Variables & Report Structure:

After the Data cleaning process for superstore fictional data, I have chosen 3 quantitative & 3 qualitative variables to analyze, interpret the visuals & draw conclusions.

The 6 chosen variables were:

1. Ship Mode
2. Customer Segment
3. Product Category
4. Profit
5. Sales
6. Quantity ordered

The report will consist of 3 parts:

1. Analyzing qualitative variables
2. Analyzing quantitative variables
3. A combination of both qualitative and quantitative

1. Analyzing qualitative variables:

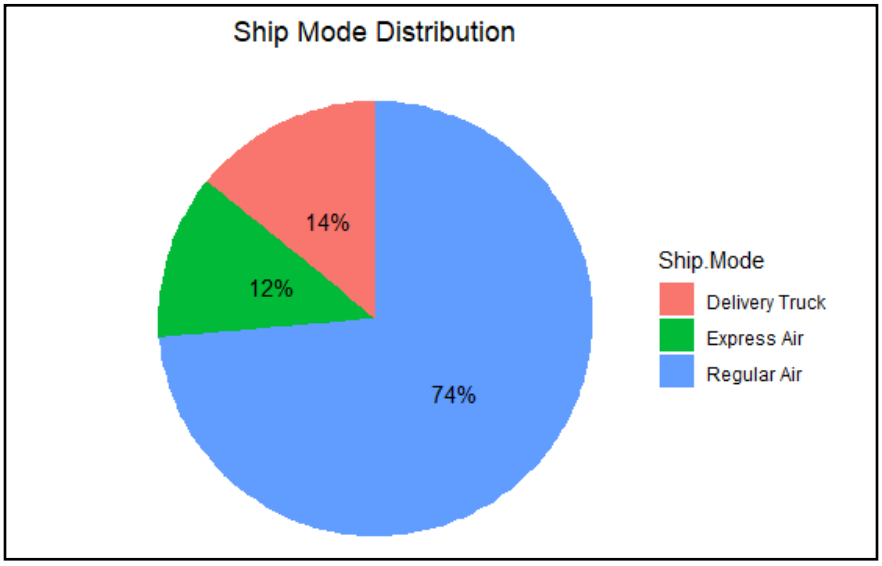

Figure 1:Ship Mode Pie chart

Most of our sales are done using the 'Regular air' ship mode with around 74% of the orders done through 'Regular air' while the 'Delivery truck' and 'Express Air' come in the second and third place respectively.

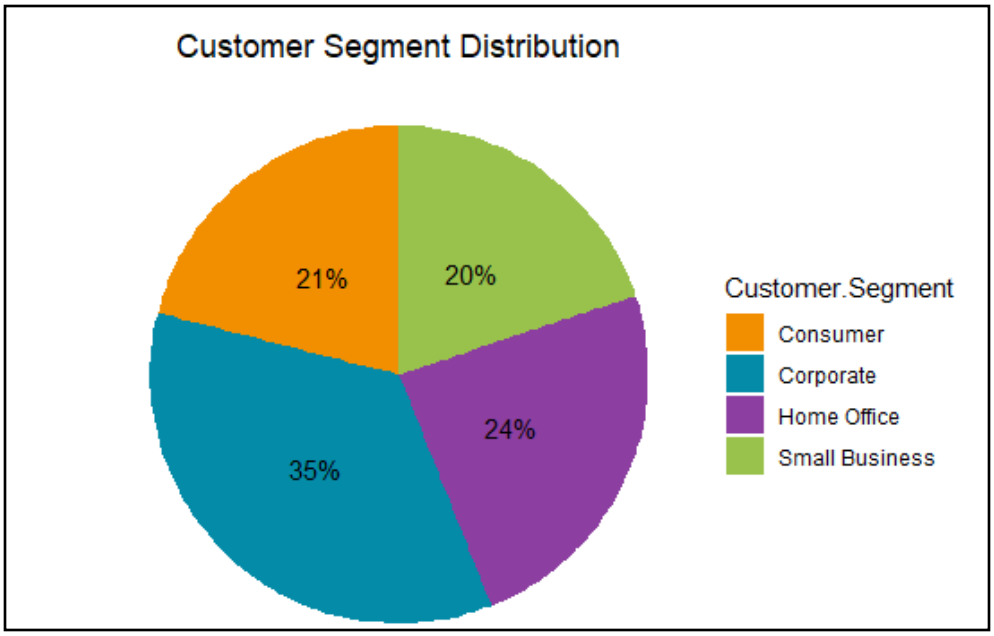

Figure 2:Customer Segment pie chart

The pie chart shows that around 35% of our customers are from corporates followed by the home offices, consumers & small businesses.

