

**Final Exam**

**INFO8675-22S-Sec4-Advanced Data Analytics**

**Olutoyin Taiwo**

**August 21, 2022**

# SuperStore Sales Report

## Introduction

We have been provided with historical sales data for 45 stores located in different regions- each store contains a number of departments.

We are to analyze the data and present a dashboard and report to the management providing recommendations for business growth and expansion.

## Data Exploration and Descriptive Statistics

### Variables Definitions and Data Types

There are 8048 rows in the Superstore Exam Data file with the first row being the header (labels) and the remaining 8047 being the dataset. We have 15 variables (columns) altogether.

**Nominal (categorical) Variable** - meaning they are non-numerical and have no clear ordering of the categories

**Ordinal Variable** - meaning they are categorical but have a clear ordering of the categories

**Discrete Variable** - meaning they are numerical but at set intervals

**Continuous Variable** - meaning they are numerical and can theoretically take any value in a range

S/No	Variable	Description	Number of Non-Missing Values	Number of Missing Values	Data Type
1.	Order ID	A unique ID given to each order placed.	8047	0	Categorical
2.	Order Date	The date at which the order was placed	8047	0	Date
3.	Customer Name	Name of the customer placing the order	8047	0	Categorical
4.	Country	The country to which the customer belongs to	8047	0	Categorical
5.	State	The state to which the customer belongs from the country	8047	0	Categorical
6.	Region	Contains the region details	8047	0	Categorical
7.	Segment	The ordered product belongs to what segment	8047	0	Categorical
8.	Category	Contains the details about what category the product belongs to	8047	0	Categorical

9.	Sub-Category	Contains the details about what sub - category the product belongs to	8047	0	Categorical
10.	Product Name	The name of the product ordered by the customer	8047	0	Categorical
11.	Discount	The discount applicable on a product	8047	0	Numerical (Continuous)
12.	Sales	The actual sales happened for a particular order	8047	0	Numerical (Continuous)
13.	Profit	Profit earned on an order	8047	0	Numerical (Continuous)
14	Quantity	The total quantity of the product ordered in a single order	8047	0	Numerical (Discrete)
15.	Feedback?	The feedback given by the customer on the complete shopping experience. If feedback provided, then TRUE. If no feedback provided, then FALSE	8047	0	Categorical

