Superstore Sales Report

IA640: Information Visualization

Project -1 (Report)

Instructor - Thevasha Sathiyakumar

Rahul Tammalla – 1010517, Clarkson University

Project Overview:

The project aims to create a data visualization that provides insights into the profits and losses across different regions and product categories for the year 2019. The goal is to help stakeholders quickly identify which categories are performing well and which regions are showing potential areas for improvement.

Dataset Overview:

The dataset, SuperStore Sales DataSet, contains detailed sales transactions from a retail store. Key attributes in the dataset include:

- Region: The geographic location of the sale (Central, East, South, West).

- Category: Product categories, including Technology, Office Supplies, and Furniture.

- Profit: Financial data representing either profit (positive values) or loss (negative values).

- Sales: The number of sales made for each transaction.

- Other Columns: Includes details such as Order ID, Ship Mode, Customer Information, and Quantity Sold.

Data Type: The dataset consists of both categorical and numerical data. Key numerical columns include Sales, Profit, and Quantity, while categorical data includes columns like Order ID, Ship Mode, Customer Name, and Product Category.

Data Preprocessing:

- Removed Unnecessary Columns: Eliminated columns like ind1 and ind2 to clean the dataset.
- Handled Missing Values: Replaced missing values in the Returns column with 0 to ensure no gaps in the analysis.
- Renamed Columns: Changed ambiguous column names for clarity, improving readability.
- Ensured Categorical Data Consistency: Checked and maintained consistency in categorical variables like Category, Region, and Ship Mode.
- Converted Date Columns to Date Format: Ensured date fields were in the correct format for time-based analysis.

- Ensured Numeric Data Consistency: Checked numeric columns to ensure they were ready for analysis and visualization.

Visualization Idiom:

A stacked bar chart was selected to show the distribution of profits and losses across categories and regions.

- The x-axis represents regions (Central, East, South, West).
- The y-axis represents profit values (including losses).
- Each bar is divided into sections corresponding to the product categories (Technology, Office Supplies, Furniture), with distinct colors assigned to each category.

Why stacked bar chart?

- A stacked bar chart allows us to effectively display the breakdown of profits and losses across multiple categories within each region. This idiom is well-suited for showing both the overall profit and the individual contributions from each product category.
- The stacked bars allow for easy comparison between categories and regions, and the color coding ensures that the different product categories are easily distinguishable.
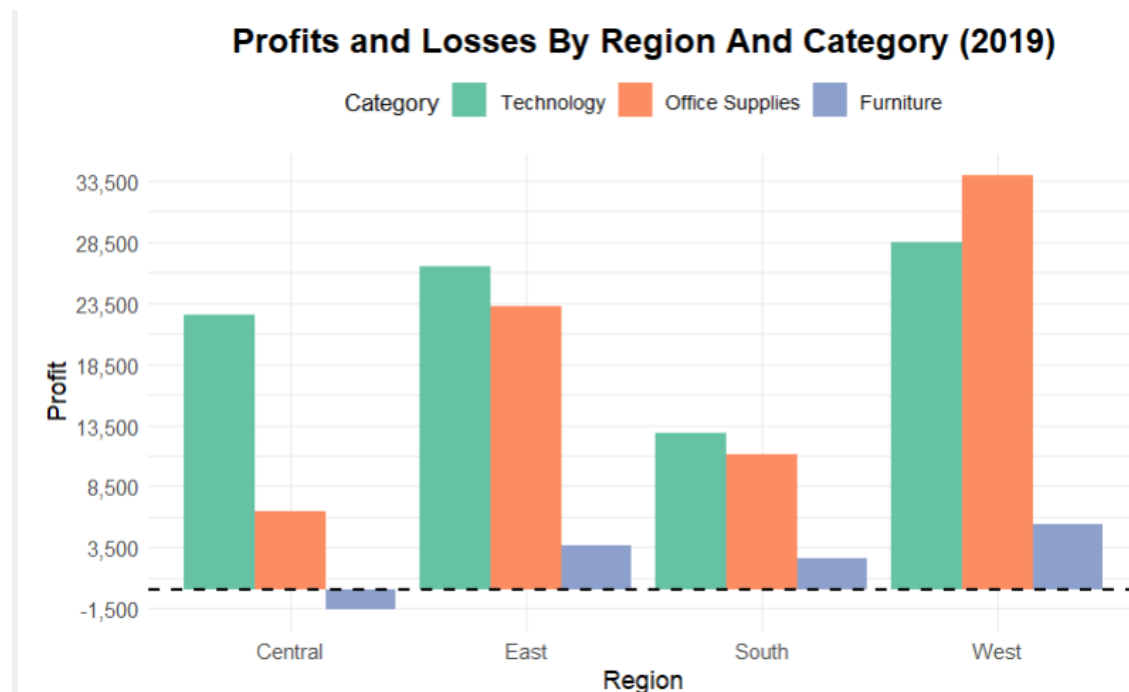
Interpretation:

Region-Wise Breakdown

- Central Region: Technology leads in profit, Office Supplies shows moderate profit, while Furniture incurs losses.

- East Region: Technology dominates profit, Office Supplies contributes moderately, and Furniture shows minimal profit.

- South Region: Small profits in Technology and Office Supplies; Furniture contributes very little.

- West Region: The highest profits, led by Office Supplies, followed by Technology; Furniture shows small but higher profits than other regions.

Category-Wise Analysis

- Technology: Top profit contributor, strongest in the West and East regions.

- Office Supplies: Profitable across regions, highest in the West but second to Technology.

- Furniture: Loss-making in the Central region, with highest profit in the West.

Design Choices:

- **Color Palette:**

  The color palette (light green for Technology, light orange for Office Supplies, and light blue for Furniture) ensures clarity while maintaining an aesthetic balance. It minimizes visual overload while making the chart easy to interpret.

- **Data-Ink Ratio:**

  The chart follows the principle of a high data-ink ratio by minimizing unnecessary elements like excessive gridlines and focusing on the data itself.

- **Strategic Placement:**

  Data labels are strategically placed on top of the bars for easy reading, and the legend is positioned above the chart for clarity.

Conclusion

- This report and visualization provide a comprehensive analysis of profits and losses across product categories and regions. By following best practices in data visualization, the stacked bar chart offers a clear and easy-to-interpret view of how each category contributes to overall profits or losses by region. The insights derived from the chart can help stakeholders make informed decisions to improve performance in underperforming regions and categories.