We can infer from the plot that most of our customers are categorized as corporates or home offices(with around 59% of the customers) where both fall into the category that involves paper work & using office supplies.

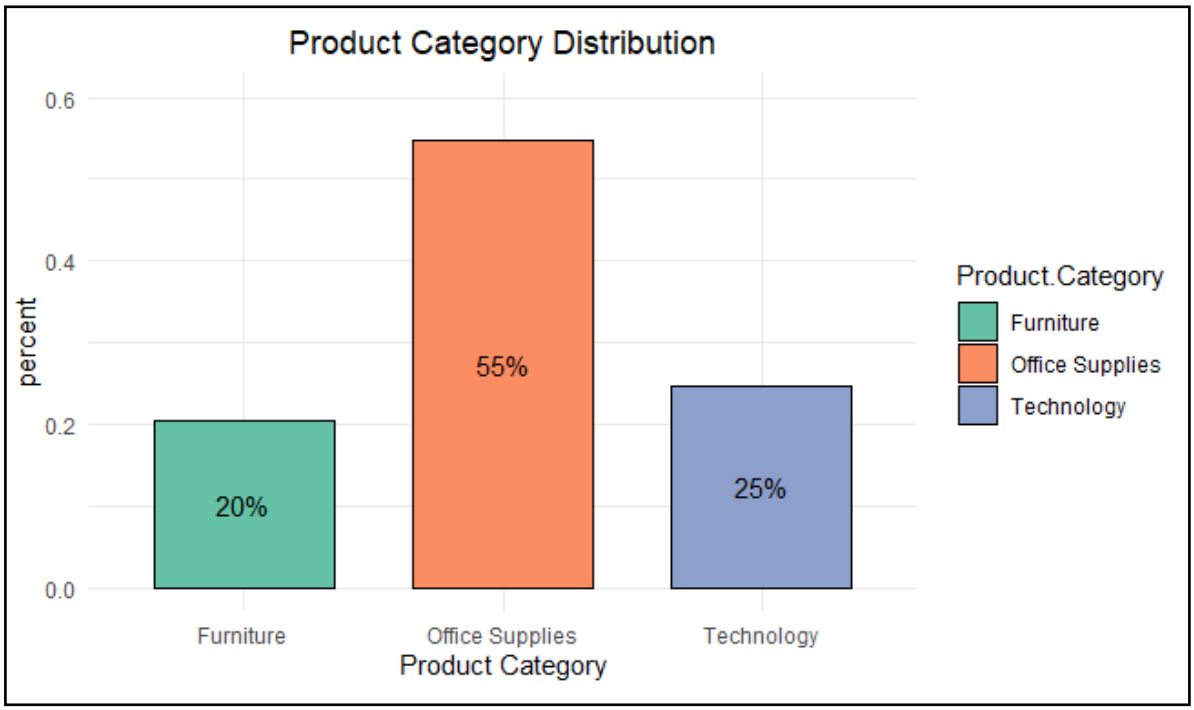

Figure 3: Product Category Bar chart

Most of our sold products are office supplies representing 55% of the total sales with 25% for the Technology category followed by 20% for furniture.

The bar chart above complements what have we said before in figure 2, that around 59% of the customers are home offices and corporates which is reflected on the product category since most of our sales are office supplies with 55%.

