Sathwik Ch

17PW32

M.Sc Software Systems

EXPLORATORY DATA ANALYSIS

Data Set Description:

I have chosen Super Shop Dataset from Kaggle which gives information about different shops around the USA.

It consists of 13 columns:

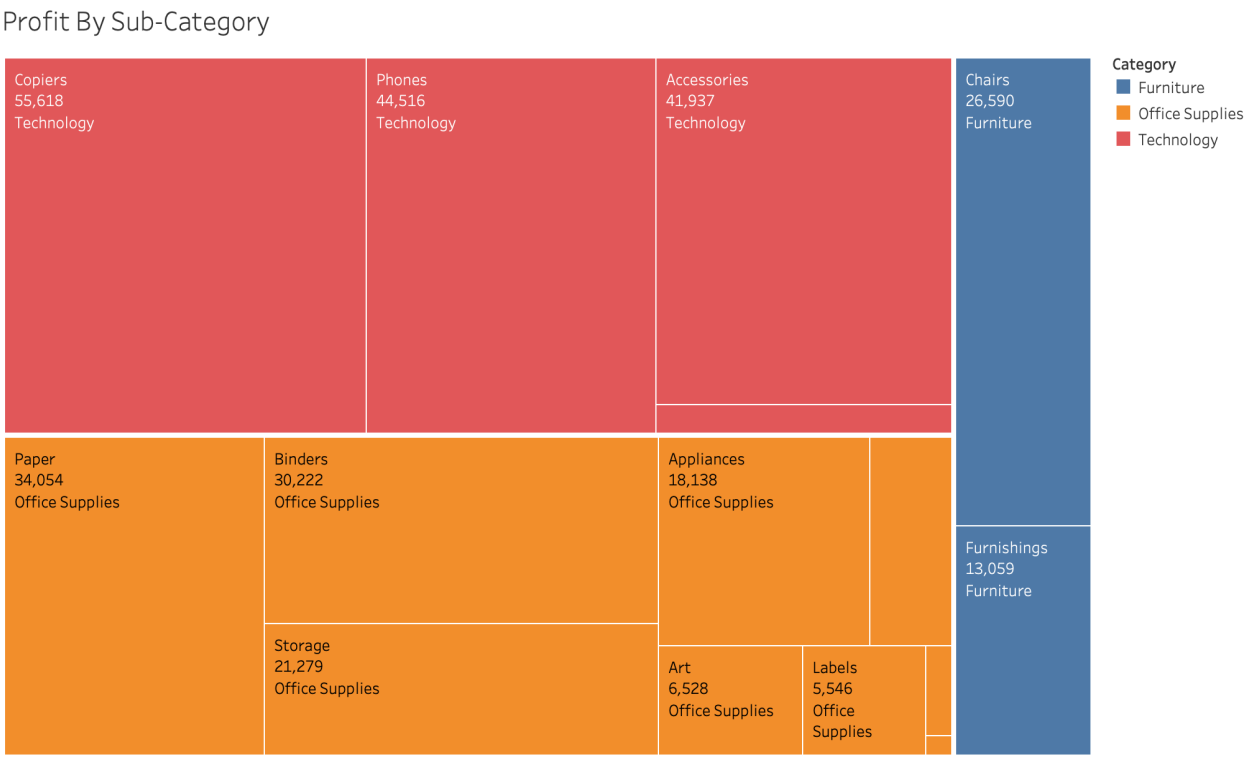
- Ship Mode
- Segment
- Country
- City
- State
- Postal Code
- Region
- Category
- Sub-Category
- Sales
- Discount
- Profit
- Quantity

Questions:

1. Which Regions made the highest and lowest based on sales and profits?
2. State Which made the highest sales?
3. Which products made the highest and lowest profits?

