, which had an impact on sales. Ship mode was added as a slicer to have a clearer view on the impact of discount on shipping.



**Figure 16**

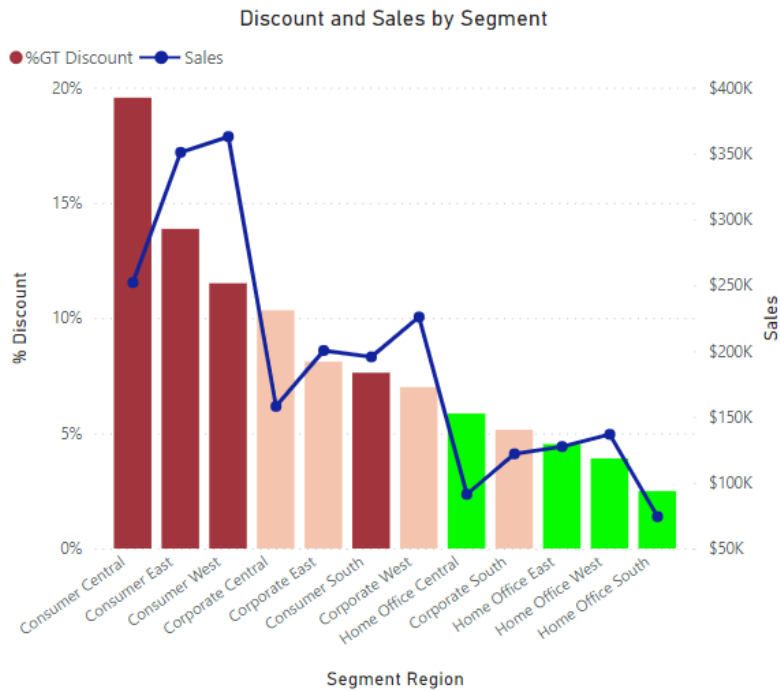Each chart was further drilled down by region and product sub-category, for expanded details as seen in figure 17 and 18.
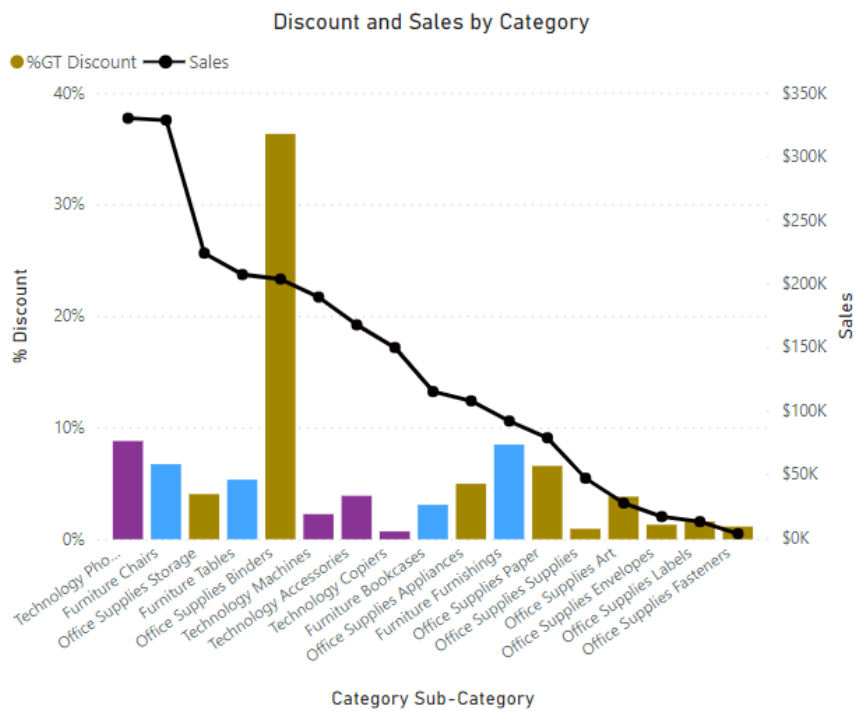


**Figure 17**



**Figure 18**

## 4.0. Conclusion

The designed visualisations were excellent in showing the relationships between features in the dataset. It was realized that sales and profit performance increased gradually from 2014-2017. A larger portion of this performance was attributed to the sales of technology products irrespective of its minimal discount offers. Also, a larger portion of sales and profit, came from California located in the western region of the United States. Office supplies were the most purchased products. Consumers patronized the business more and were offered the most discount offers. All forms of customers opted mostly for 5-day standard shipping, compared to other shipping options. Lastly, it was evident that discount offers had little impact on sales performance. These insights will therefore be of great value to the Superstore, for strategic business decisions.

## Reflections

This dataset had a lot of rows and product names. All order and shipping dates had to be converted to a unified format as different rows had different date formats. It was also a struggle to clearly represent cities in the visualisation.

## 5.0. References

davidiseminger (n.d.). Data types in Power BI Desktop - Power BI. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-data-types#:~:text=Fixed%20Decimal%20Number%20%E2%80%93%20Has%20a [Accessed 13 Apr. 2022].

maggiesMSFT (n.d.). Slicers in Power BI - Power BI. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-slicers.

www.oreilly.com. (n.d.). Use fixed decimal number for precision - Microsoft Power BI Cookbook [Book]. [online] Available at: https://www.oreilly.com/library/view/microsoft-power-bi/9781788290142/e5e174f1-5dac-4293-b477-0c1c0290e522.xhtml.

## 6.0. Appendix

The dataset is publicly available at:
https://www.kaggle.com/datasets/vivek468/superstore-dataset-final?resource=download

| S/N | Column Name | Description |
| --- | --- | --- |
| 1 | Row ID | Serial number for each row |
| 2 | Order ID | Unique ID per order |
| 3 | Order Date | Date of product order |
| 4 | Ship Mode | Shipping types |
| 5 | Ship Date | Product shipping date |
| 6 | Customer ID | Unique ID per customer |
| 7 | Customer Name | Full customer names |
| 8 | Segment | Customer class |
| 9 | Country | Customer location by country |
| 10 | State | Customer location by state |
| 11 | City | Customer location by city |
| 12 | Postal Code | Customers postal code |
| 13 | Region | Customer location by region |
| 14 | Product ID | Unique ID per product |
| 15 | Category | Product category |
| 16 | Sub-Category | Product sub-category |
| 17 | Product Name | Names of ordered products |
| 18 | Sales | Sales of products |
| 19 | Quantity | Product quantity ordered |
| 20 | Discount | Discount available |
| 21 | Profit | Profit/Loss incurred |