

# SUPERSTORE DATA VISUALIZATION DASHBOARD

## PROCESS BOOK

### Group - 5:

Name: Sai Sriharsha Griddaluru

CUID: C15358926

Email: [sgridda@clemsn.edu](mailto:sgridda@clemsn.edu)

Name: Kanish Vijay Arunkumar

CUID: C46866282

Email: [karunku@clemsn.edu](mailto:karunku@clemsn.edu)

Name: Tejdeep Uckoo

CUID: C51451099

Email: [tuckoo@clemsn.edu](mailto:tuckoo@clemsn.edu)

Dataset: <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final/>

Website: <https://harsha9554.github.io/gr-05-datviz-project/src/index.html>

Repository: <https://github.com/Harsha9554/gr-05-datviz-project>

Video: <https://drive.google.com/file/d/1gBAtpK6-Vv23AyDUMMAOGrpCd-yUJifH/view>

Overview and Motivation
Related Work
Questions
Data
Exploratory Data Analysis
Design Evolution
Implementation
Evaluation

## Overview:

This project is a data visualization of the Superstore dataset created using d3.js. The project explores sales trends, customer behavior, and product performance through various interactive charts and filters.

## Motivation:

The retail industry is a constantly evolving landscape, and understanding customer behavior and sales trends is crucial for businesses to remain competitive. This project aims to analyze the Superstore dataset, a widely used benchmark for data visualization tasks, to gain insights into key aspects of retail performance.

By leveraging data visualization techniques, we hope to create an interactive dashboard that provides users with the ability to identify top-selling products and categories, gain insights into customer demographics and purchase preferences. Such insights can be invaluable for businesses in making informed decisions regarding product development, marketing strategies, inventory management, and resource allocation.

Inspiration from Existing Work:

Several existing data visualizations of the Superstore dataset served as inspiration for our project. These include:

Superstore Sample Data Visualization with Tableau:

<https://www.tableau.com/data-insights/dashboard-showcase/superstore>

Interactive Data Visualization with the Superstore Dataset:

<https://m.youtube.com/watch?v=TwALdb9RIvE>

These resources provided valuable insights into effective visualization techniques and design approaches for exploring the Superstore dataset.

Overall, this project is motivated by the desire to leverage the power of data visualization to unlock valuable insights from the Superstore dataset. By providing an interactive and user-friendly interface, we aim to empower users to explore the data independently and gain a deeper understanding of retail sales trends and customer behavior. This knowledge can be instrumental for businesses in making data-driven decisions that optimize operations and enhance profitability.

## **Related Work :**

We explored existing data visualizations of the Superstore dataset, such as a Tableau dashboard showcasing various chart types and interactive features, and a YouTube video demonstrating D3.js implementation for an interactive dashboard. These resources offered valuable insights into design choices and visualization techniques, inspiring us to incorporate interactive filtering and clear organization by product categories and sub-categories in our own dashboard. The visualizations that inspired us -

<https://www.tableau.com/data-insights/dashboard-showcase/superstore> , and

<https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/Superstore-Data-Visualization-with-Power-BI/td-p/2977837>

## **Questions:**

- 1. In which states are the superstore actually doing good by having high sales/profit? and which states it needs to work on (low profit/sales or loss)?**
- 2. How are the customers distributed in the regions of US and who gives the highest profit and who is creating a loss in a particular sale for the superstore?**
- 3. What is the profit/loss for the items in the superstore?**
- 4. What is the sales of the products in the superstore and are they going in profit or loss?**

## Data:

Our exploration of the Superstore began with a crucial first step: acquiring the data. From the vast repository of datasets on Kaggle, we downloaded the Superstore dataset. This raw data, housed in the "superstore.csv" file, represented the unpolished potential awaiting our analysis. However, upon closer inspection, we encountered inconsistencies and missing values within the data. This necessitated a meticulous data preprocessing phase; we started data cleansing and refinement, employing various techniques to remove missing data points, clean and standardize inconsistencies, and ultimately transform the data into a format conducive to visualization.

This process resulted in the creation of several key files. The "superstore-subset-processed.csv" file emerged as the cornerstone of our analysis, representing the refined and ready-to-explore version of the data. We also preserved a backup of the processed data in the "superstore-subset.csv" file for safekeeping. Additionally, the "cities-location.csv" and "US\_states.json" files provided valuable spatial context for our visualizations, enabling us to explore trends across geographical regions using map i.e., geojson file of the states of the US.

