**TASK 2: VISUALIZATION OF SUPERSTORE DATASET**

**Dataset source:**

<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final?select=Sample+-+Superstore.csv>

**Cleaned dataset:**

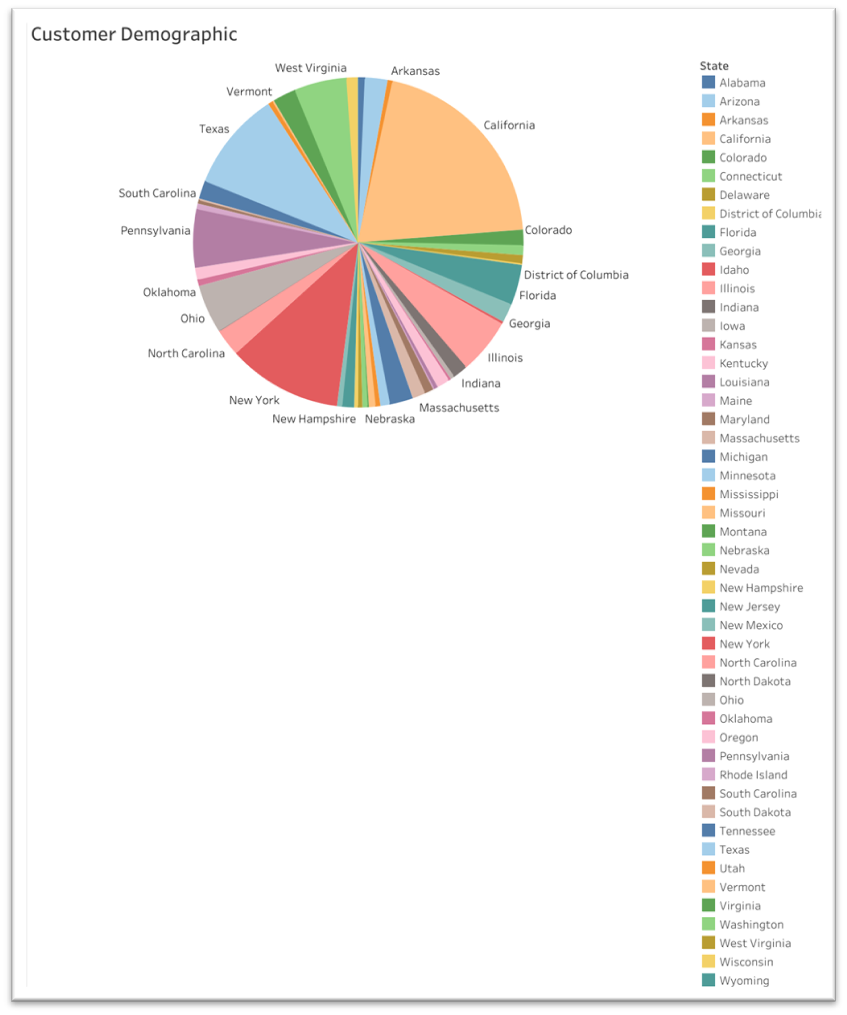
<https://docs.google.com/spreadsheets/d/15BeDA8kQ6lN_FkOiYgXhT0LnzvtRFs6q/edit?usp=drive_link&ouid=106760994882511478209&rtpof=true&sd=true>

**Data Visualization using Tableau:**

<https://public.tableau.com/views/Superstore-Visualization_17586529430500/CustomerDemographic?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