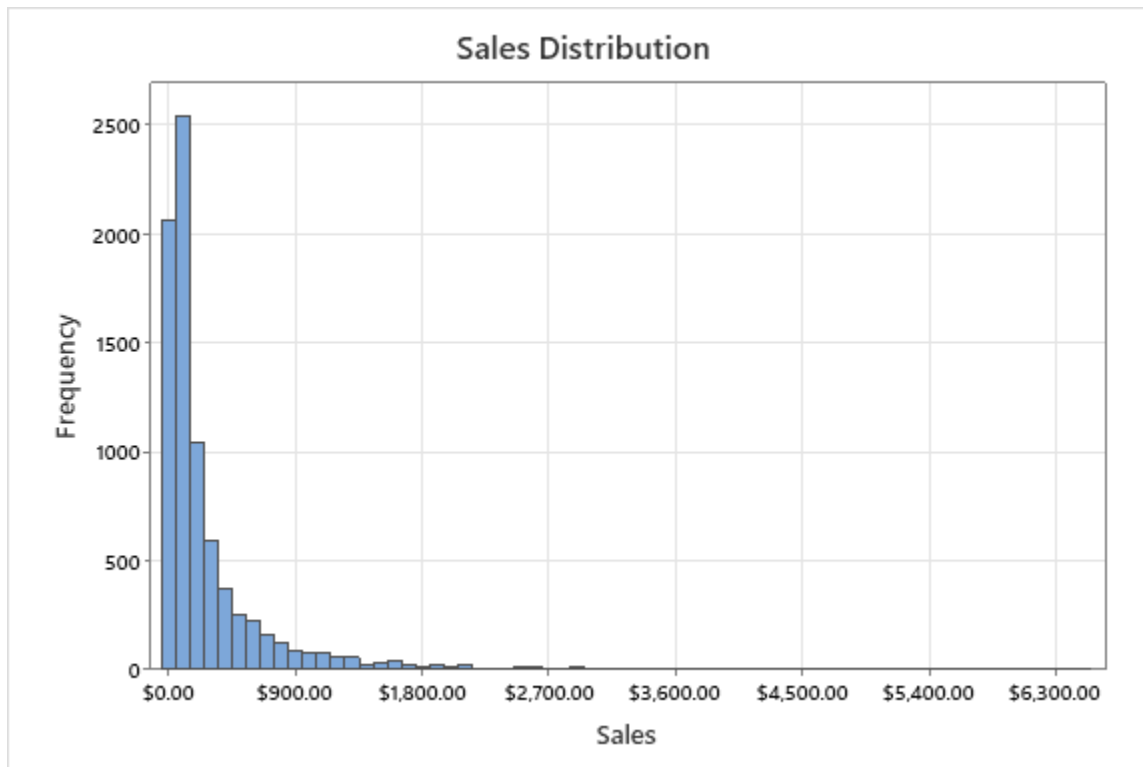
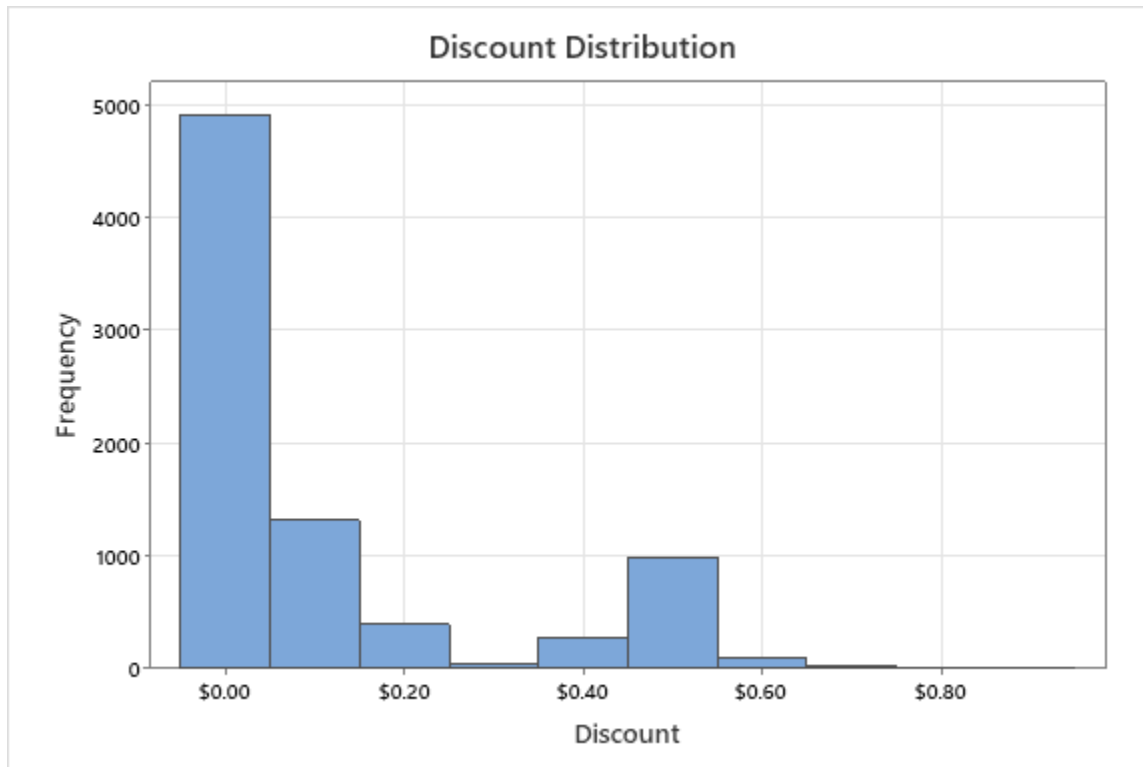
There were no missing values in the dataset. The values with "\$ -" were actually \$0 which makes sense for the numerical variables.