## Exploratory Data Analysis:

Our exploration began with analyzing data, uncovering hidden patterns and relationships through descriptive statistics and compelling visualizations. Scatter plots, bar charts, and heatmaps which we first made in Tableau revealed fascinating trends across categories, regions, and time, leading to valuable insights for targeted marketing and product optimization. Further analysis delved into granular customer segments, uncovering specific behaviors and preferences that informed personalized understanding and interactive exploration. Unexpectedly we had new findings from this, highlighting underperforming categories, regional variations, and profitable sales channels. These insights served as stepping stones, guiding us towards actionable recommendations for enhancing the Superstore dashboard.

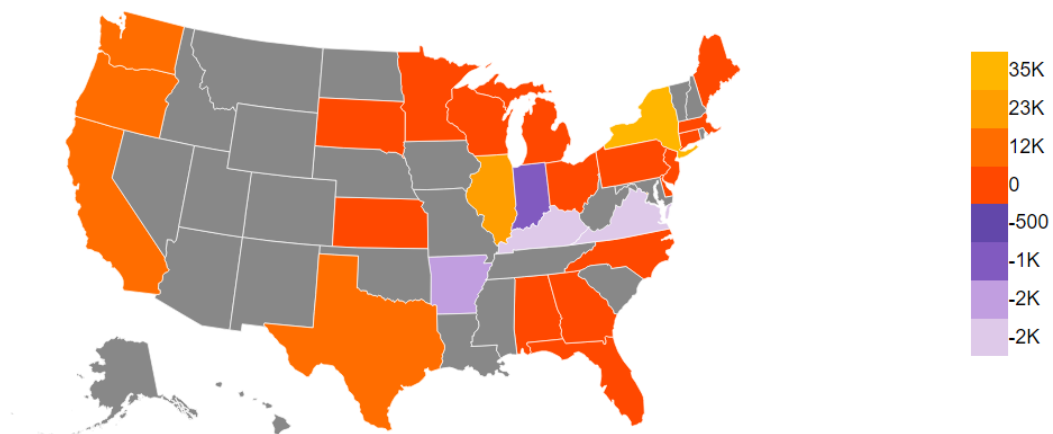
## Design evolution:

After choosing the dataset we planned on doing a heatmap of US, bar chart, pie chart and a sankey graph which represents different parts of the dataset like profit and shipment method, and factors like discount too, but Dr.Fredrico suggested in class that the visualizations are supposed to be interactive and they should answer reasonable questions and not just that, We should make it so that the dashboard should narrate a story about the dataset and we started to brainstorm for new ideas and we still held on to the idea of the heatmap because we felt it gives a fun touch to the dashboard and it also would help the person understand the dashboard better. Along with that we we planned on making a pie chart, scatter plot and a bar chart.

A basic explanation of the graphs and why and how we came up with them are given below and the changes done after suggestions from Dr. Fredrico are mentioned.

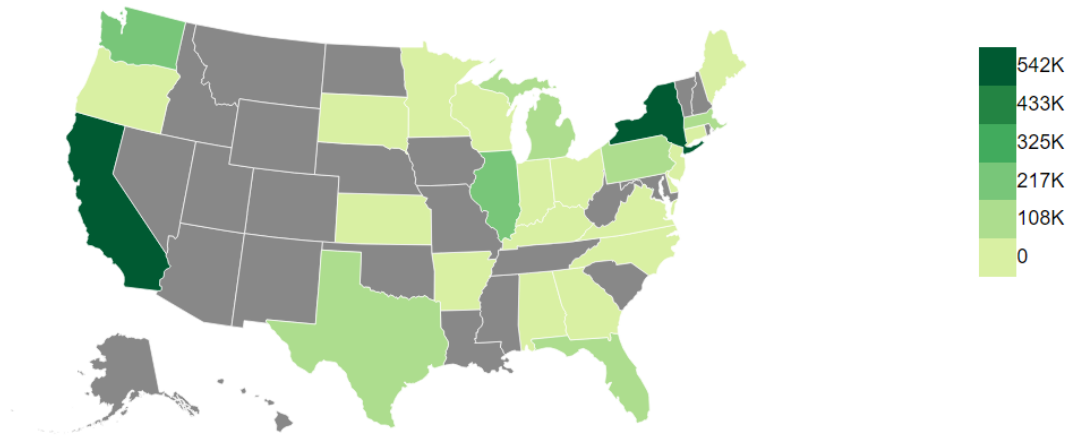
The geospatial visualization/heat map of the US map shows the profit and loss across different states in the map orange represents profit and purple represents loss