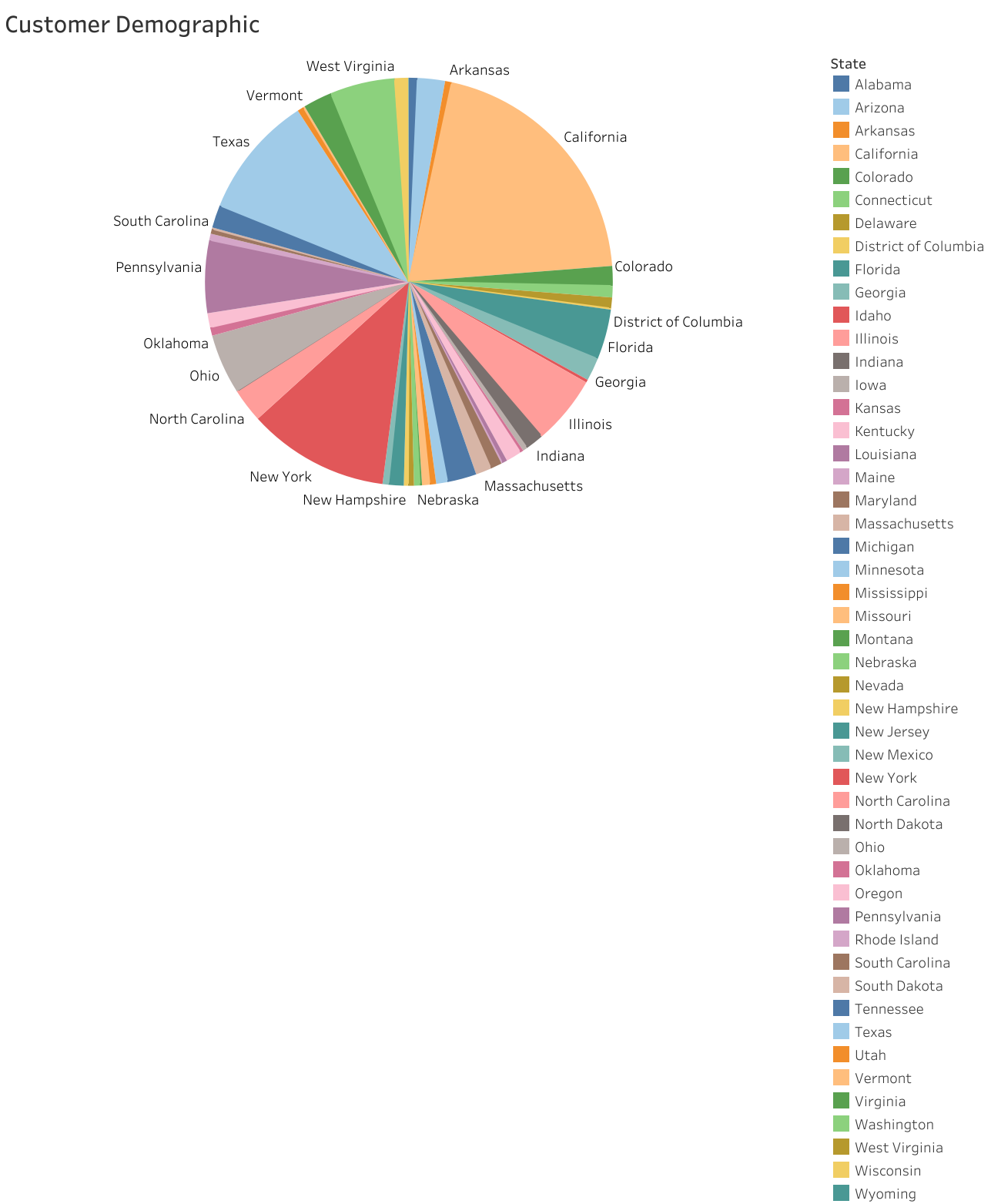
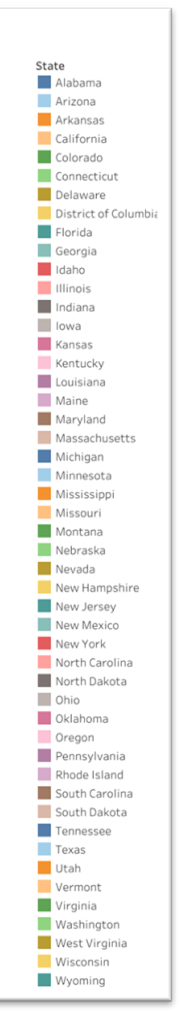
By Kusum Talreja

**Customer Demographic:**

The pie chart below shows the distribution of unique order IDs with respect to the USA states. Basically, it shows the number of customers in each state.

