Exploratory Data Analysis – Retail

❖ Goal:

To identify various patterns in the data set given and suggest or take necessary measures after analysing the data set. Also, to find out weak areas where we can work to earn more profits.

❖ About:

→ Data set URL https://bit.ly/3i4rbWl
→ Rows are 9994 and columns are 13. And they are as follows:

```
$ Ship.Mode  : chr  "Second Class" "Second Class" "Second Class" "Standard Class" ...
 $ Segment    : chr  "Consumer" "Consumer" "Corporate" "Consumer" ...
 $ Country    : chr  "United States" "United States" "United States" "United States" ...
 $ City       : chr  "Henderson" "Henderson" "Los Angeles" "Fort Lauderdale" ...
 $ State      : chr  "Kentucky" "Kentucky" "California" "Florida" ...
 $ Postal.Code : int 42420 42420 90036 33311 33311 90032 90032 90032 90032 90032 ...
 $ Region     : chr  "South" "South" "West" "South" ...
 $ Category   : chr  "Furniture" "Furniture" "Office Supplies" "Furniture" ...
 $ Sub.Category: chr  "Bookcases" "Chairs" "Labels" "Tables" ...
 $ Sales      : num 262 731.9 14.6 957.6 22.4 ...
 $ Quantity   : int 2 3 2 5 2 7 4 6 3 5 ...
 $ Discount   : num  0 0 0 0.45 0.2 0 0 0.2 0.2 0 ...
 $ Profit     : num  41.91 219.58 6.87 -383.03 2.52 ...
```
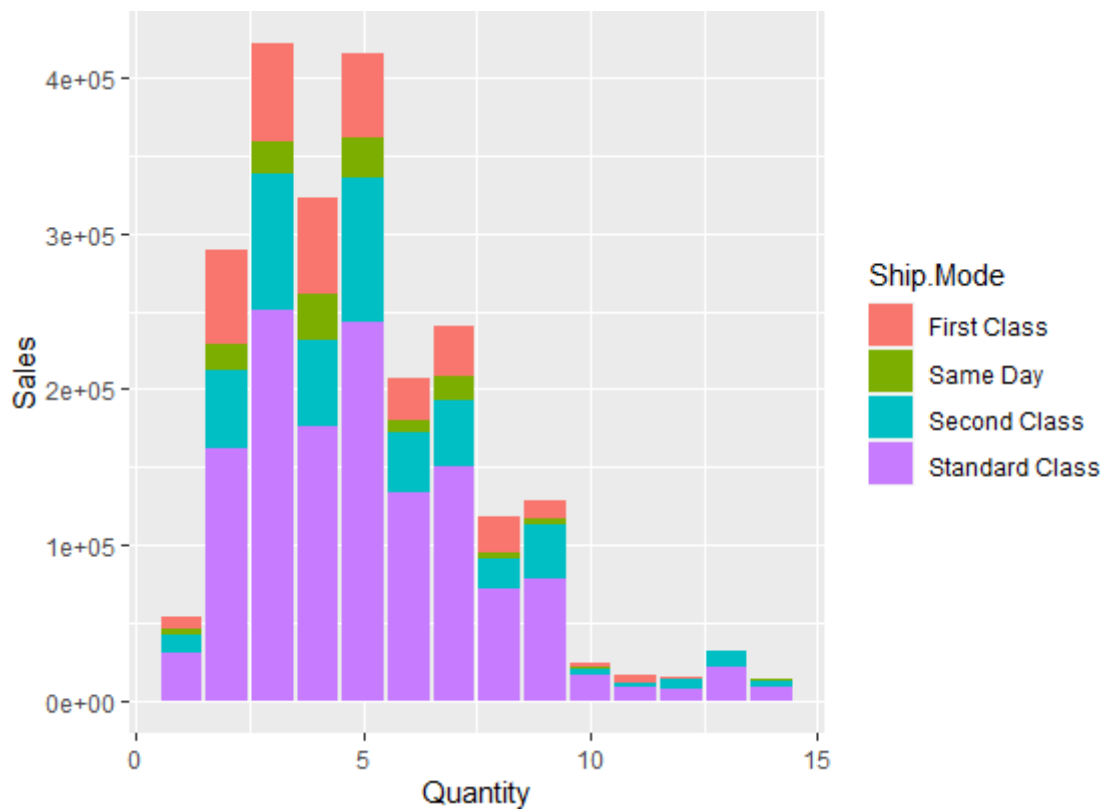
→ There are no missing values
→ Country – United States
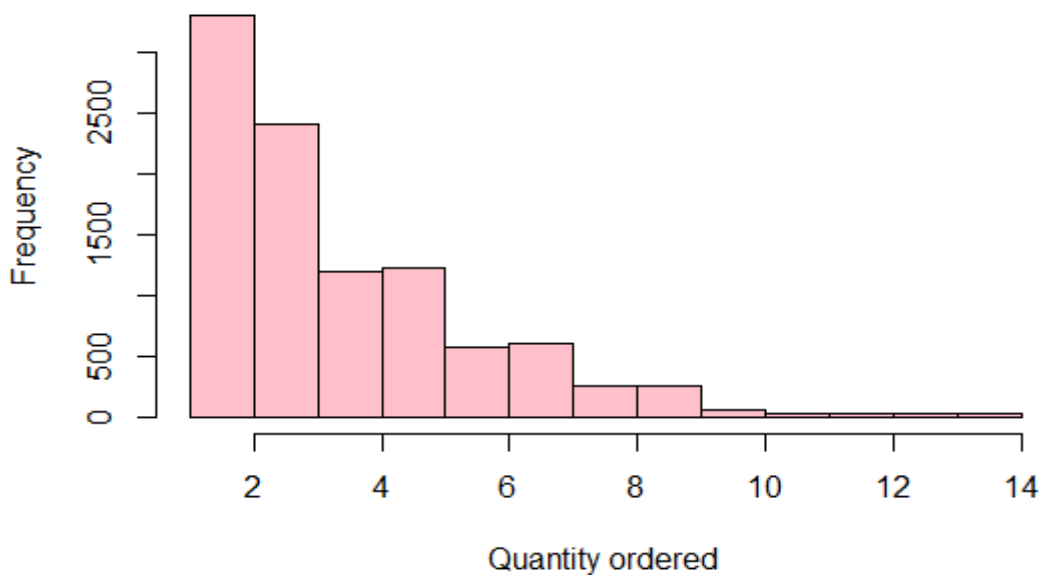→ Removed 'Country' and 'Postal Code' column