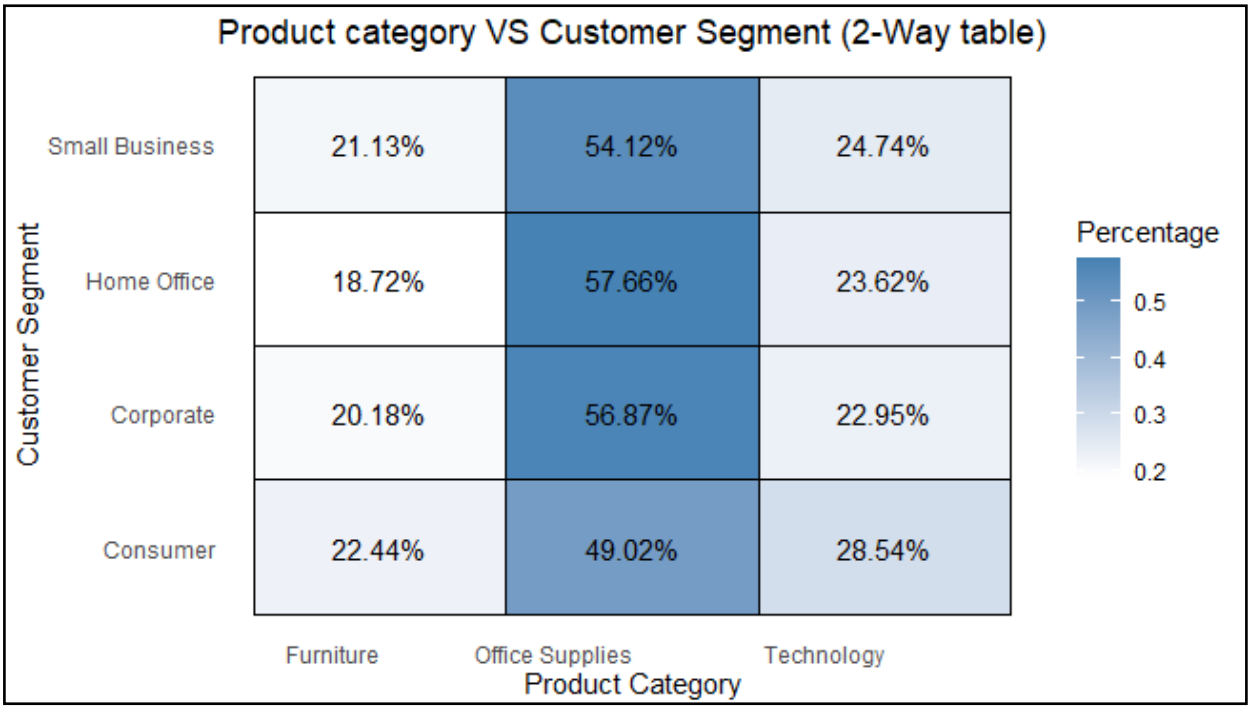

Table 1: Product category Distribution among customer segments

The table above shows the percentage of the 3 product categories purchased by each of the 4 customer segments, we can say that office supplies are the top seller products for the store since its percentage is greater than 50% for **almost** all customer segments followed by technology category which is the second most bought product category, then comes in last the Furniture category

