

"Exploratory Data Analysis (EDA) Report." **The Sparks Foundation**

Author
Kartik Parlikar
(Business Analyst & data science intern)



Client: **"Superstore Manager"**
Date of Submission: 08 June 2023
Tool: Tableau

Internship Task

As a business manager, we need to find out the weak areas where we can work to make more profit. I am mentioning in detail some of our areas where we make a negative profit which is a loss.

Dear [Manager],

Thank you for providing us with the datasets from [Dataset](#). The below table highlights some details from the datasets received.

- The dataset has 9994 rows and 13 columns.
- the data being analyzed was already in good shape, meaning it did not have any missing or empty values (NULL).
- The data did not require any cleaning or fixing because it was already.

clean from the start. Additionally, there was no need to change the data types (type casting).

The only issue found was some duplicate values, which were removed. As a result, a new calculated field was created based on the data to perform additional calculations or analysis.



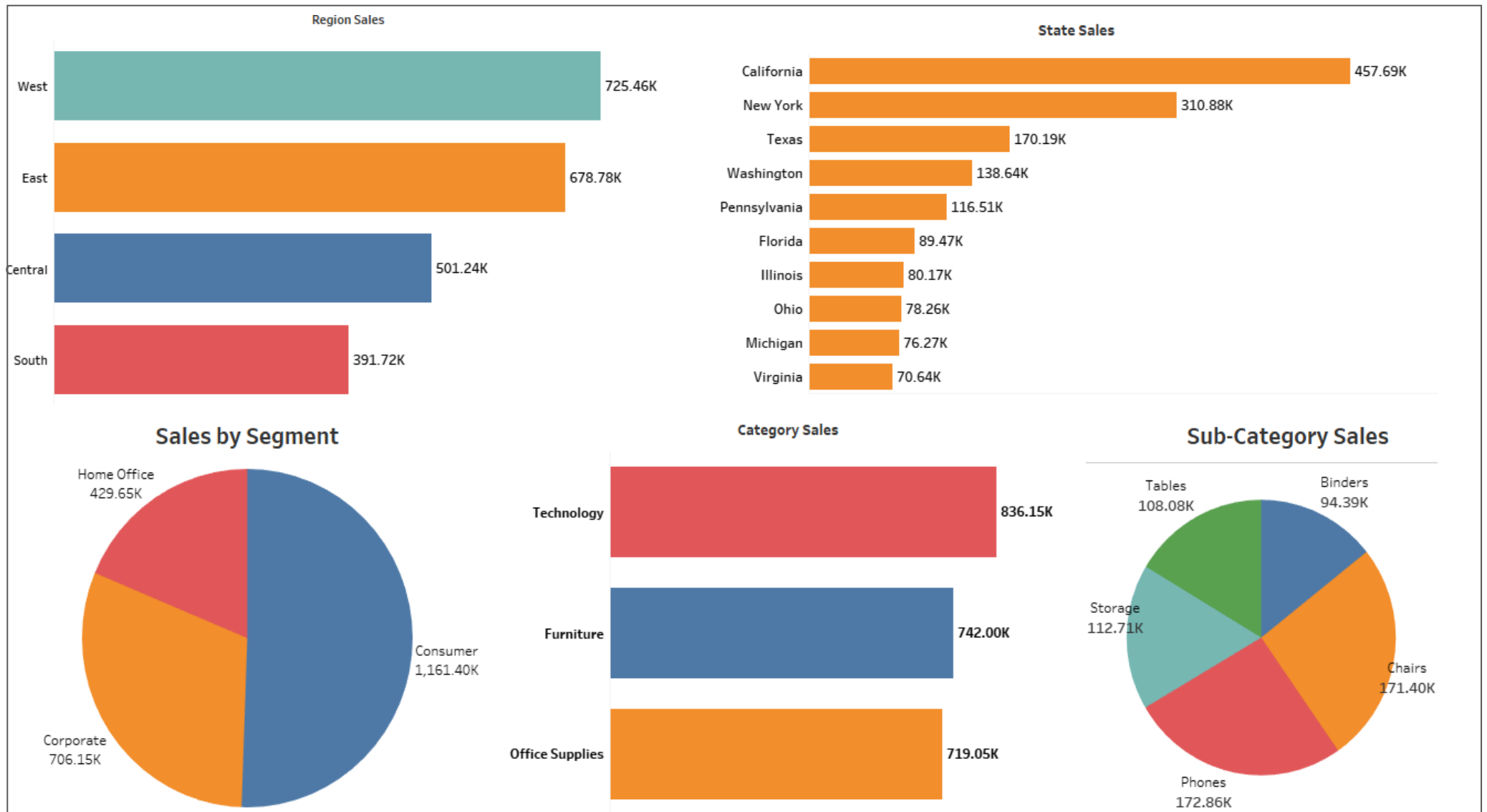
```
df.shape
```

```
(9994, 13)
```

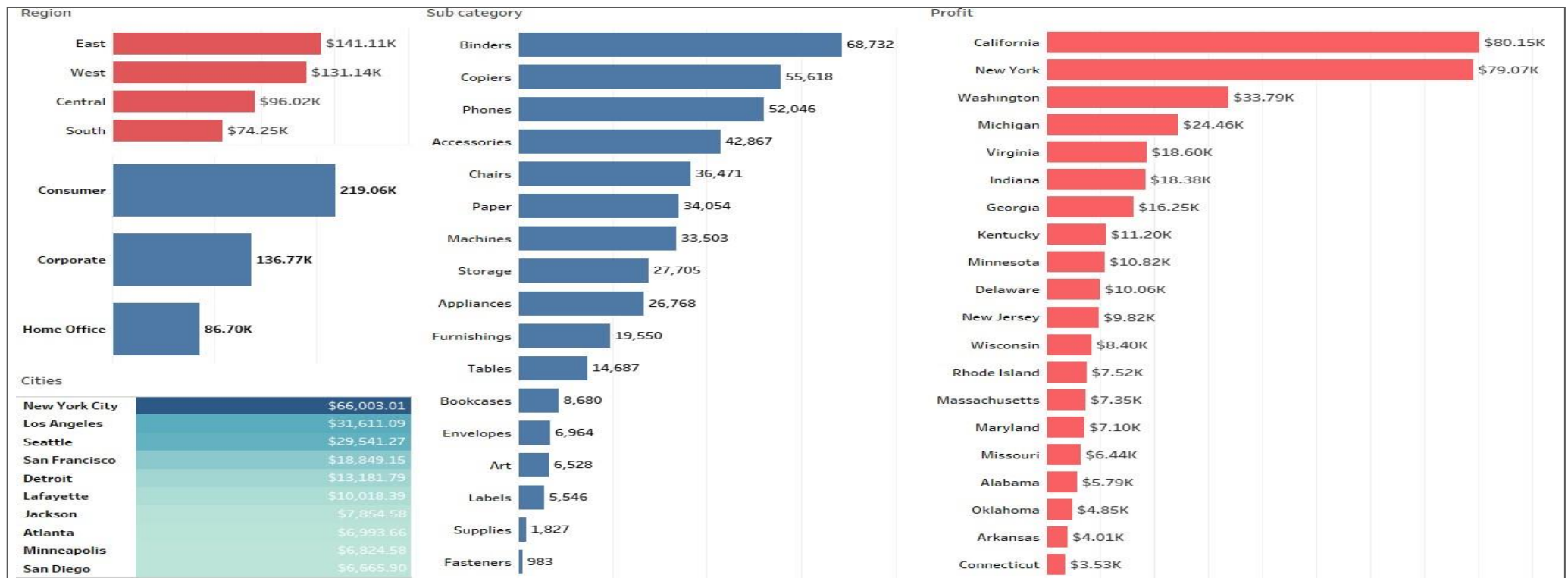


```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Ship Mode       9994 non-null   object
1   Segment         9994 non-null   object
2   Country         9994 non-null   object
3   City            9994 non-null   object
4   State           9994 non-null   object
5   Postal Code     9994 non-null   int64
6   Region          9994 non-null   object
7   Category        9994 non-null   object
8   Sub-Category    9994 non-null   object
9   Sales           9994 non-null   float64
10  Quantity        9994 non-null   int64
11  Discount        9994 non-null   float64
12  Profit          9994 non-null   float64
dtypes: float64(3), int64(2), object(8)
memory usage: 1015.1+ KB
```