Analysing the patterns

1. Sales and Quantity within the Shipment mode

From the above graph we can analyse that the shipment mode "Standard Class" has more number of sales than the other three.
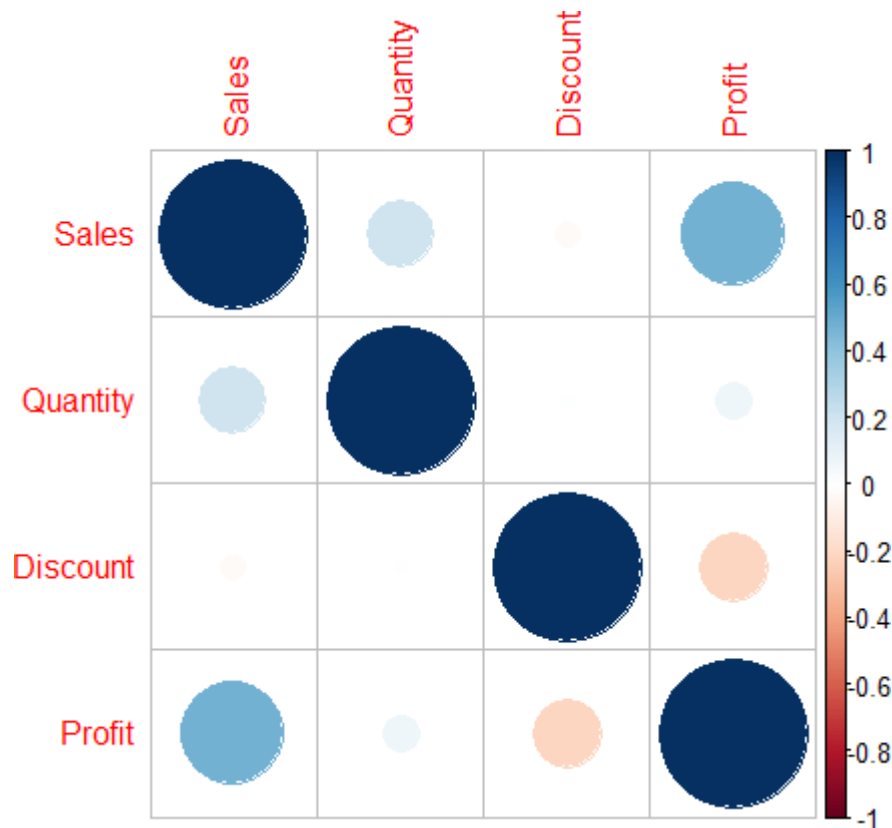
2. Frequency distribution of quantity ordered