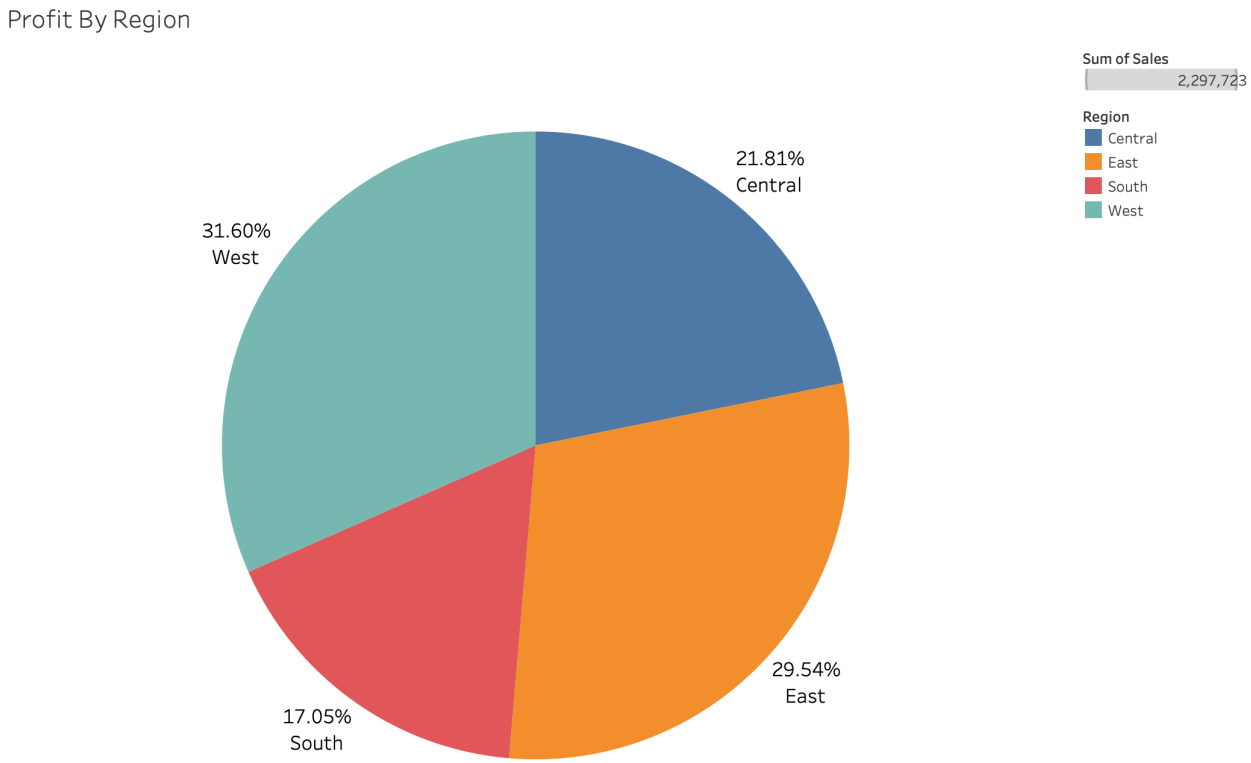
Each Page consists of a Chart and a description.

TreeMap:



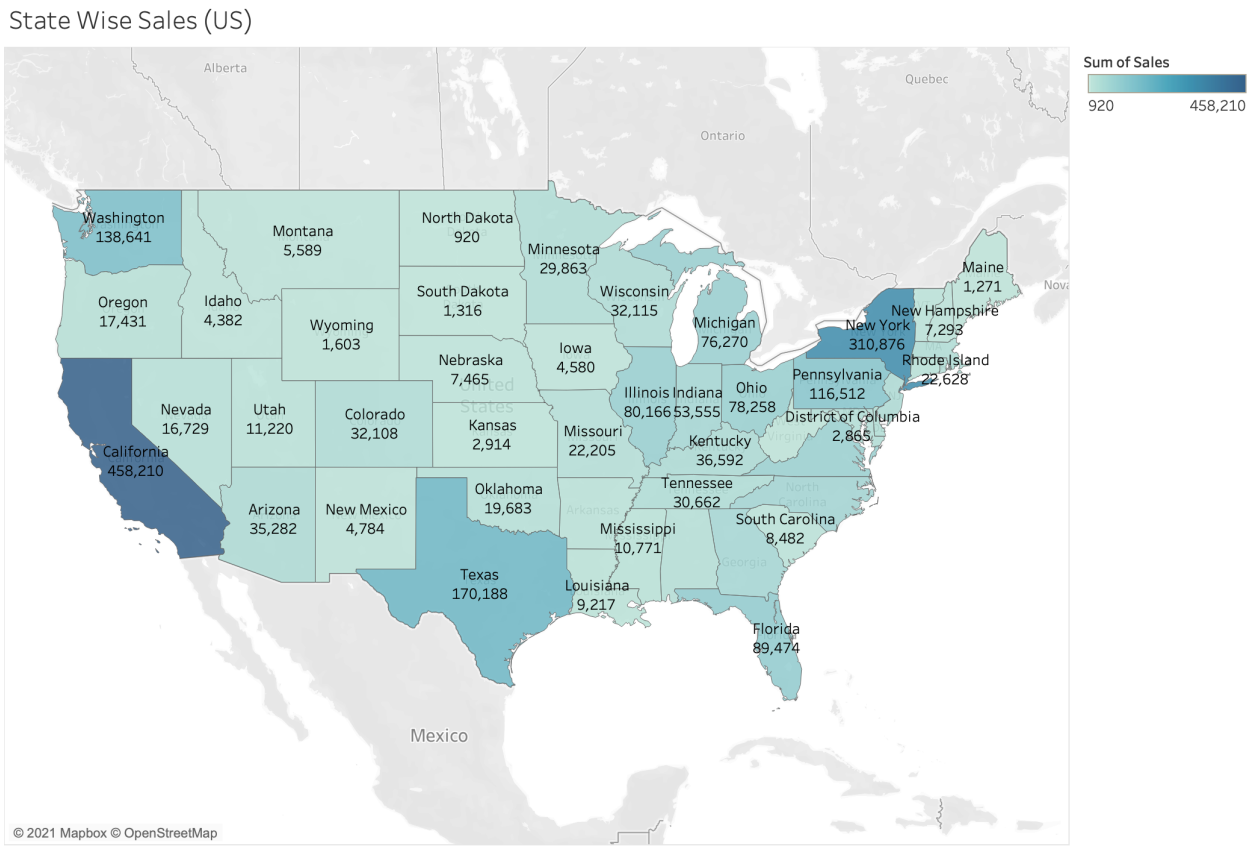
The above TreeMap shows profits made by each Category where blue represents **Furniture**, orange represents **Office Supplies** and red represents **Technology** and each category is again sub-divided into sub-categories. From this chart, we can say that profits made by **Technology** and **Office Supplies** are almost equal and then we have **Furniture**. In **Technology** as the numbers suggest **Copiers** have made the highest profit, In-**Office Supplies** it's **Paper** and in **Furniture** it's **Chairs**.

Pie-Chart:



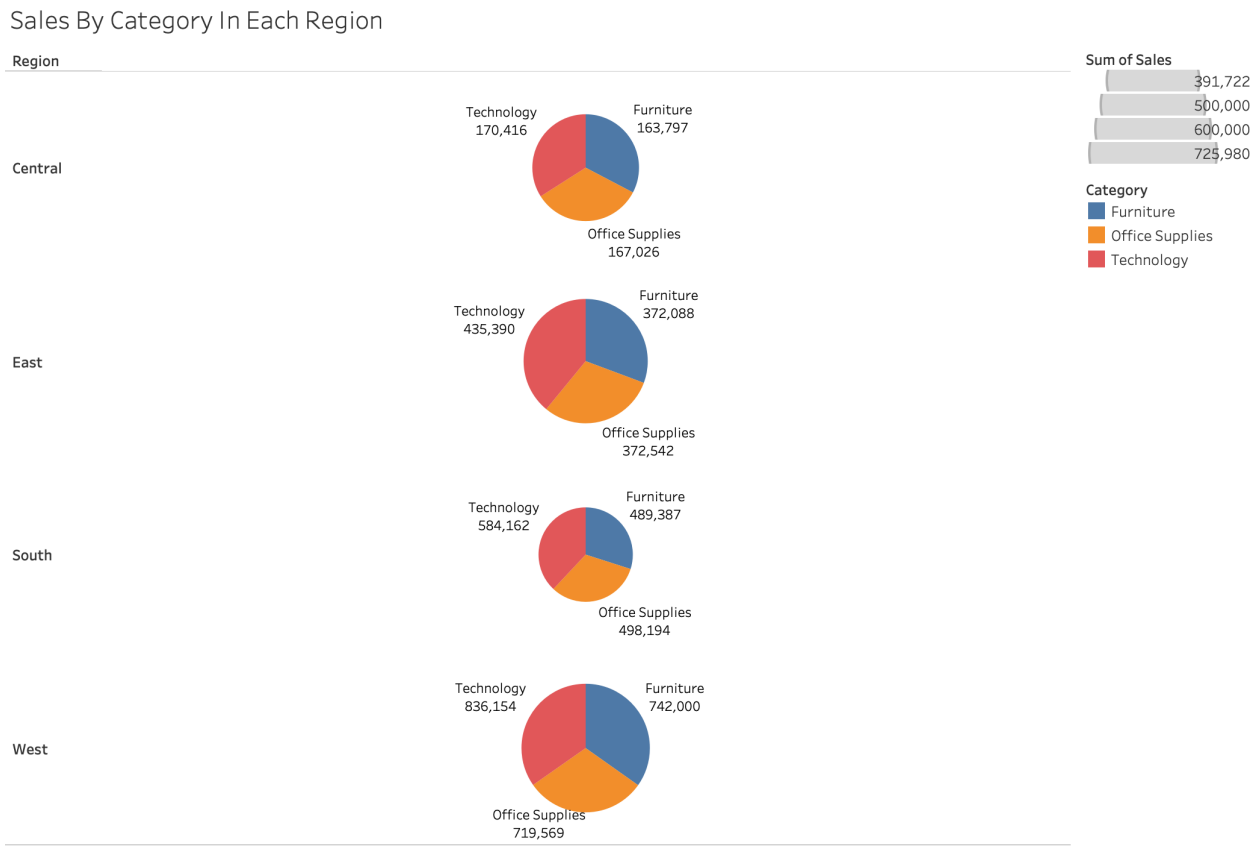
The above Pie-Chart shows profit made in each region (Central, East, South and West). As the numbers suggest we **West** region making the highest profits and then followed by **East**, **Central** and **South** region making the least profits.