2. Analyzing quantitative variables:

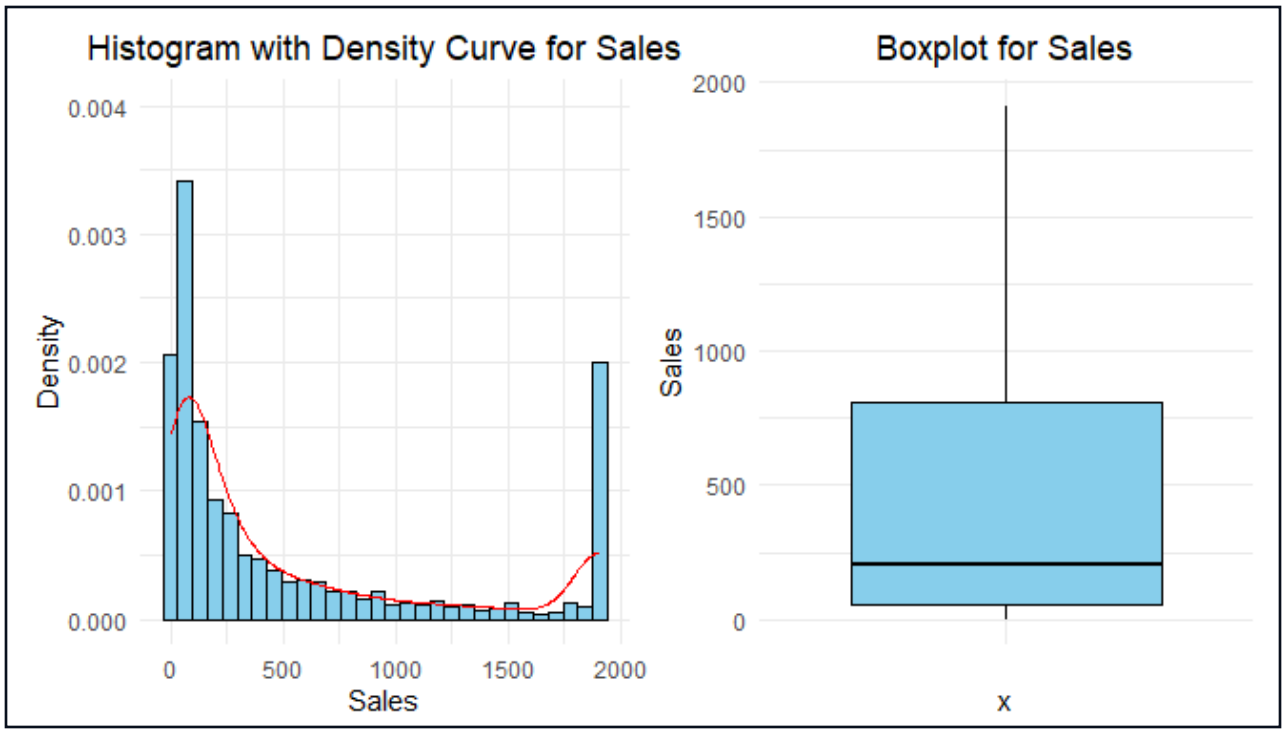

Figure 4: Sales Histogram & Boxplot

Min	Q1	Q2	Q3	max	mean	Std.dev	skewness
2.25	58.89	203.01	804.72	1913.84	543.46	664.97	1.1994

Table 2: Sales summary Statistics

The histogram shows right skewness to the distribution of sales which also goes along with the box plot, that means most of the sales happened on low values, around 50% of the sales were below 203.01$ (assuming sales are in dollars) while 25% of our sales were below 58.89$ & 75% of the sales were below 807.72$, the average sales was 543.46$.