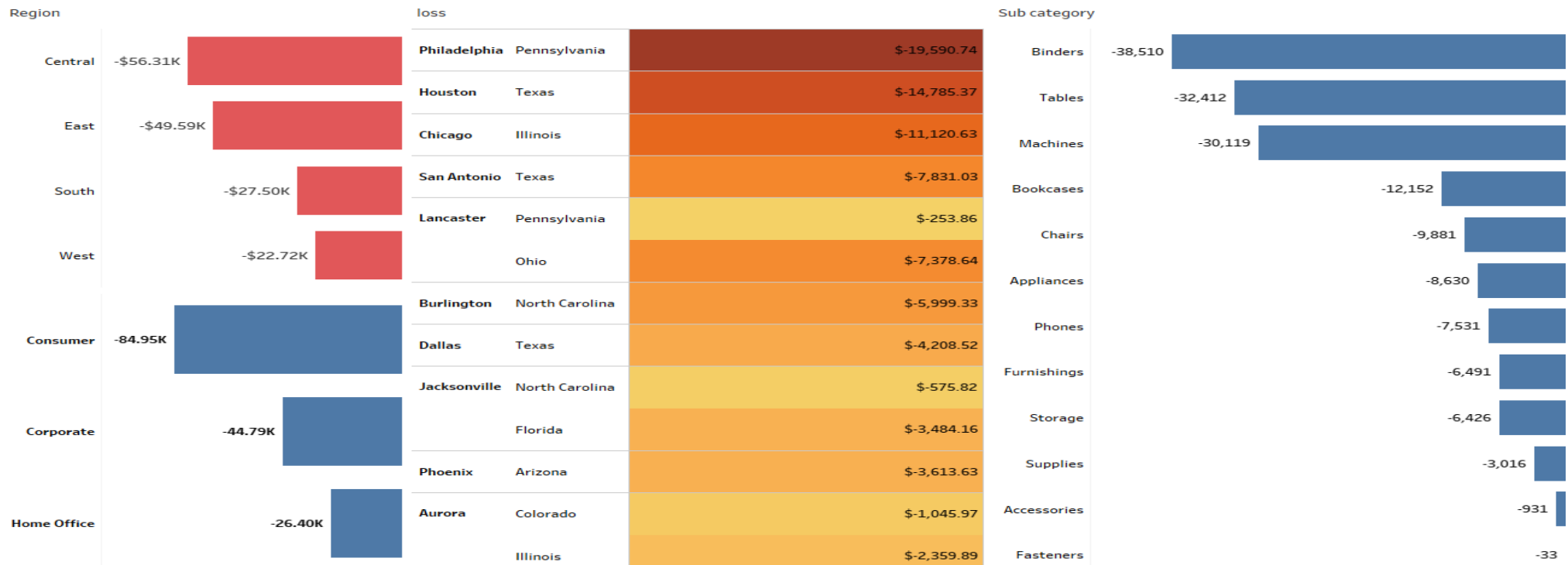
The states of California, New York, Texas, and Washington (which are in both the West and East regions) are generating the highest sales in the technology category. This refers to the amount of money earned from selling technology products. The sales are specifically in the standard class mode, which means they are not specialized or premium sales.



These are the 5 top cities that have given us the most Profit.

- The highest profit is being generated by customers from the Consumer segment in the states of California, New York, Washington, Michigan, and Virginia.
- The profit is specifically related to certain sub-categories of products, including Binders, Copier, Phone, Accessories, Chair, and Papers. These products are contributing significantly to the overall profitability in those regions and segments.

	State	Cities	profit
1.	New York	New York city	\$66000
2.	California	Los Angeles	\$31611
3.	Washington	Seattle	\$29541
4.	California	San Francisco	\$18849
5.	Michigan	Detroit	\$13182



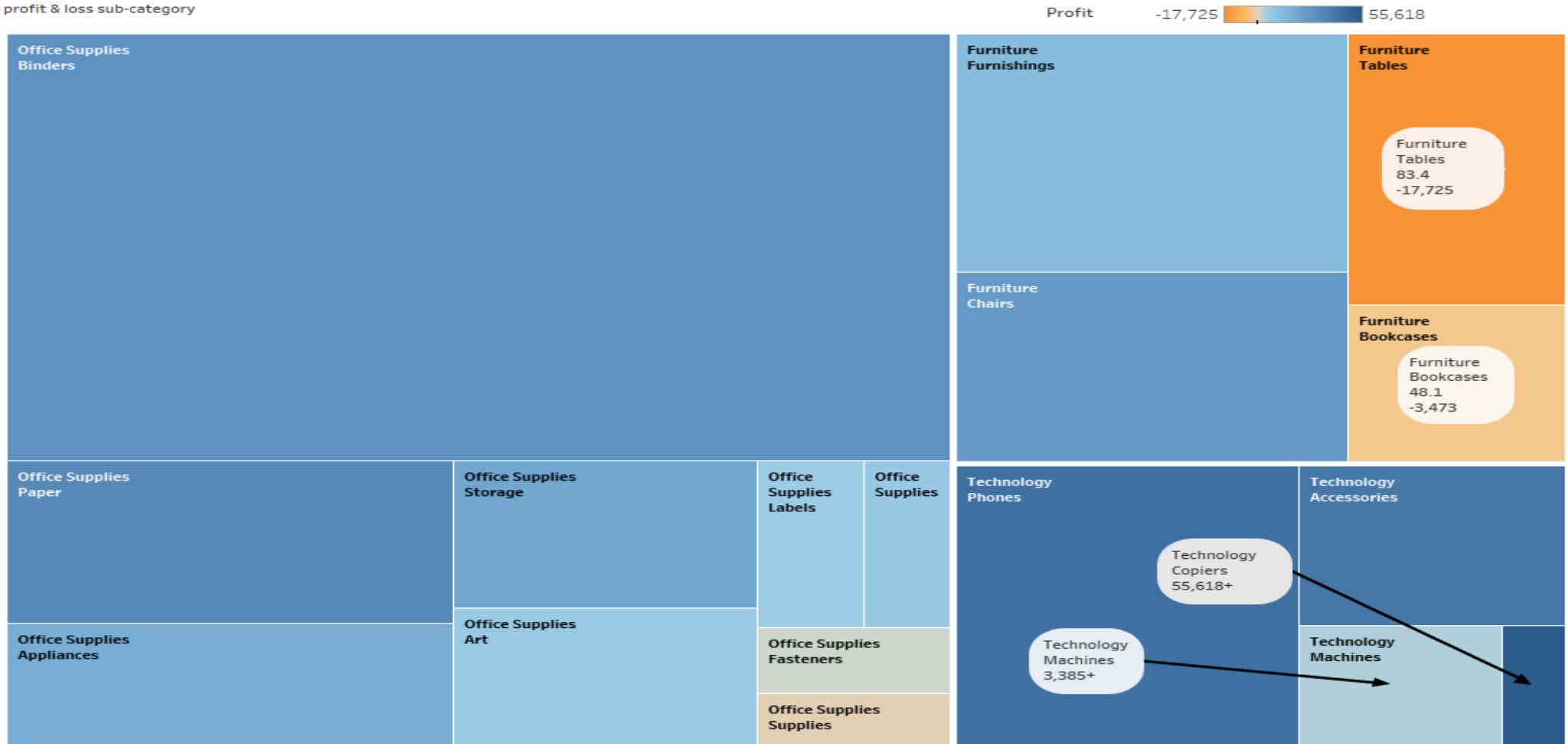
- Texas and Illinois in the central region, Ohio and Pennsylvania in the east, North Carolina, Florida, and Tennessee in the south, and Colorado in the west region are the states where the most losses are occurring.
- This means that these states are experiencing financial losses or negative profitability compared to other states.