The maximum ordered quantity is 1 which is 300 as we can see from the above histogram.

3. Correlation

This is the correlation diagram which shows red colour as negative correlation and blue as positive. The variables are sales, quantity, discount, profit. Sales and Profit have high positive correlation whereas Discounts and Profits have negative.

4. Sales and Discounts within Shipment mode
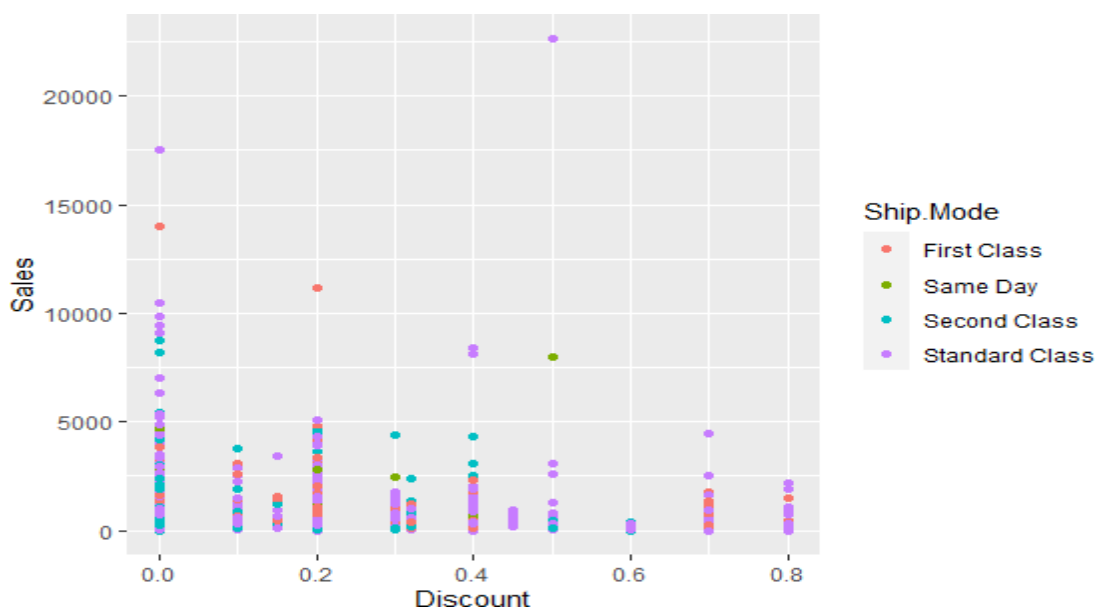
Even with no discounts, sales are high from the standard class shipment mode. As the discounts increase, we can see that sales of standard shipment mode are still high.