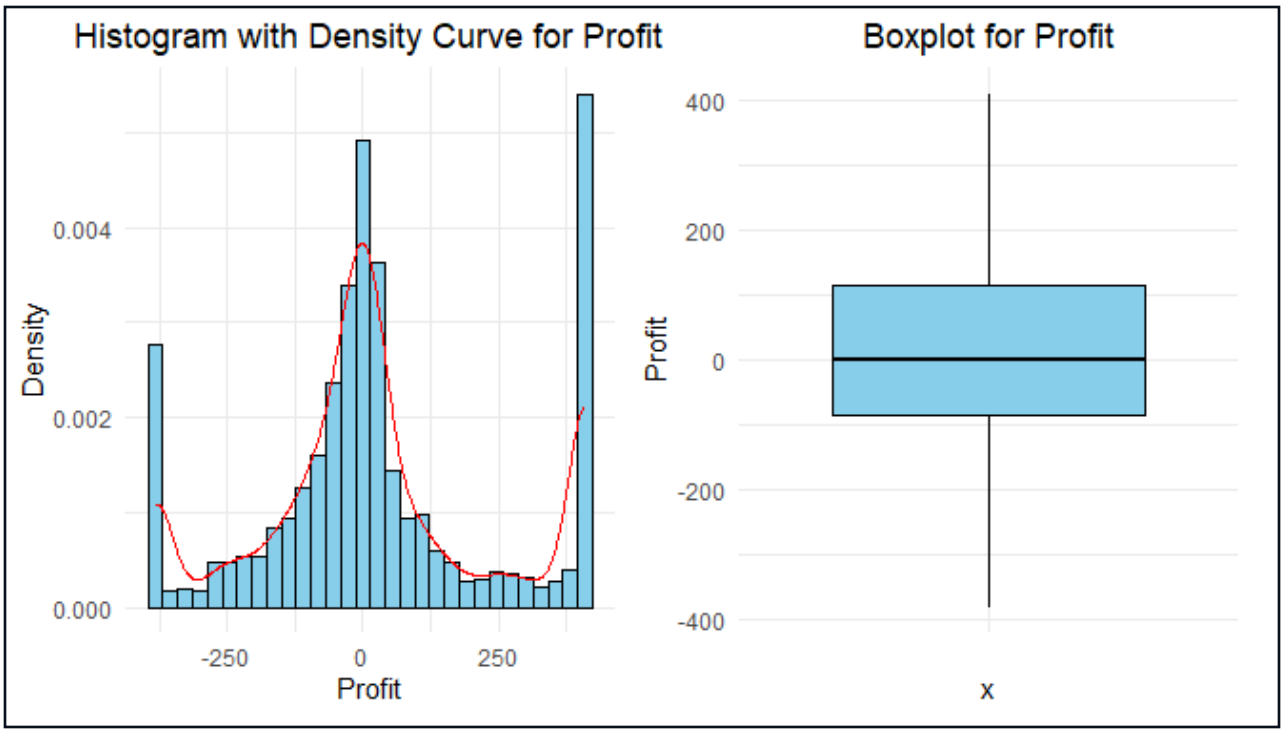

Figure 5: Profit Histogram & Boxplot

Min	Q1	Q2	Q3	max	mean	Std.dev	skewness
-381.84	-84.53	1.48	116.30	411.52	27.71	221.90	0.23005

Table 3: Profit summary Statistics

The histogram shows almost a symmetric distribution around the Zero(skewness is almost 0) which also goes along with the boxplot in the right, we can say that the profit for 50% of the total sales was below 1.48$ while the profit for 75% of the total sales was below 116.30$ but we didn't profit from 25% of the total sales since the first quartile was -381.84 (negative profit) which means that we lost from the 25% of sales, while the maximum profit was 411.52$ compared to the biggest loss from sales which is 381.84$(minimum of profit).

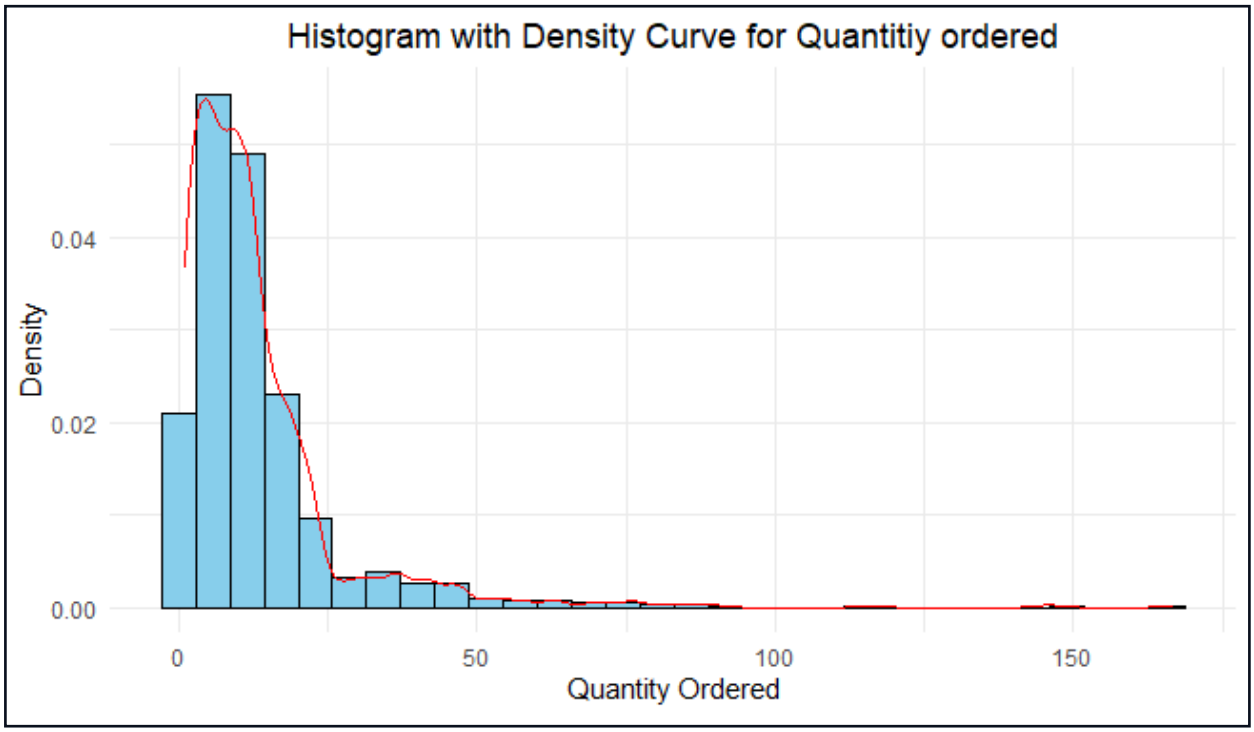

Figure 6: Quantity ordered Histogram.