Map-Chart:



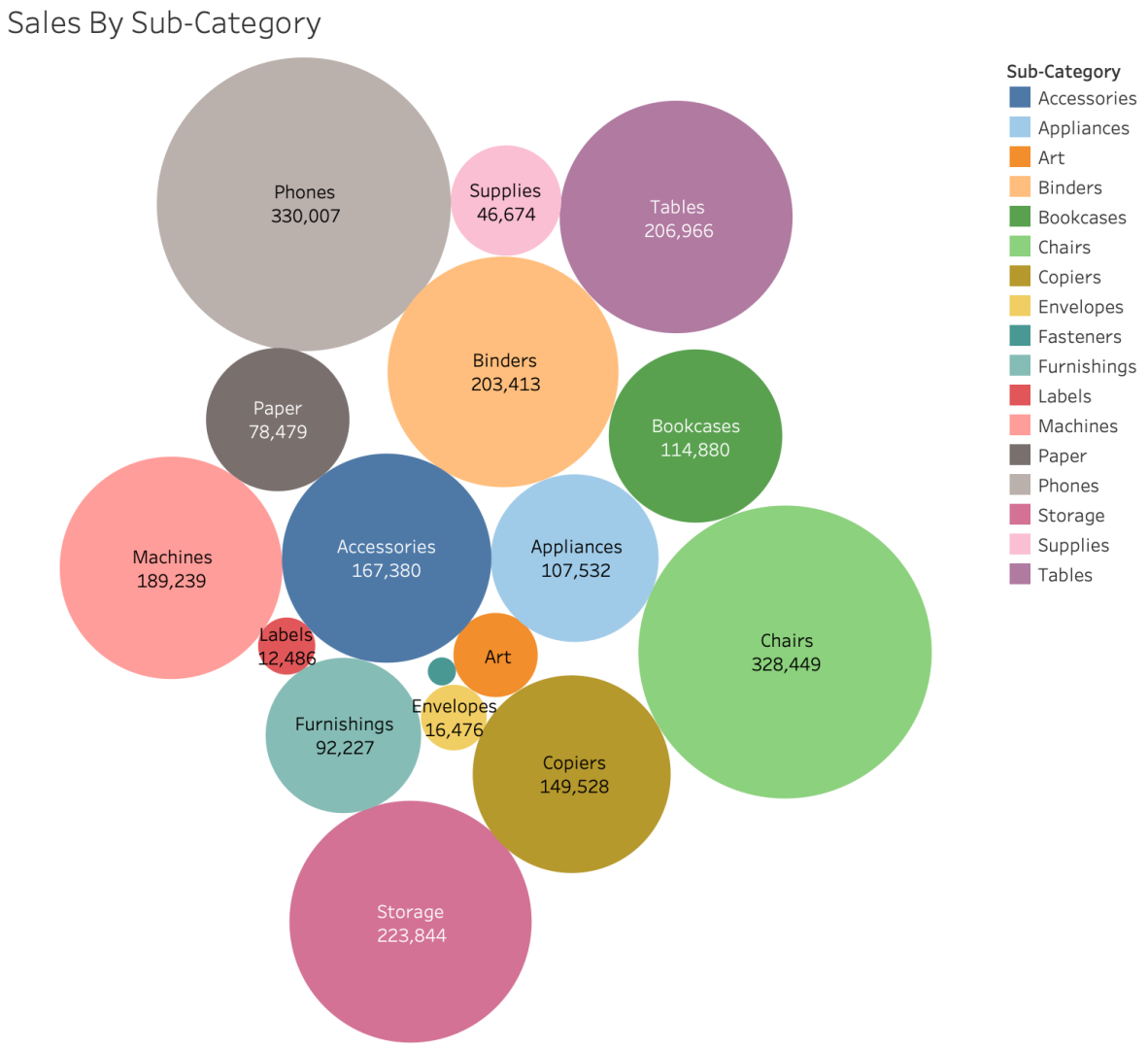
The above Map-Chart represents the sales made by each state in the US. The highest sales were made in **California** with a total of **458, 210** sales. Map-Charts help us in finding places where sales of profits are the highest or lowest.