5. Profits and Discounts within Shipment mode
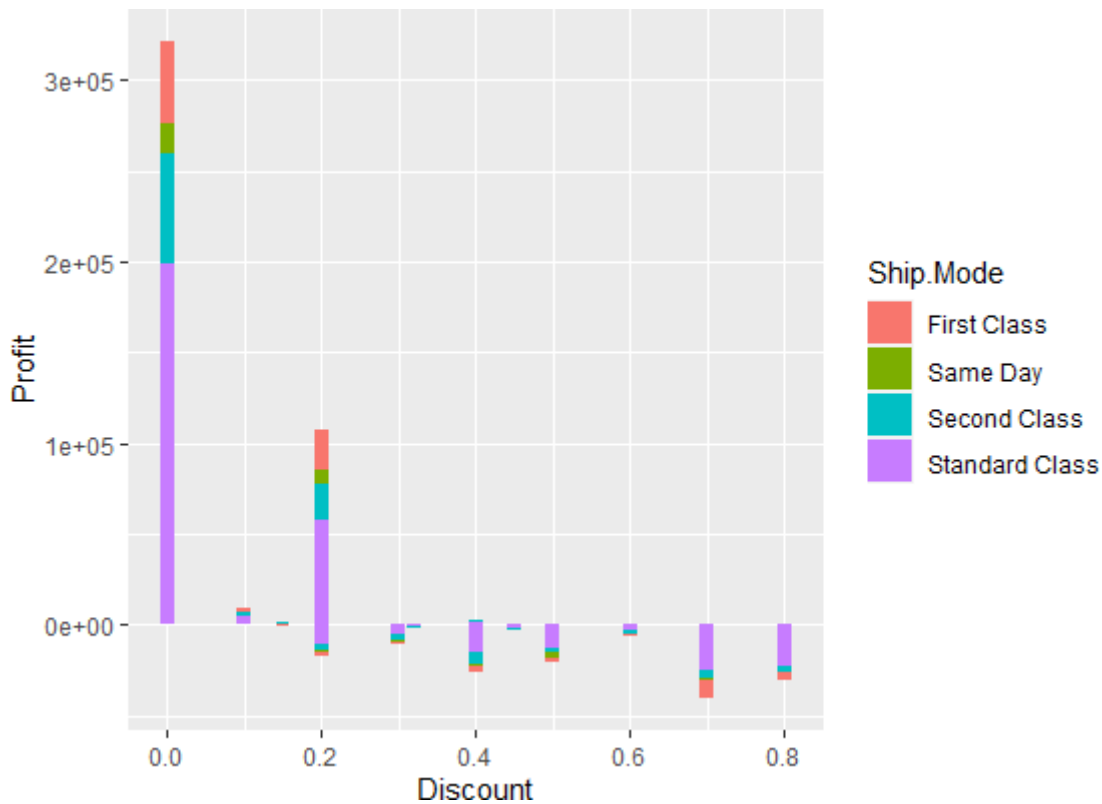
With the increase in discounts, profits are decreasing as we can see from here. Products from different shipment mode especially Standard class and first class are incurring losses with the increase in discounts.
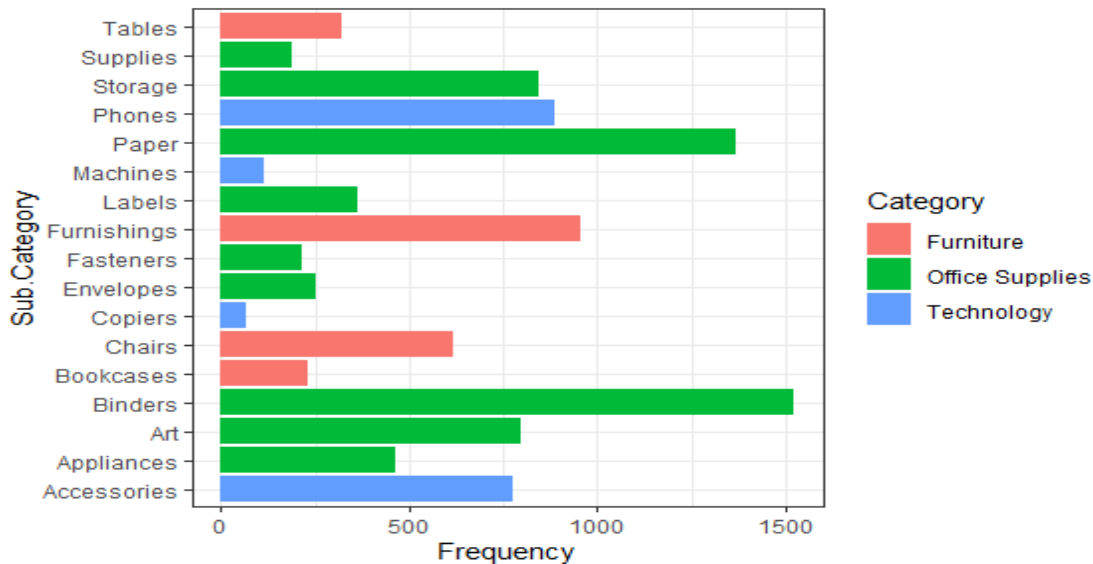
6. Profits within the sub category and region
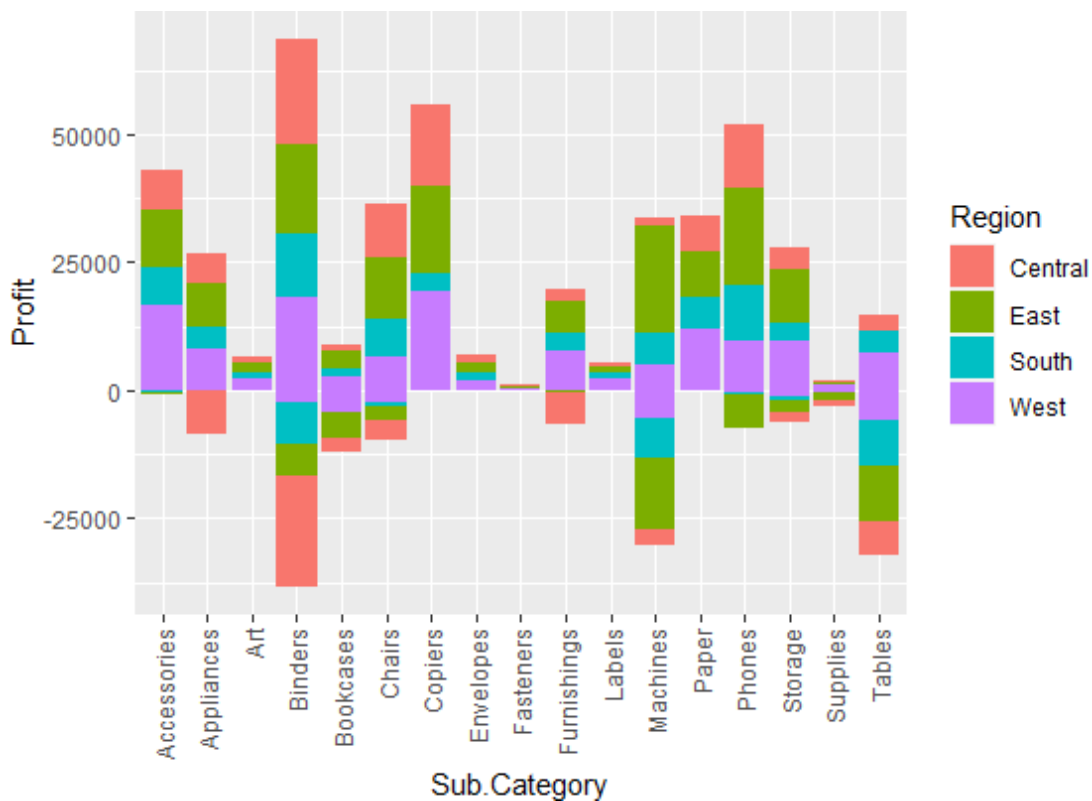
Here we have distributed region and subcategory with the profits. We can see here that the profits are highest in Binders Sub category and lowest in fastners. But we can also see that losses are also high in Binders.