Quantity Ordered	Percentage of total Quantity ordered
0 – 10	54.3%
10 – 20	30.7%
20 – 30	7%
30 – 40	3.2%
40 – 50	2.5%
More than 50	1.6%

Table 4: Quantity ordered Frequency.

Min	Q1	Q2	Q3	max	mean	Std.dev	skewness
1	5	10	16	167	12.93	13.86	3.69

Table 5: Quantity ordered summary Statistics

The histogram shows right skewness where most of the quantities ordered happening in low values which is represented in the percentage table where the quantity (0 – 10) form 54.3% of the total quantities ordered followed by quantity (10 – 20) which form 30.7% of the total quantities ordered, for quantity orders that is more than 50 quantities they only form 1.6%.

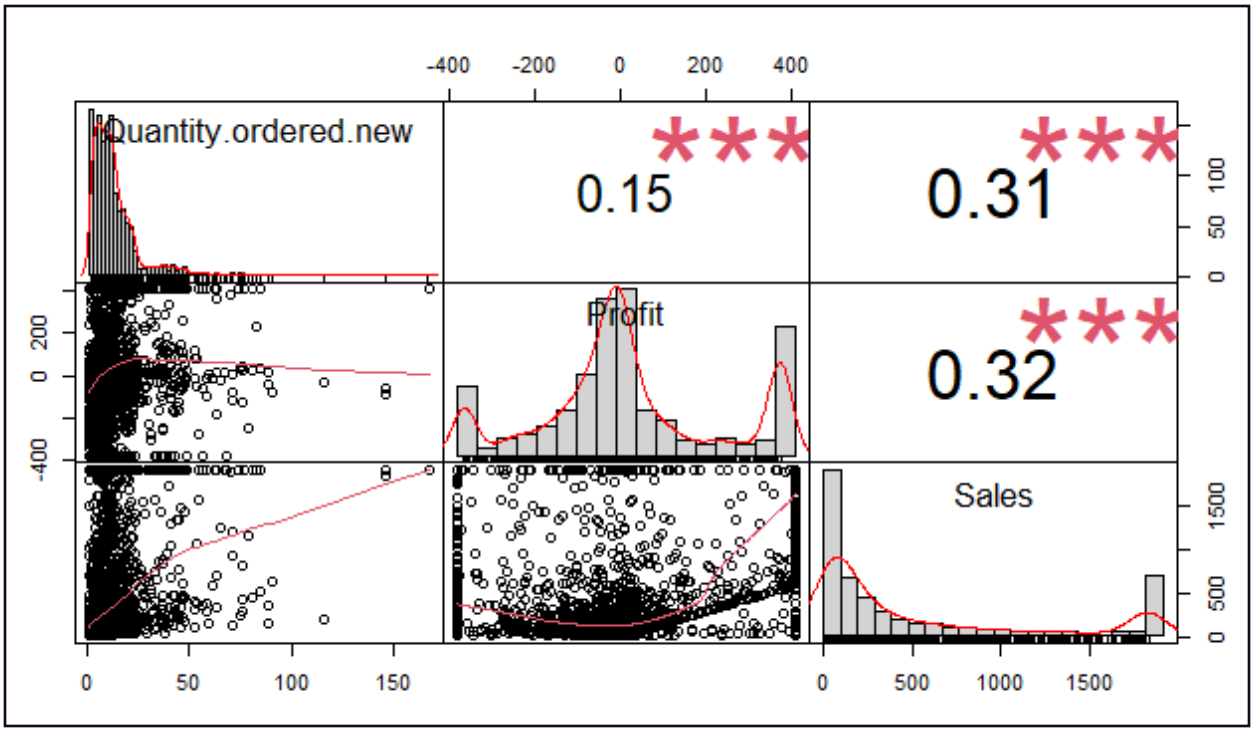

Figure 7: Pairwise Scatterplot

The scatterplot matrix in figure 7 shows random scattered points for all plots among the 3 variables which is also reflected on the pearson correlation coefficient such that **0.15** indicates that there's a weak positive linear association between quantity ordered and profit, while **0.31** indicates that there's a moderate linear association between quantity ordered and sales which is reasonable because when quantity ordered increase a customer has to pay more, **0.32** indicates that there's a moderate linear association between Profit and sales which is also reasonable since profit is a function of sales.