● Profit ○ Sales

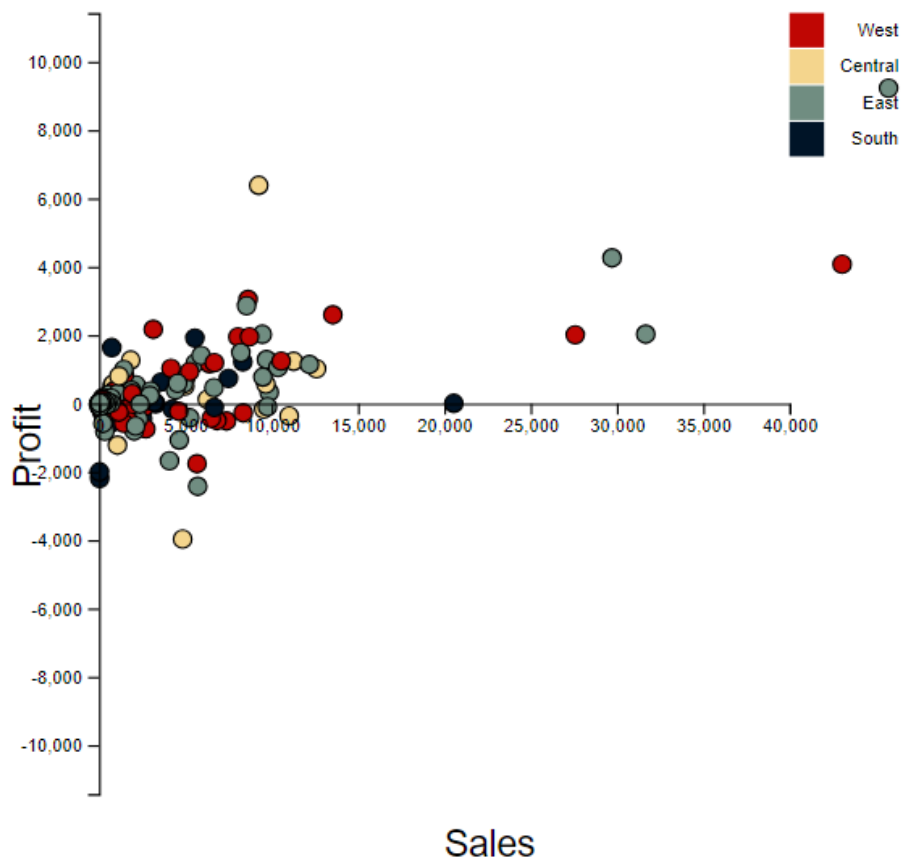


In the image below we can see the sales amount across the map from light green representing low to dark green representing high sales.