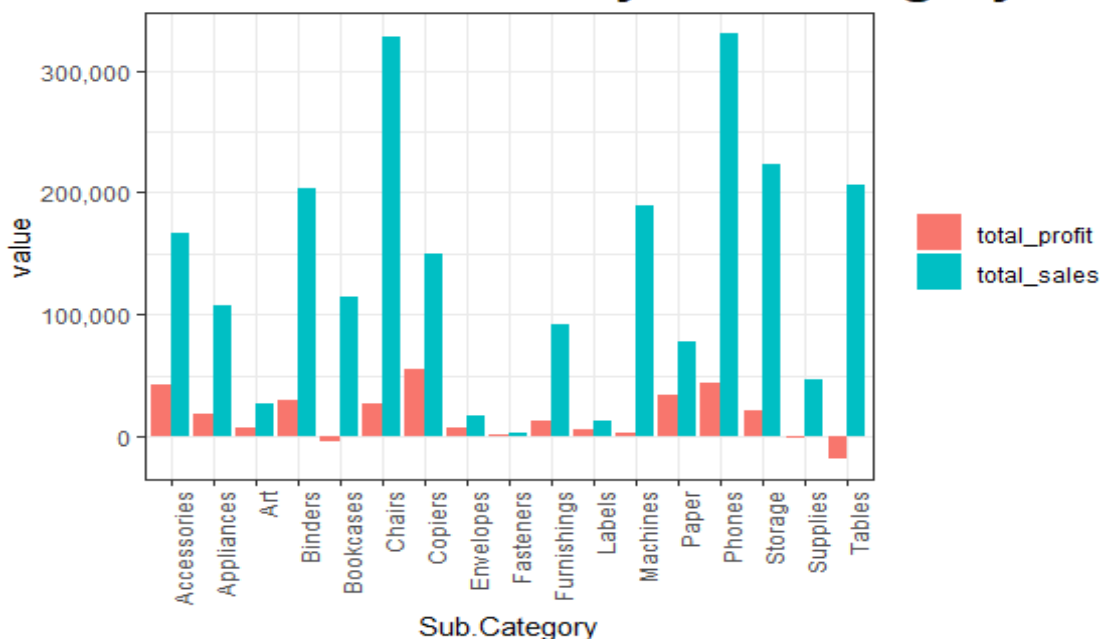7. **Frequency of sub category within the category**