3.A combination of both qualitative and quantitative

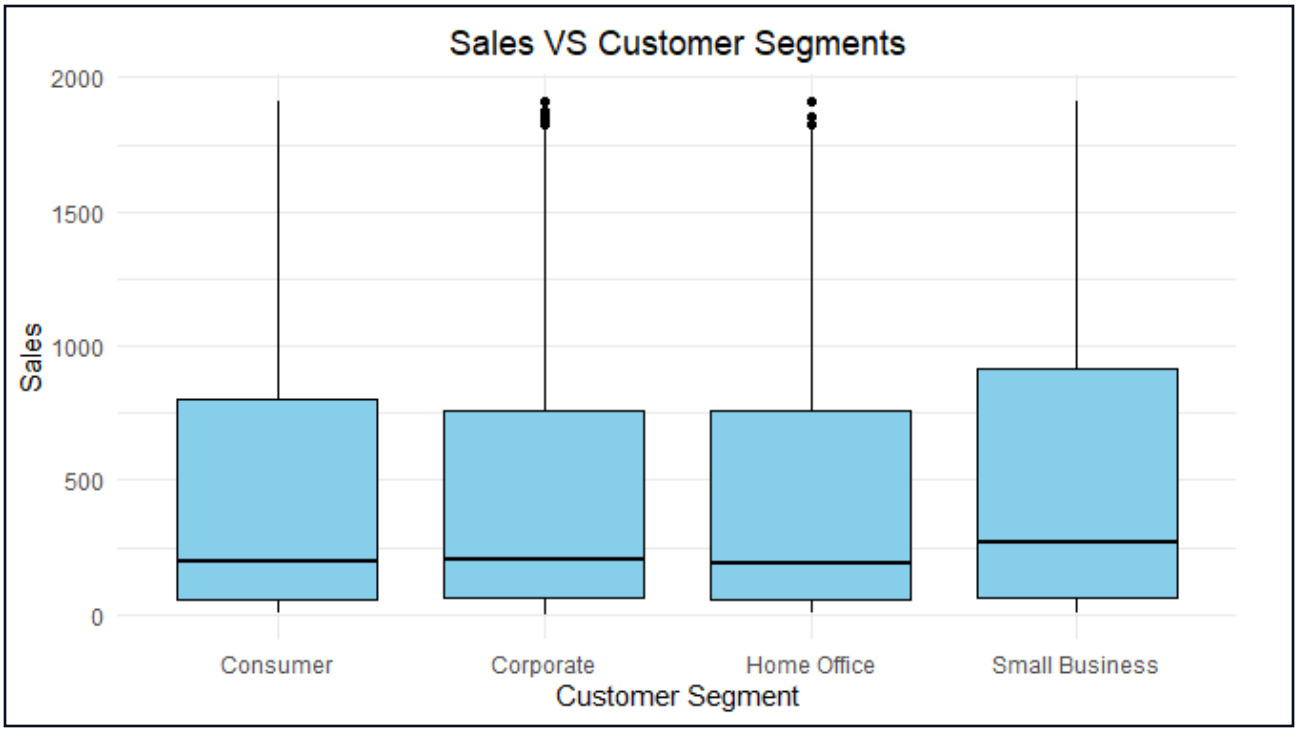

Figure 7: Sales VS Customer Segments Boxplot.

Customer Segment	Min	Q1	Q2	Q3	max	IQR
Consumer	3.08	57.73	198.28	804.98	1913.84	747.24
Corporate	2.25	59.29	204.76	760.50	1913.84	701.2175
Home Office	3.13	56.52	188.01	760.61	1913.84	704.0975
Small Business	4	66.34	267.75	915.49	1913.84	849.15

Table 6: Sales Summary Statistics for Customer Segments.

The table and figure above shows that small businesses has the biggest contribution to sales since it has the largest min,Q1,Q2, Q3 as well as the IQR which shows the spread of small businesses on our sales, Consumer category has the second biggest contribution to sales followed by home office and corporate.