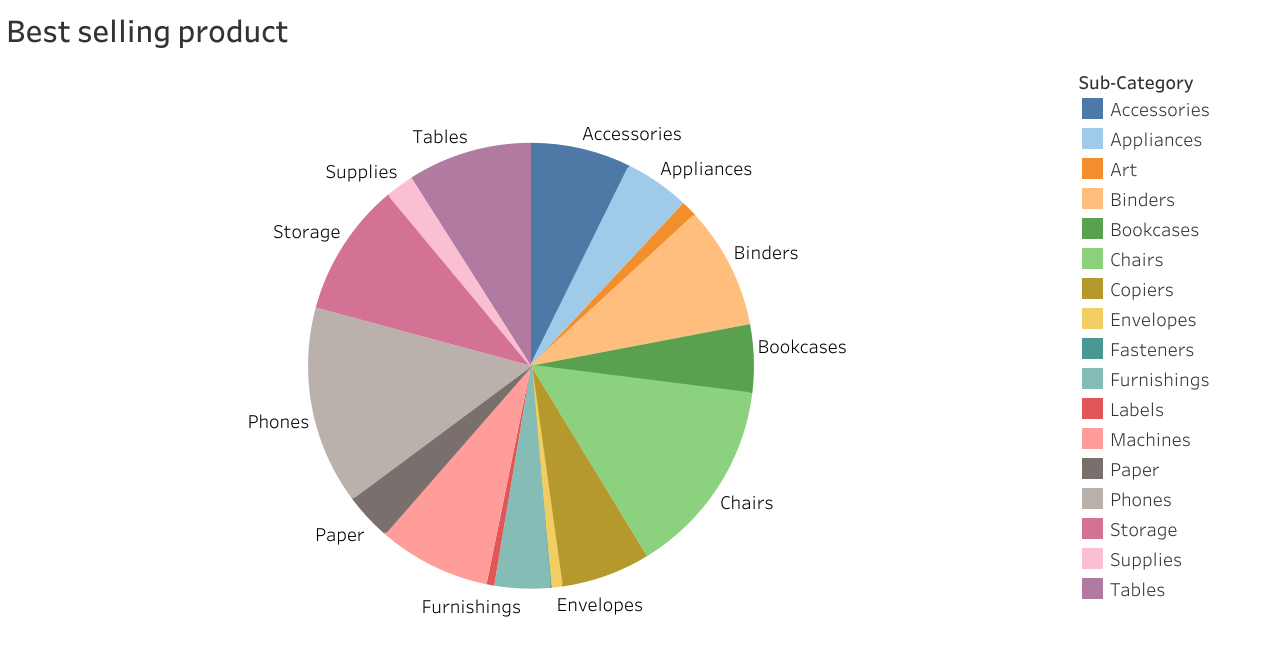
**Most number of customers reside in California**, followed by New York and Texas, which is surprising, we will later see why.

This means Superstore has a strong customer base in those states so they can establish more stores in those areas.



**Best-selling product:**

Phones and chairs are the top 2 best-selling products. The stores should stock up more of these to make sure they never run out if a customer shows up looking for any one of them.