There are 3 categories. Furniture, Office supplies and Technology. The frequency of office supplies is high, and Furniture is low.

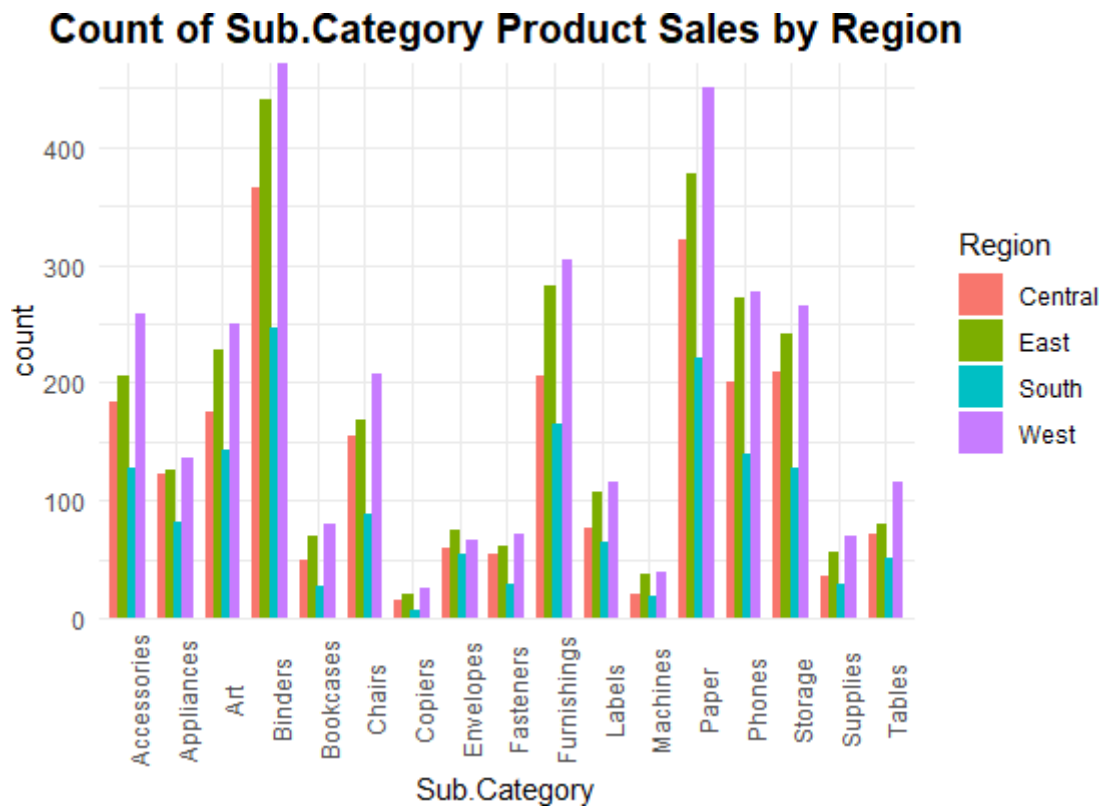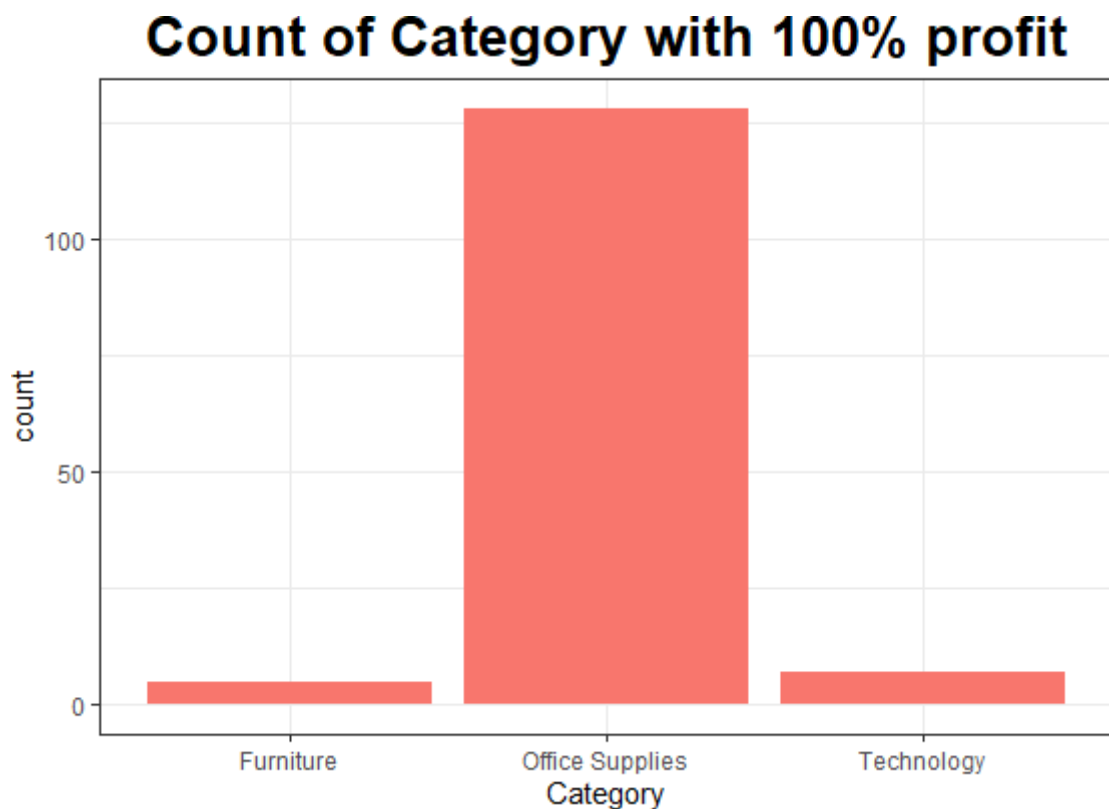8. **Profits and Sales**