Pie-Charts:



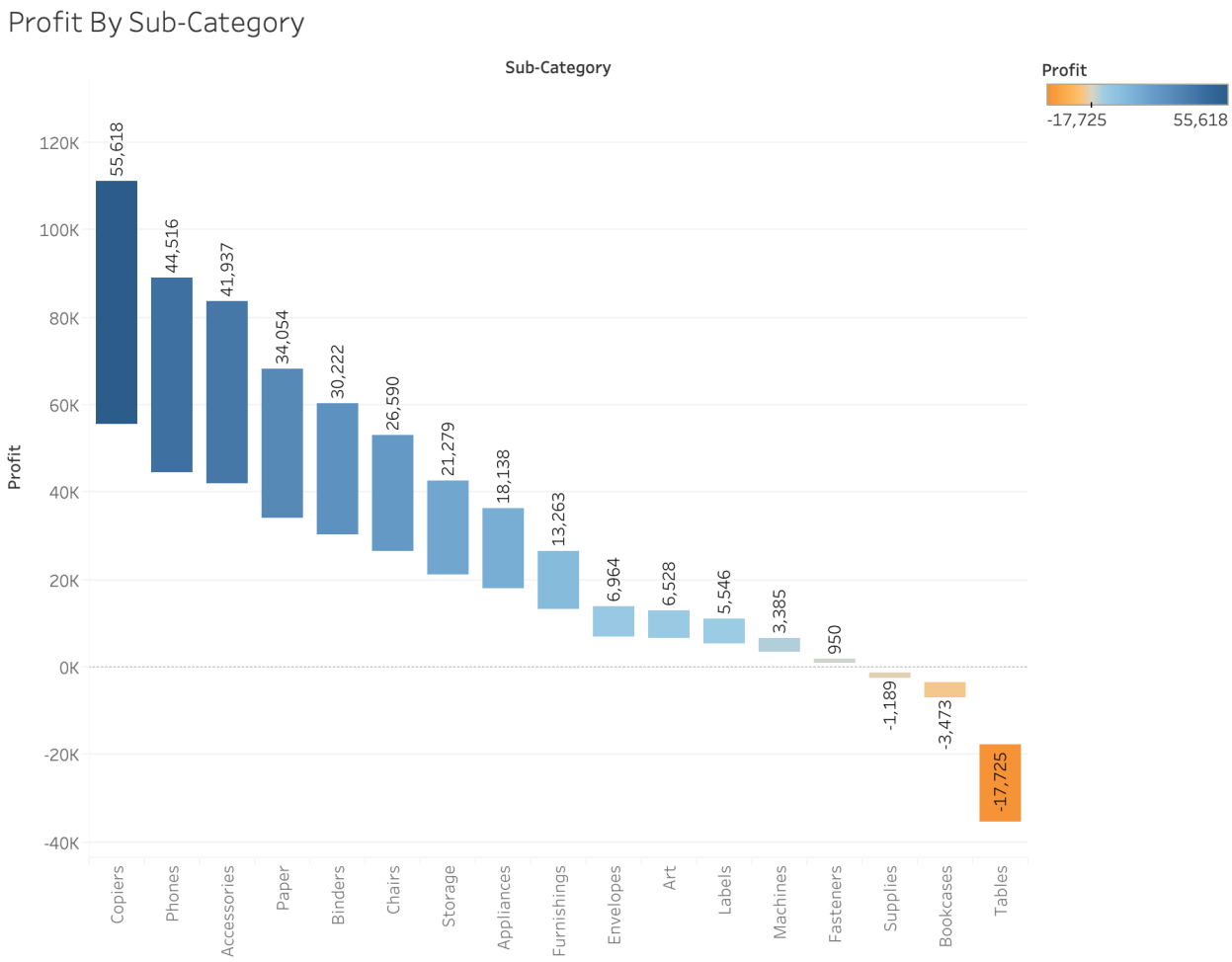
The above Image has multiple Pie-Charts based on region, the size of a Pie-chart is decided by sales done in that particular region and each Pie-Chart depicts the sales made by each category. From this, we can say that the **West** region has the highest sales and **Technology** is making more sales in the West region, and the South region has the lowest sales but even in the **South** region **Technology** is making the highest sales.