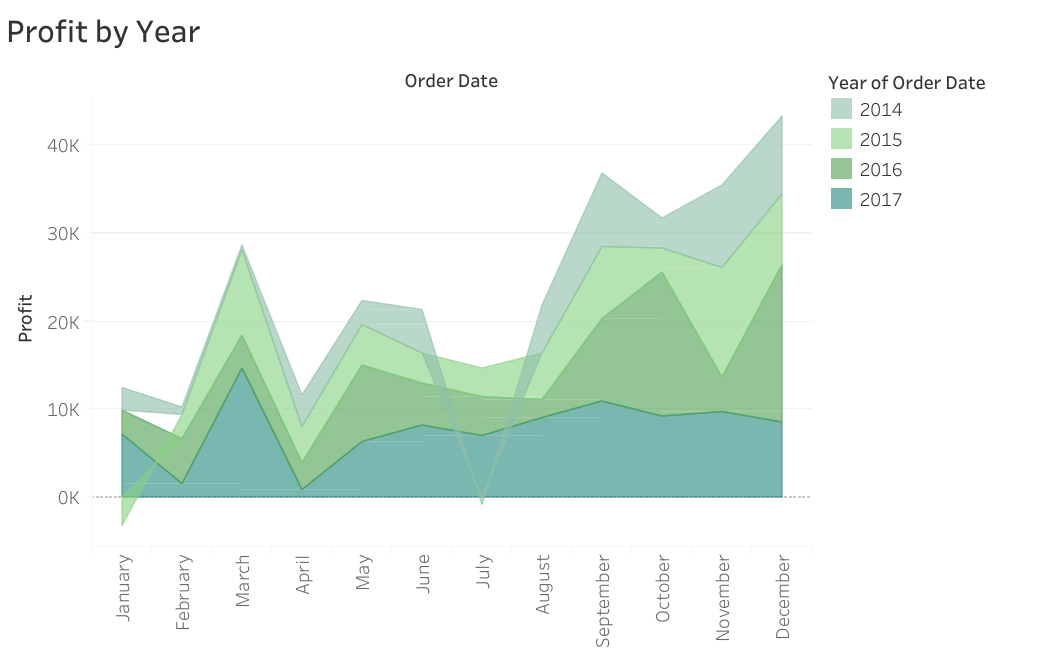
****

**Profit by Year:**

First observation is that **the month of February has very low profit compared to the other months for all the years.** The low sales have been consistent so one advice for the company would be to look more into this. A good suggestion would be to offer more discounts during that month to boost sales. They can also advertise products which are only bought in February, like gifts for Valentine’s Day. They can stock up relevant things for such important occasions.