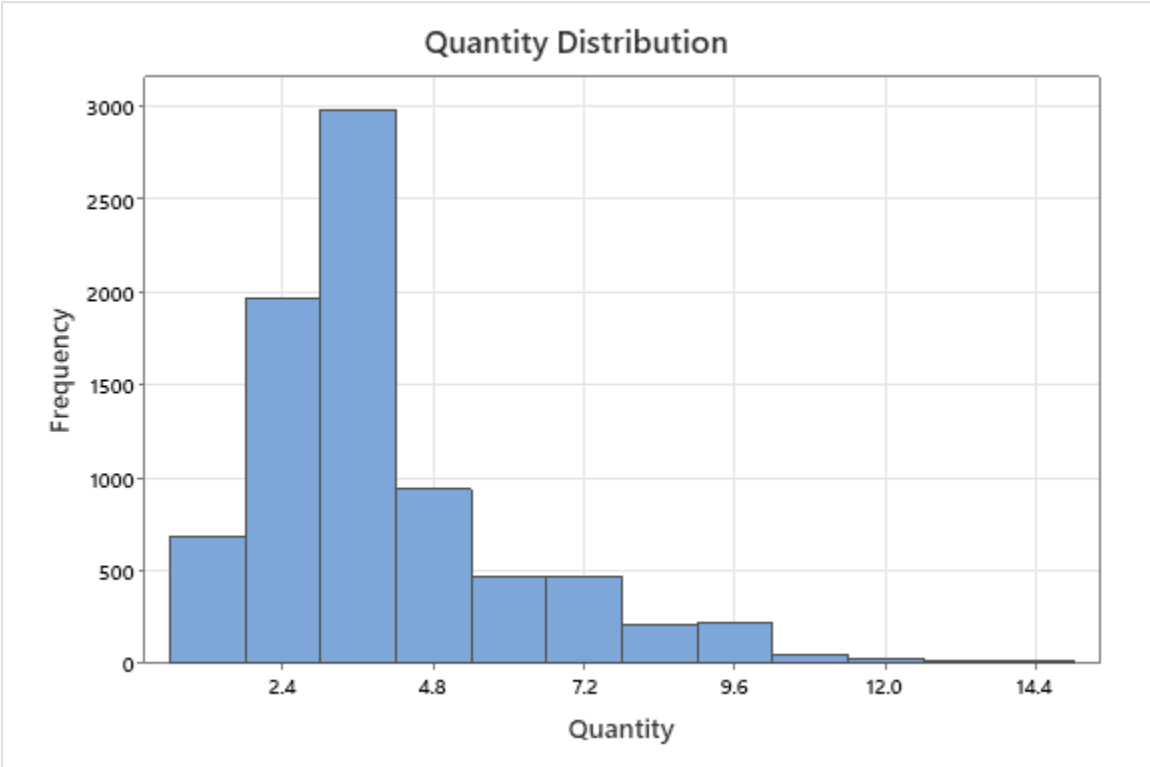
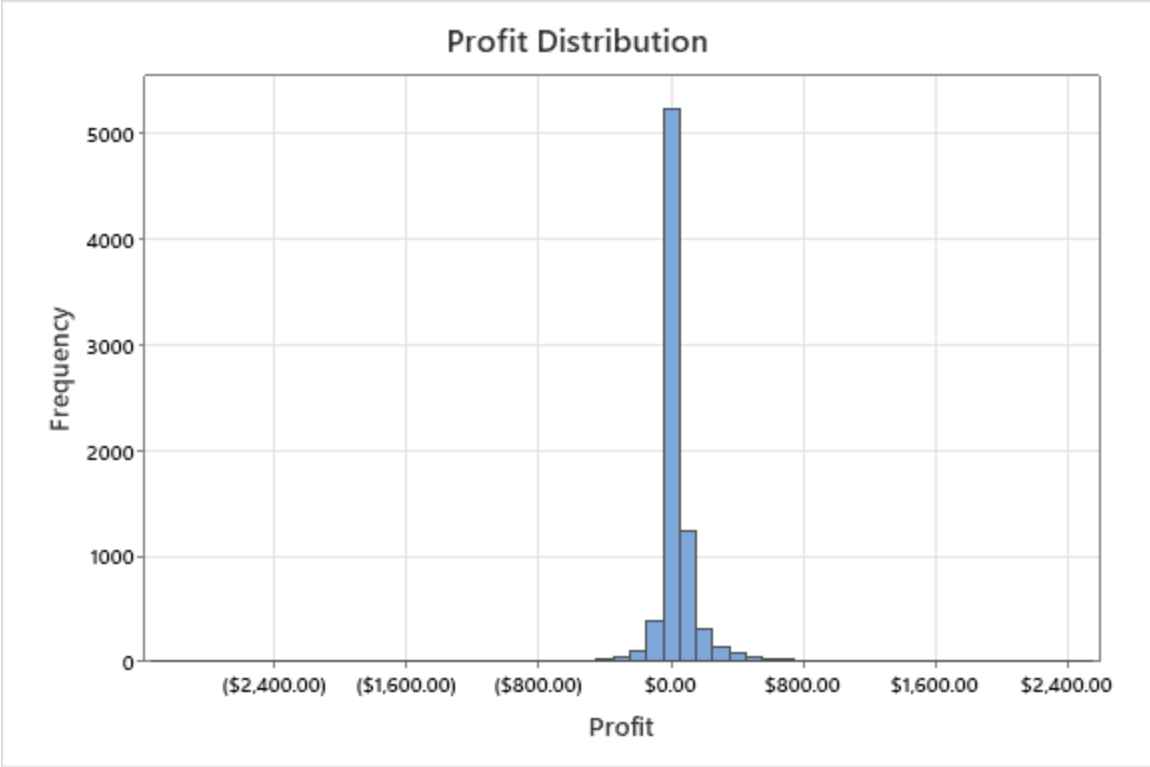
### Descriptive Statistics