○ Profit ● Sales

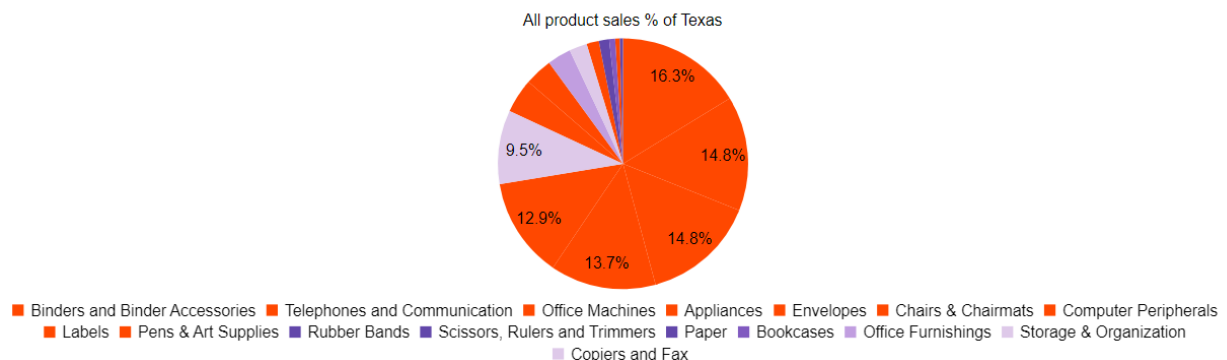


This scatter plot is plotted between profit and sales and is colour coded with east, south, central and west parts of US. each point represents a customer, earlier we attached names with the points, but to respect peoples privacy and as per the curriculum of the class Dr.Fredrico guided us to exclude the customers name from the graph.

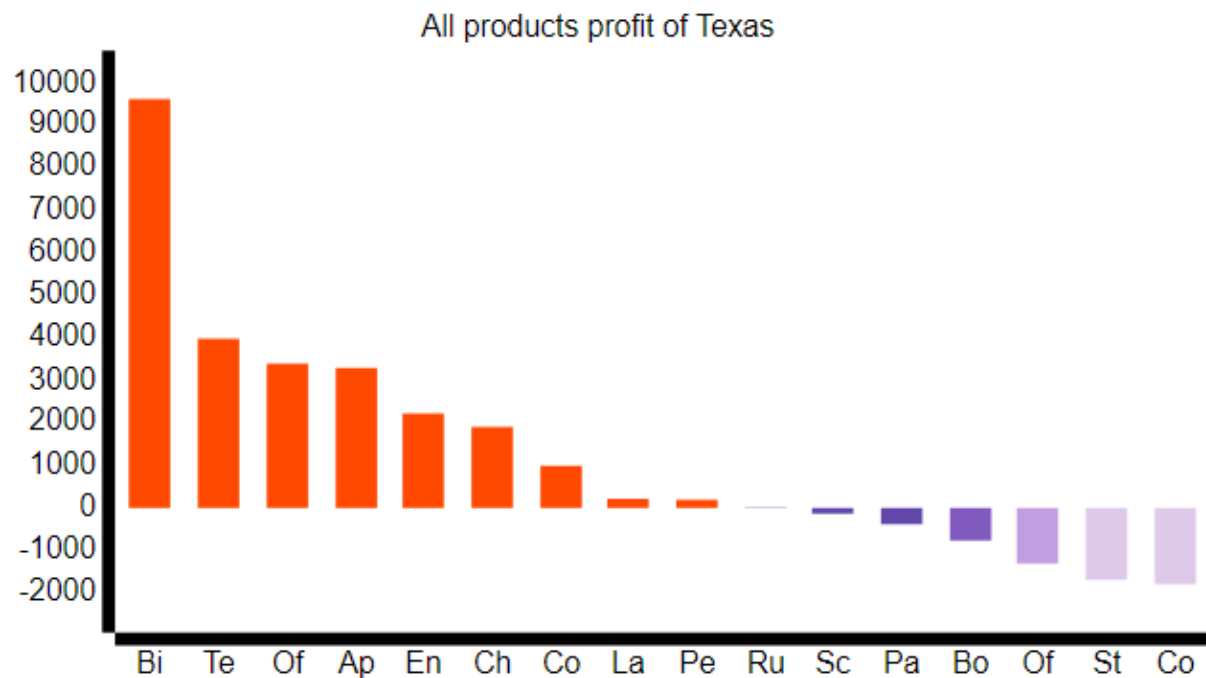


In this graph we compare the sales of all types of products sold by the super store and is colour coded by profit or loss according to the previous heatmap

This helps us understand which product has the highest amount of sales and if the store has profited from that or are they suffering from a loss. This single graph answers the most crucial questions which arise when looking at the dataset



In this graph we visualize the profit and loss of all the products available in the superstore. This graph supports the previous pie chart and helps people understand it even better. The previous graph and this graph together provides clear insight on the sales, **profit and loss across regions for specific items sold by the superstore.**