Here we can look at the total profits and sales in the sub category. Chairs and Phones have the highest sales whereas Fastners have the lowest sales. The total profit if copiers is high whereas tables are incurring heavy losses.
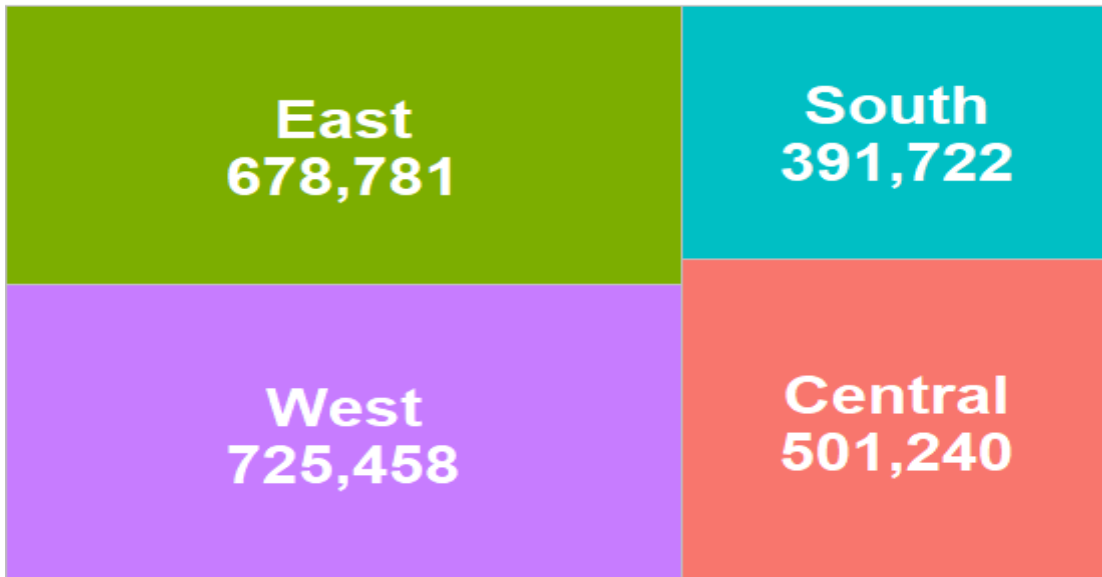
9. Sub category sales within the region