Bubble-Chart:



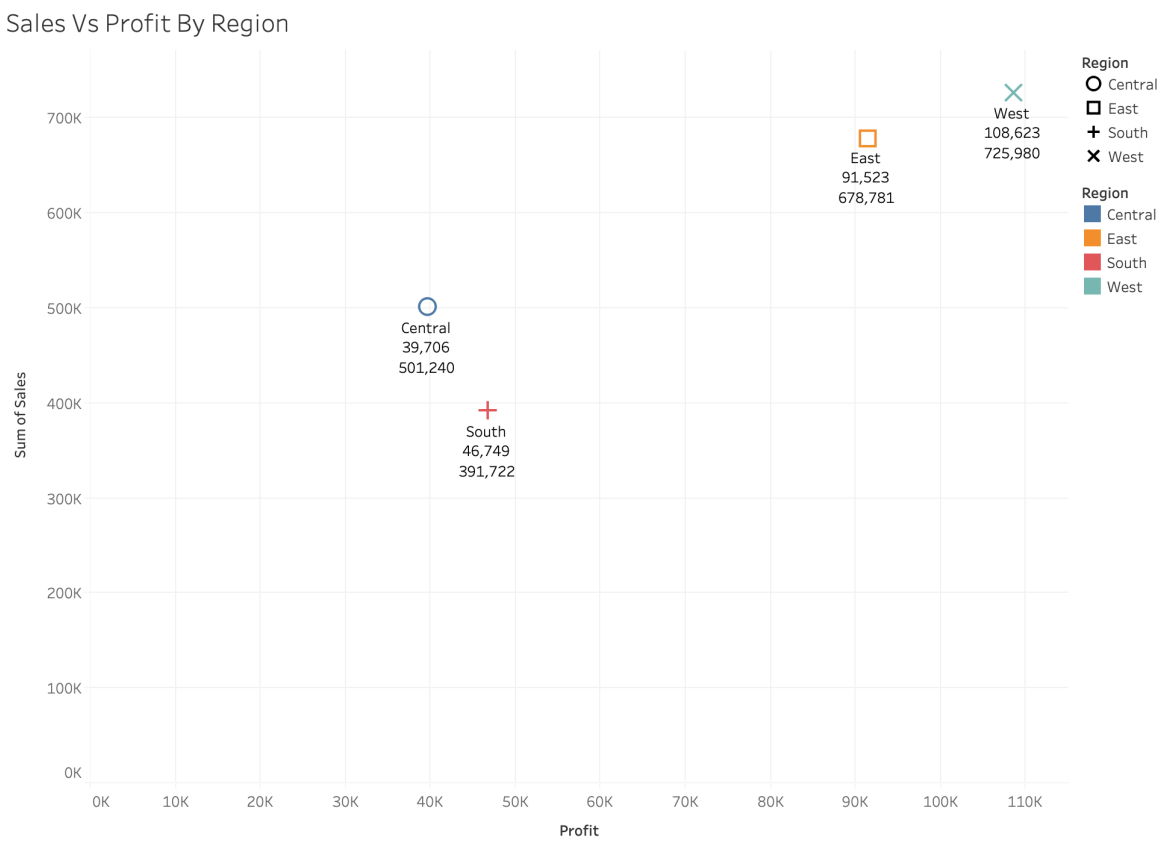
The above Bubble-Chart shows Sales made by each product where each bubble represents a product and the size of the bubble is proportional to the sales made by those products. The largest bubble over here is **Phones** and it has made the **highest** sales compared to other products and the smallest bubble is **Fasteners** which made the **lowest** sales.