## Implementation:

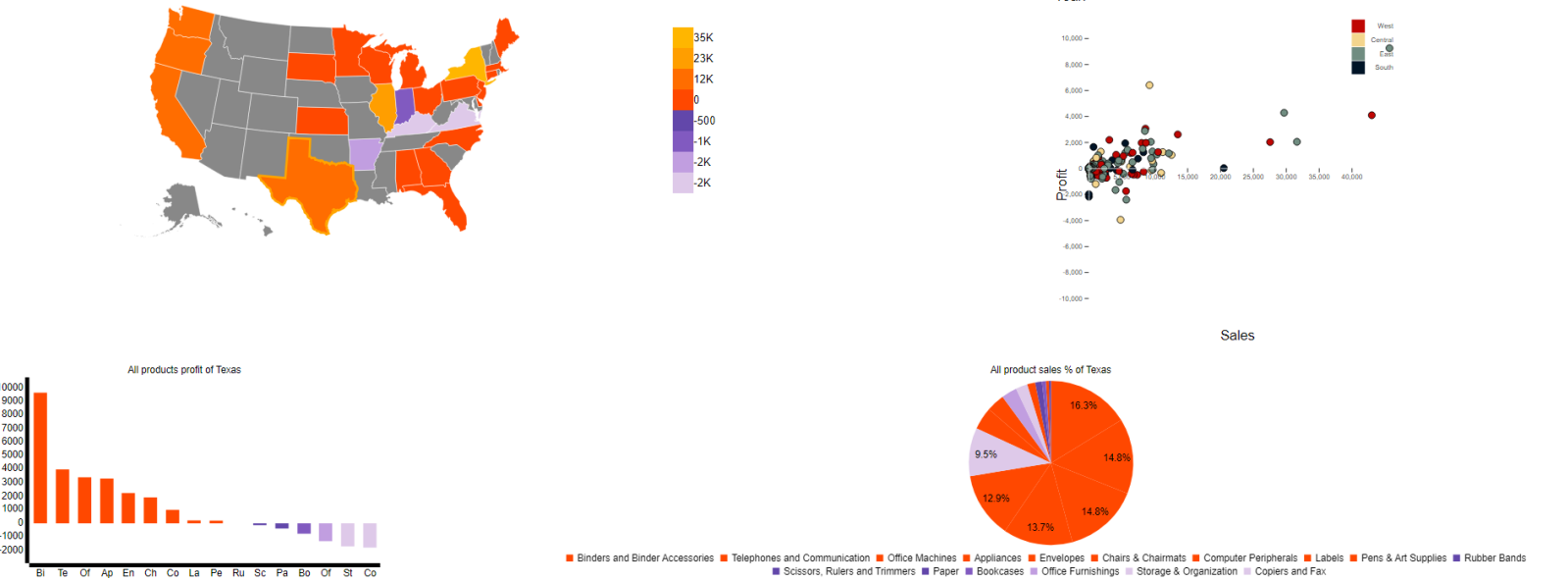
The previous subheading explains the changes made and gives a breif overview of the graphs

After class presentation and getting suggestions from Dr.fredrico and consulting other peers for more suggestions we came up with our final dashboard



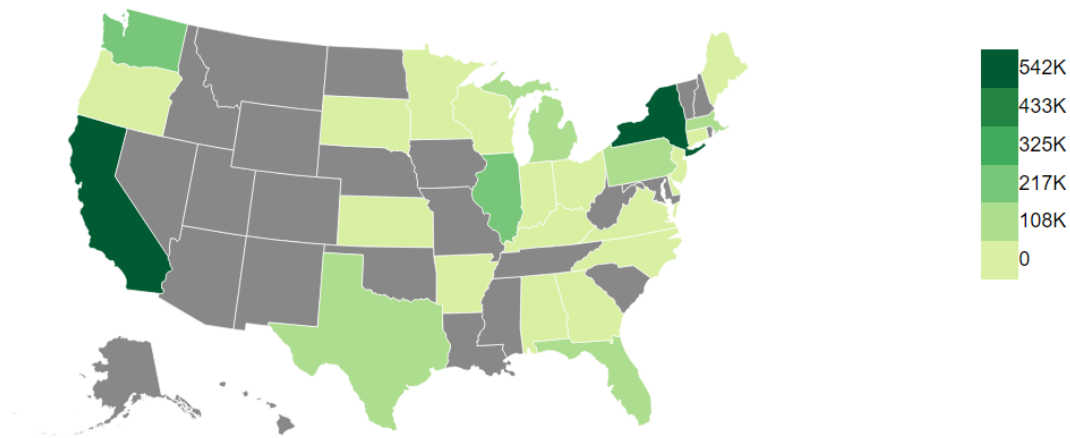
Superstore Data Visualization Dashboard

Year: ▼ All  
☒ Profit ☐ Sales



The heatmap is the main visualization in this dashboard as it interacts with the other visualizations if we click on a particular state in the visualization the below graphs changes according to the state, for instance in the above image ive clicked on texas and the other graphs shows content for texas.