Loss By Cities

These are the 6 top cities that have given us the most losses.

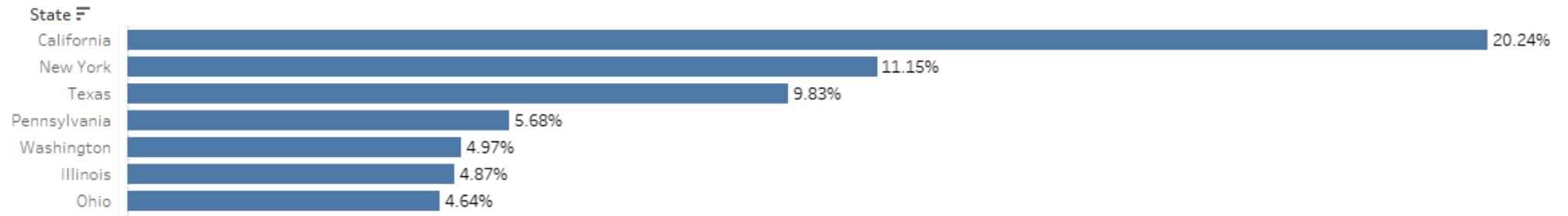
State	Cities	loss
1. Pennsylvania	Philadelphia	\$19591
2. Texas	(Houston, San Antonio, Dallas)	\$26824
3. Illinois	Chicago	\$11120
4. Ohio	Lancaster	\$7378
5. North Corallina	Burlington	\$6000

profit & loss sub-category



1. There is a loss of \$17,725 in Furniture category, meaning the company is not making a profit from selling furniture.
2. The total discount given on the furniture is \$83, but this discount does not have an impact on the overall loss. The reason for the loss is selling the furniture at a price that is lower than the cost, resulting in a loss for the company.

state wise percentage of total order



The percentages of orders from each state:

- California: 20.24%
- New York: 11.15%
- Texas: 10%
- Florida: 3.64%
- Washington: 5%
- Pennsylvania: 5.68%

This means that California has the highest percentage of orders at 20.24%, followed by New York at 11.15%. These percentages represent the portion of total orders that come from each state.

Conclusion:

1. There is a lot of potential to increase profitability in some areas. In the technology category, California, New York, Texas, and Washington stand out as the regions with the highest sales, with phones being the best-selling items.
2. Additionally, the consumer segment in states such as California, New York, Washington, Michigan, and Virginia is making profits in sub-categories such as blinders, copiers, phones, accessories, chairs, and paper.
3. Tables, bookcases, and supplies are giving us net loss, so for this we need to make a strategy so that we can make profit.
4. Texas and Illinois in the central region, Ohio and Pennsylvania in the east region and North Carolina, Florida, Tennessee in the south region, along with Colorado in the west region are experiencing significant losses. It is crucial to address these weak areas and derive strategies to improve profitability and reduce losses.

Click here for a deeper understanding of data visualization. 

Kind regards,
[Kartik Parlikar]

Thank you