Gantt-Chart:



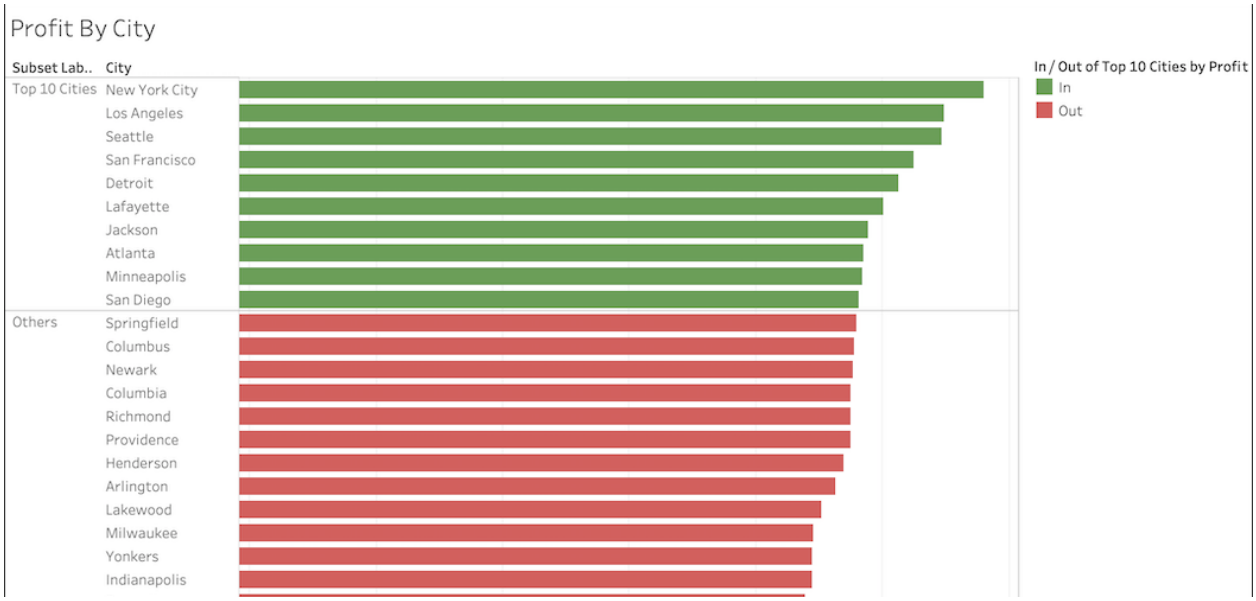
The above Gantt-Chart is used for visualizing profits made by each Sub-Category. On the y-axis, we the **Profit** scale and on the x-axis, we have the **Product** names. Gantt-Chart can be used to negative values too, we can see that **Tables** have made the highest negative profit and the **Copiers** made the highest profit. So, In TreeMap we cannot show negative values in that case if we want to observe negative values we can go with Gantt-Charts.

Scatter-Plot:



The above plot is a Scatter-Plot which shows **Sales** on the y-axis and **Profit** on the x-axis. The different regions that have been plotted are used to show both the sales and profits in a particular region. In the **West** region, we have the highest sales and profits and the **South** region having the least sales and profits.

Bar-Chart:



The above Bar-Chart is used to visualize the profits made by each city in the US. The green bars are the **Top 10 Cities** by Profit and the red bars are other cities with varying profits.