☐ Profit ☒ Sales



We can also visualize the sales heatmap to compare the data with the profit / loss graphs below.

Finally after suggestions we made it so that we can only visualize a specific year for the scatter plot as each point represents a customer and if we try to visualize all the years/ customers throughout the years it will be clustered and clumsy and we felt it will do no good.

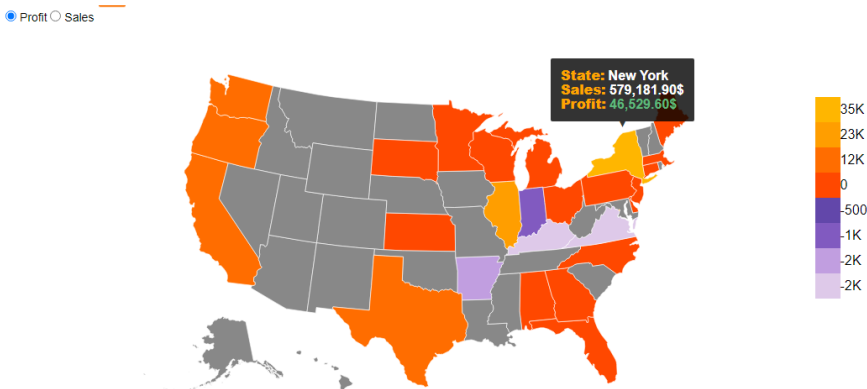
## Evaluation:

Insights:

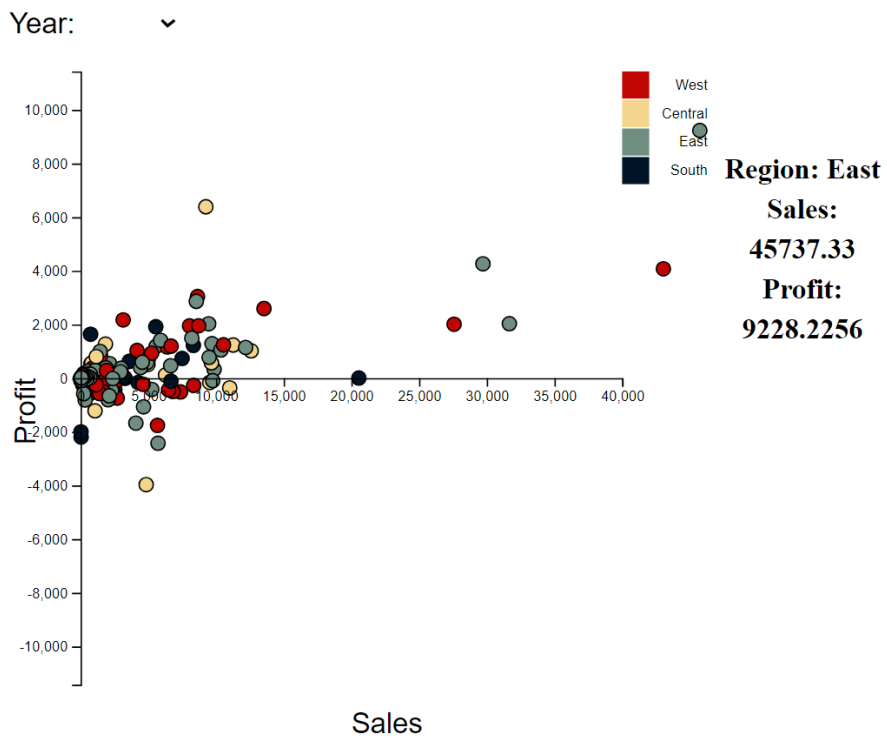
- This dashboard is to be a valuable asset to the superstore and this dashboard's format can be used to create various other dashboards for other businesses
- Firstly we can identify which region has the highest sales/profit and which region has the lowest sales/profit or even loss and the superstore can take actions accordingly
- We can also see the customers and which region of the country they belong to, if we split them according to their state it would have been clumsy, so we decided to split them based on their region, as this is a project we anonymised the customers name. The super store can identify regular customers(using the year filter), customers who provide them with the highest sales and work on improving the relationship
- The other two graphs work together aiding the superstore to know which product to manufacture more or which product to stop manufacturing based on profit/ loss/ sales.