Count of Sub.Category Product Sales by Region

Region having the highest sales is West region in which Binders and paper are selling more.

10. Category with 100% profits
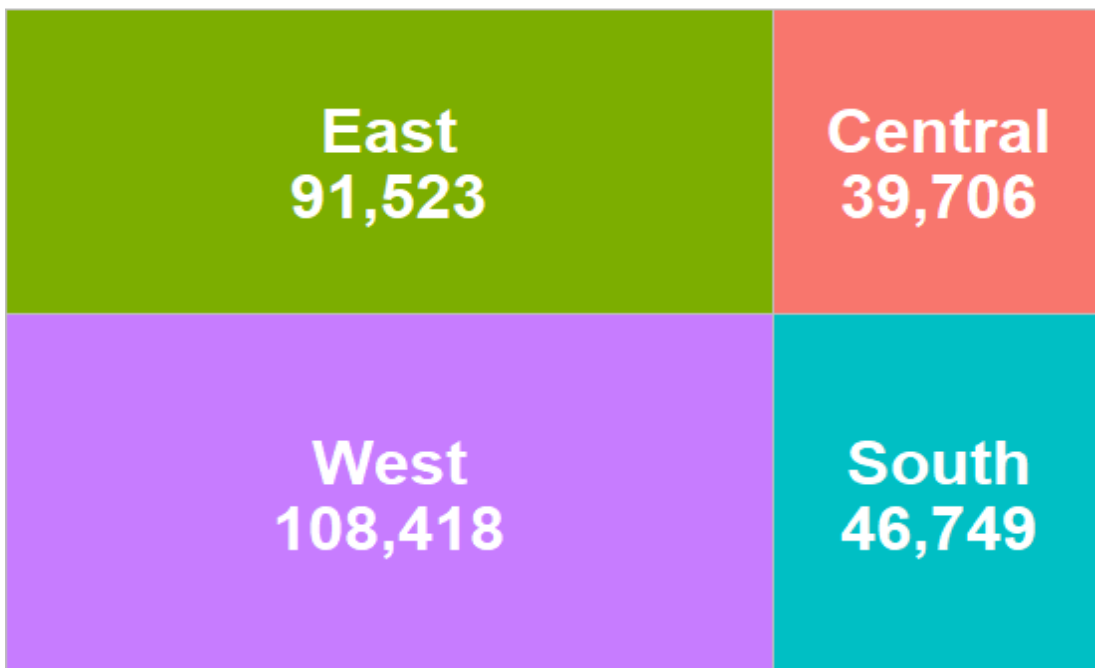

Count of Category with 100% profit

We can clearly see from here that Office supplies have 100% profits. And the category of Furniture as well as Technology should be focused upon to make more profits.

Sales by Region

East 678,781	**South** 391,722
West 725,458	**Central** 501,240

Profit by Region

East 91,523	**Central** 39,706
West 108,418	**South** 46,749

❖ Measures:

→ Same day shipment mode should be focused upon since the sales are low there if we go with the patterns

→ Should try to decrease giving discounts and should focus on sales of first class and same day shipment mode

→ Profits as well as losses both are high in Binders sub category which should be focused upon

→ In Furniture category, Tables are incurring losses which should be considered

→ South region is the weakest in terms of sales and that region should be focused upon as well to increase sales