**A huge loss is noticeable in the years 2014 and 2015 for the months July and January simultaneously**. Superstore should look into this past data to understand the reasoning behind it, in order to prevent this from happening again.

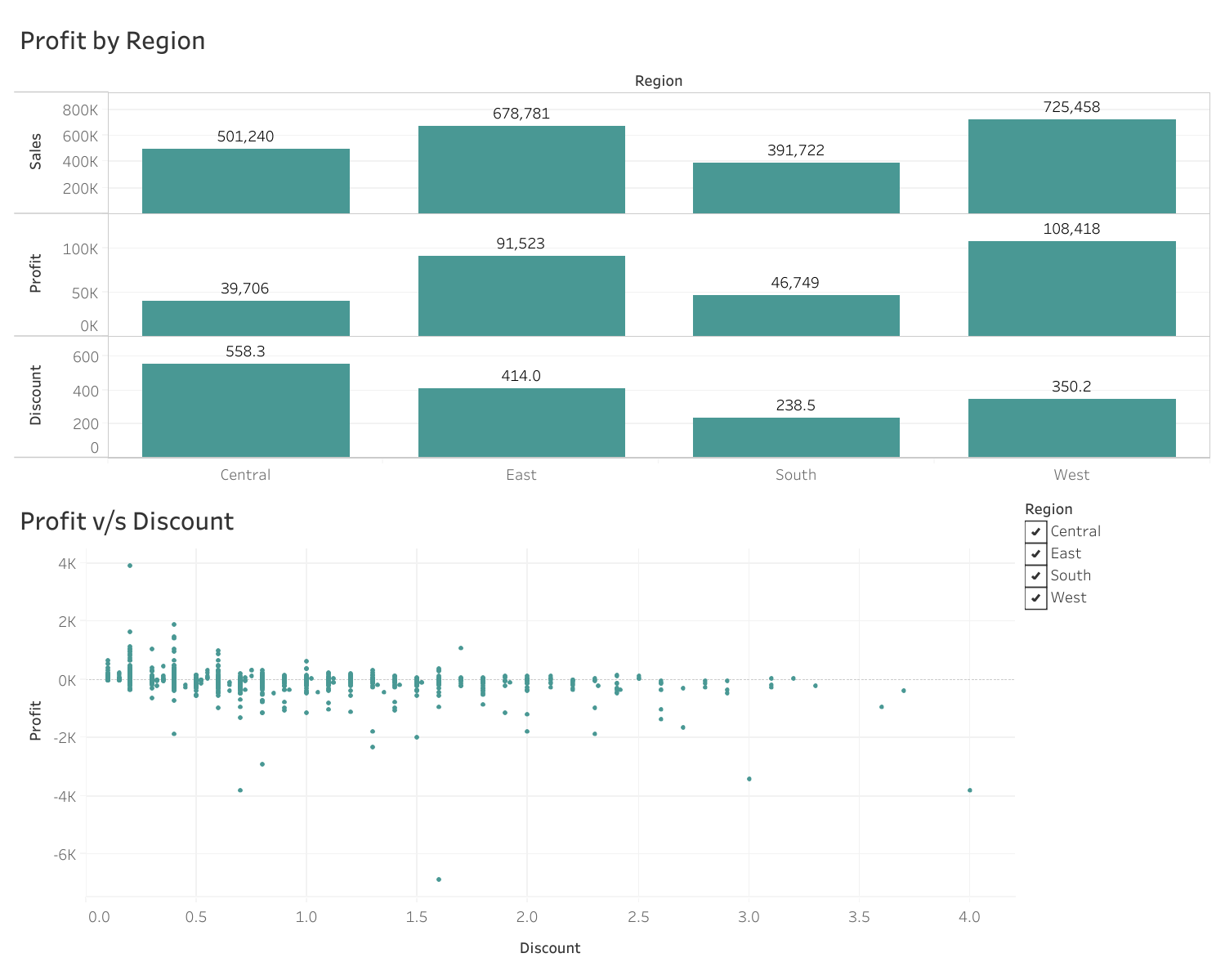
So far, no huge losses have occurred in 2016 and 2017. Also, the year 2017 occupies the most area in this area chart. **This shows Superstore has improved their profits over the years.** They should implement ways to maintain this streak.