Questions and how are they answered:

**In which states are the superstore actually doing good by having high sales/profit? and which states it needs to work on (low profit/sales or loss)?**

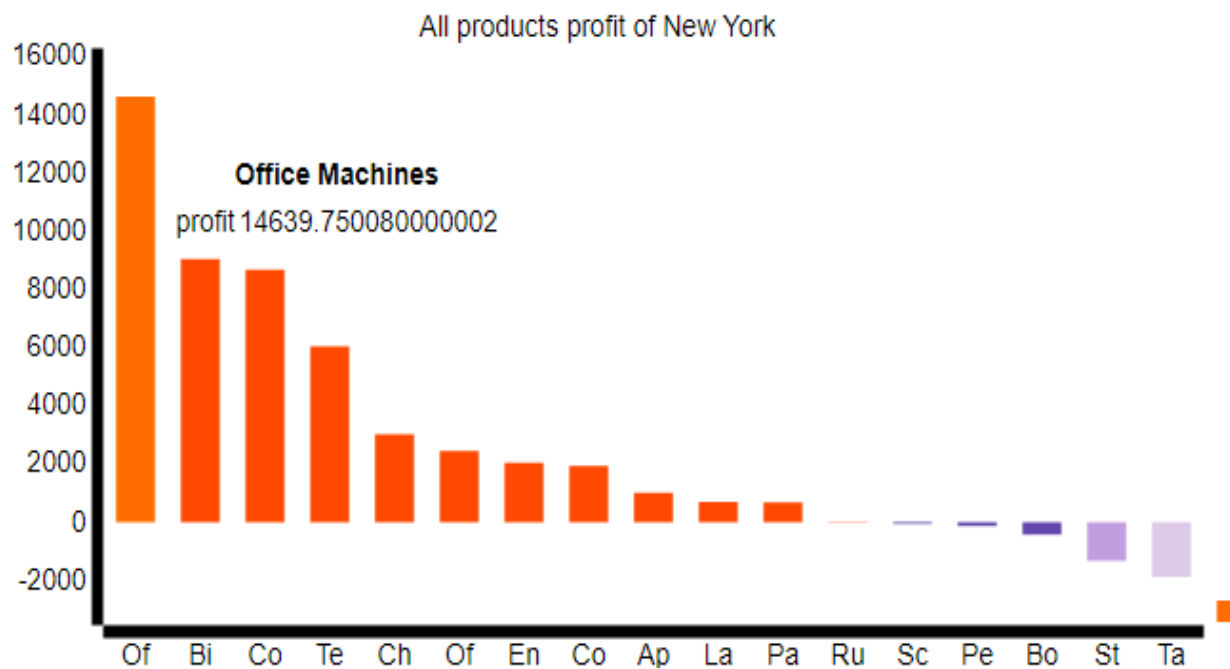


**How are the customers distributed in the regions of US and who gives the highest profit and who is creating a loss in a particular sale for the superstore?**



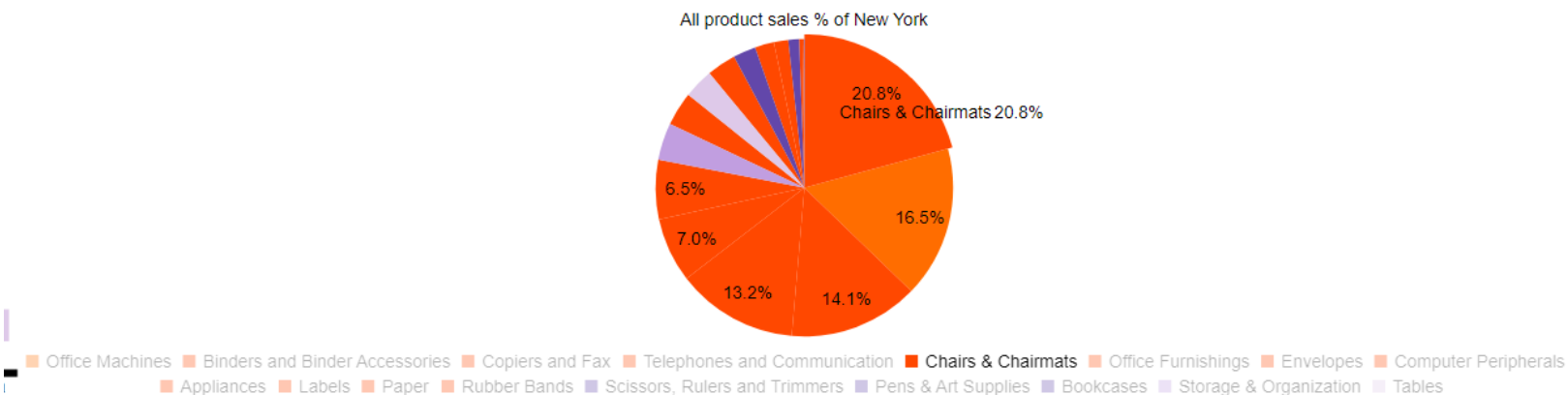
In this image we are hover on the top right point and it shows us the customers region sales and the profit they provide the store with, in an ideal situation it will show their name too but as this is a project we anonymised the customer.

**What is the profit/loss for the items in the superstore?**



Hovering across the bars gives us the names of the item and the x-axis has the abbreviation of the item.

What is the sales of the products in the superstore and are they going in profit or loss?



Hovering on the pie chart shows the name and it is color coded based on profit or loss

**Timeline & Progress :**

## Superstore Data Visualization Project - Work Report

Project Members: Harsha, Kanish, Tejdeep

Dates: October 1st - December 9th, 2023

Tools: D3.js, JavaScript

October 1st:

Harsha: Reviewed and discussed the Superstore dataset.

Kanish: Explored the attributes of the dataset as an initial step to create visualizations

Tejdeep: Started gathering resources and information on d3.js.

October 2nd - 7th:

Harsha: Conducted initial data cleaning and pre-processing.

Kanish: Brainstormed potential design ideas and visualizations.

Tejdeep: Experimented with creating simple charts.

October 8th - 14th:

Harsha: Implemented data manipulation and formatting using d3.js.

Kanish: Defined the initial layout and structure of the dashboard

Tejdeep: Developed the core functionality of the first visualization.

October 15th - 21st:

Harsha: Started developing in a new development branch and hosted the site on Github pages.

Kanish: Designed interactive filters and user interface elements.

Tejdeep: Created additional visualizations and integrated them into the dashboard

October 22nd - 28th:

Harsha: Implemented data interactivity and brushing/linking between visualizations.

Kanish: Added tooltips for enhanced data exploration.

Tejdeep: Performed unit testing and ensured consistent functionality across different browsers.

October 29th - November 4th:

Harsha: Refined the aesthetics and visual design of the dashboard.

Kanish: Prepared presentation materials and practiced demoing the project.

Tejdeep: Conducted final testing and addressed any bugs or issues identified.

November 5th - 11th:

Harsha: Revised the project documentation based on feedback.

Kanish: Rehearsed the presentation and incorporated feedback.

Tejdeep: Finalized the coder.

November 12th - 18th:

Harsha: Worked on improving the performance and responsiveness of the dashboard.

Kanish: Worked on improving the dashboard with Harsha.

Tejdeep: Wrote down additional insights and findings.

November 19th - 25th:

Harsha: Responded to feedback from friends and made final adjustments.

Kanish: Prepared the presentation for the final project showcase.

Tejdeep: Made adjustments to the code along with Harsha

.

November 26th - December 9th:

Harsha: Fielded feedback from classmates and instructors, worked on the final code and worked on merging the development branch to the master branch on GitHub using pull requests.

Kanish: Presented the project to the class, worked on showcasing project with Git

Tejdeep: Reflected on the project experience and identified areas for future improvement, Updated the project repository