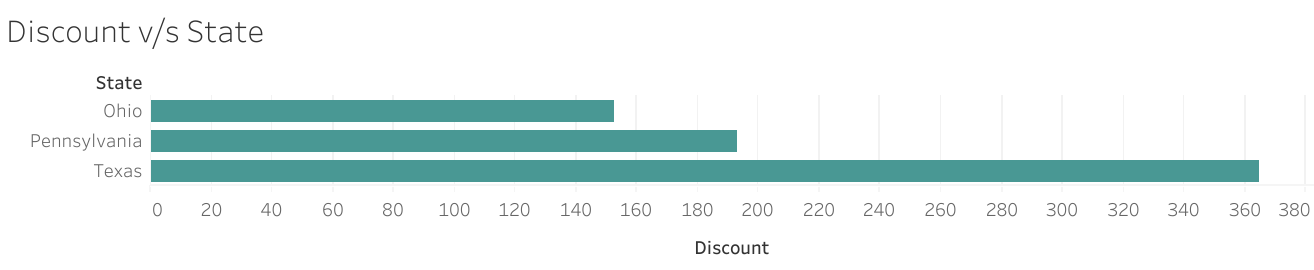
**Profit and Discount:**

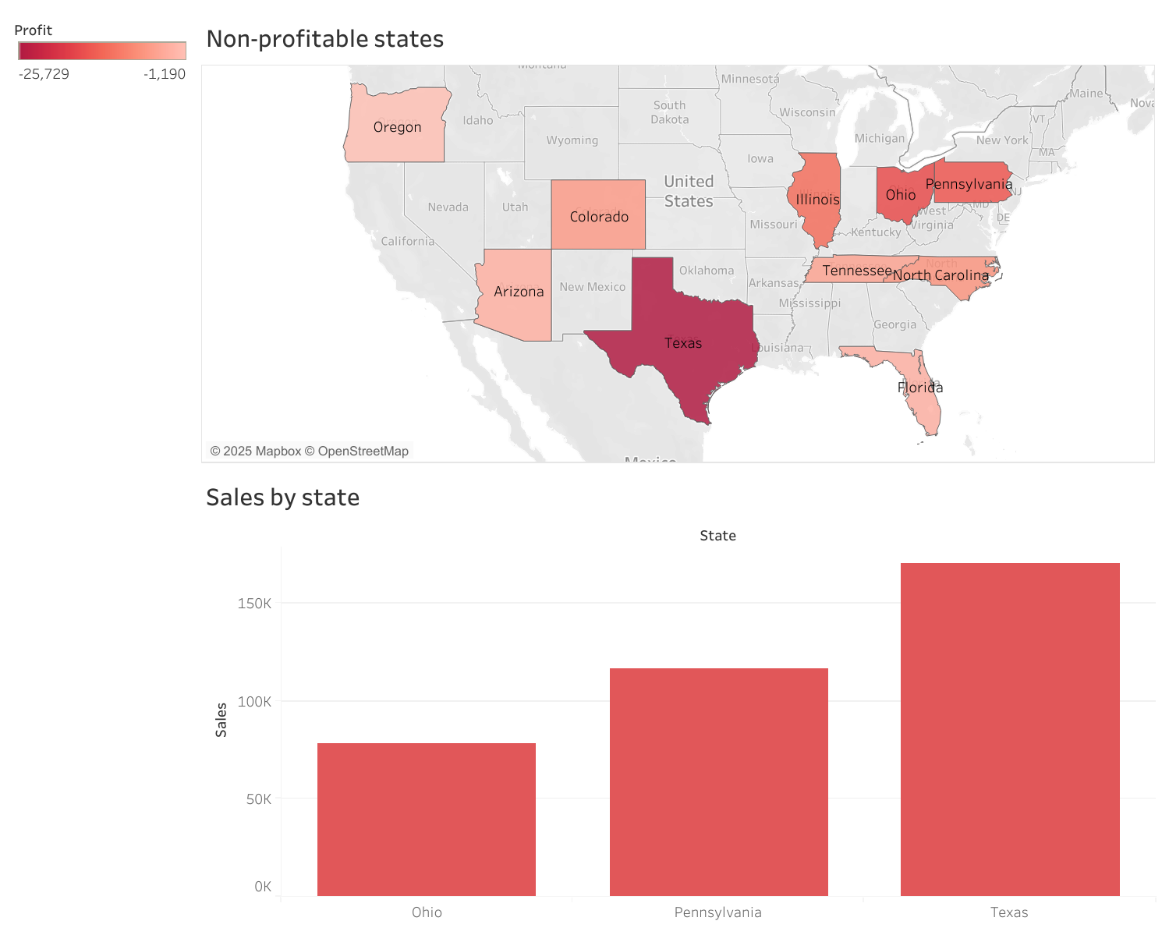
The stacked bar chart below shows us the region wise discount, profit and sales. The least profitable region is the central region followed by the south. But the sales for central region are higher than for the south. The reason behind this seems to be the discount. **More discount is offered in the central areas and hence the profit is so low, in spite of good number of sales.** The company should reduce the discount given in this region to maximize their profit because the current scheme is clearly not working in their favour.

To prove the above statement, we can refer to the scatter plot below. We can see that as discount increases; profit reduces and moves closer to the negative values. **Profit and discount have a negative correlation**. The team should come up with a smart and a better discount plan.

****

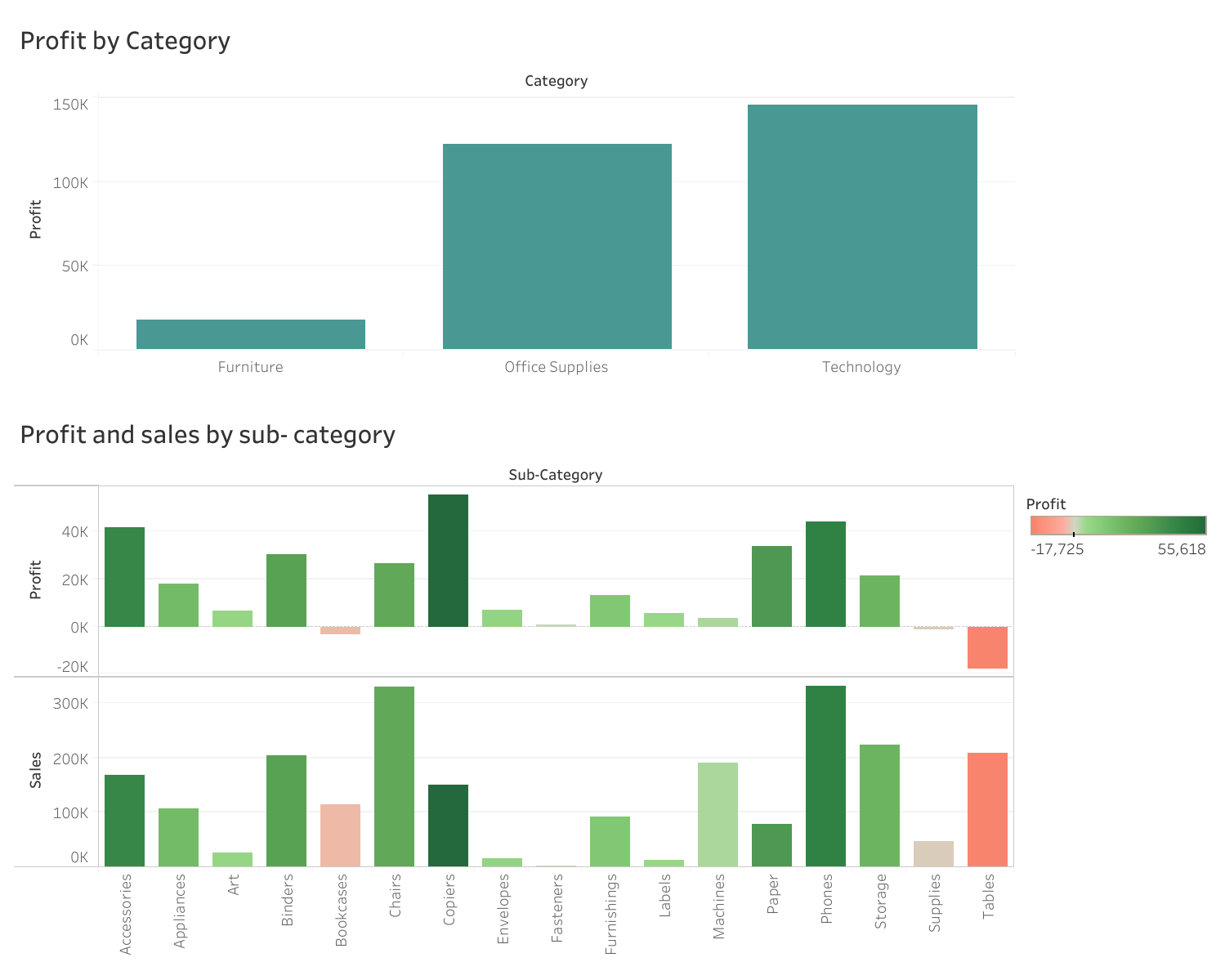
**Non-profitable States:**

Moving further into the discussion of profit. The map below shows us the least profitable states. The top 3 of them are Texas, Pennsylvania and Ohio. The next bar chart shows us the number of sales for these 3 states. **Texas has a high number of sales and yet it’s causing the most loss.** Before we saw that many customers are from Texas. Hence this is a major issue. This is a massively important topic to look into. To solve this, they can try reducing some number of stores in Texas, the stores which do not generate much profit. The green bar chart proves that the discount given in Texas is much higher. In fact, another bar chart showed us that Texas has the most discount opportunities among all the states. A change in the discount scheme is a requirement.



**Profit by Categories:**

The store sells products which are majorly divided into 3 broad categories, furniture, office supplies and technology. **The most profit is generated by technology products. The least profit is generated by furniture**. Diving deep into the categories, we see that the store sells chairs, tables, bookcases, etc. for furniture. According to the second chart, **chairs are very profitable, meanwhile the tables are responsible for a loss**. They should stop selling tables altogether, because despite many are sold, they are no good for the business. Same goes for the bookcases. Supplies are also responsible for a loss but it’s nothing major. They can decide to make changes to this sub-category or not.



**Findings and solutions:**

1. Most of the customers are from California.
2. Phones are the highest selling products.
3. February has the lowest profit for all the years. To change this, they can sell more February relevant items like Valentine’s Day products to drive more customers into the stores.
4. Years 2014 and 2015 observed a major loss for particular months. Past data should be collected and analyzed to understand the cause. This will prevent future losses.
5. Central area generates the lowest revenue because more discount is provided there. Since discount and profit are negatively correlated, the new discounts should be balanced to increase the revenue.
6. Texas has many customers but it offers no profit. Products are massively discounted in Texas. This should be changed to make the best use of the large number of sales. Another reason could be too many stores, some of which don’t experience much crowd.
7. Furniture is profitable only when it comes to chairs. Tables and bookcases do not help with generating profit. They should stop selling them or look for factors causing it. Technology is the most optimal source of income and should be more focused upon.
8. Overall, many modifications are required in their discount patterns.

Thank you.