	<i>Discount</i>	<i>Sales</i>	<i>Profit</i>	<i>Quantity</i>
<b>Mean</b>	0.110047223	291.8456568	35.19821051	3.772088977
<b>Standard Error</b>	0.002026336	5.408971242	1.985683072	0.024562371
<b>Median</b>	0	117	14	3
<b>Mode</b>	0	27	0	3
<b>Standard Deviation</b>	0.18177262	485.2121564	178.1258436	2.203369259
<b>Sample Variance</b>	0.033041285	235430.8367	31728.81616	4.854836091
<b>Kurtosis</b>	0.849252193	28.08623438	53.17520411	2.066940443
<b>Skewness</b>	1.546988789	4.308980215	0.657767434	1.310954718
<b>Range</b>	0.85	6514	5536	13
<b>Minimum</b>	0	3	-3060	1
<b>Maximum</b>	0.85	6517	2476	14
<b>Sum</b>	885.55	2348482	283240	30354
<b>Count</b>	8047	8047	8047	8047

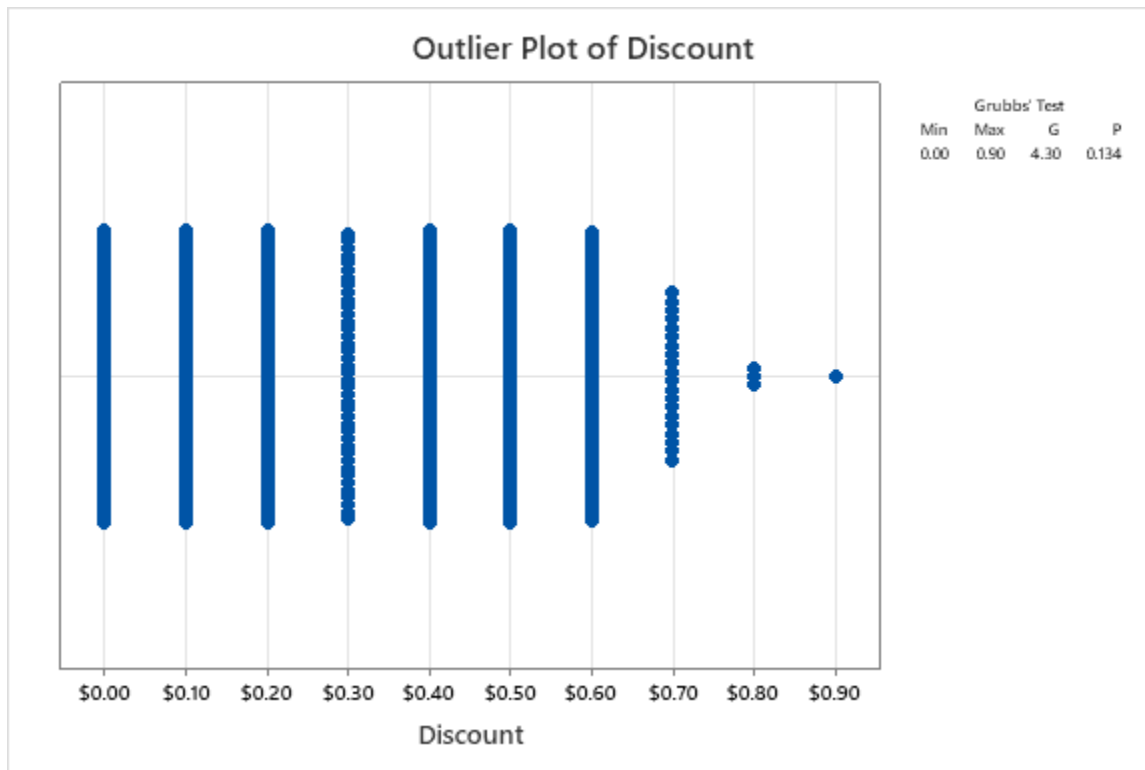
<b>Q1 Excl</b>	0	48	1	2
<b>Q3 Excl</b>	0.1	313	47	5
<b>IQR (Using Q's Excl)</b>	0.1	265	46	3
<b>Q1 Incl</b>	0	48	1	2
<b>Q3 Incl</b>	0.1	313	47	5
<b>IQR (Using Q's Incl)</b>	0.1	265	46	3

## Distribution of the Variables

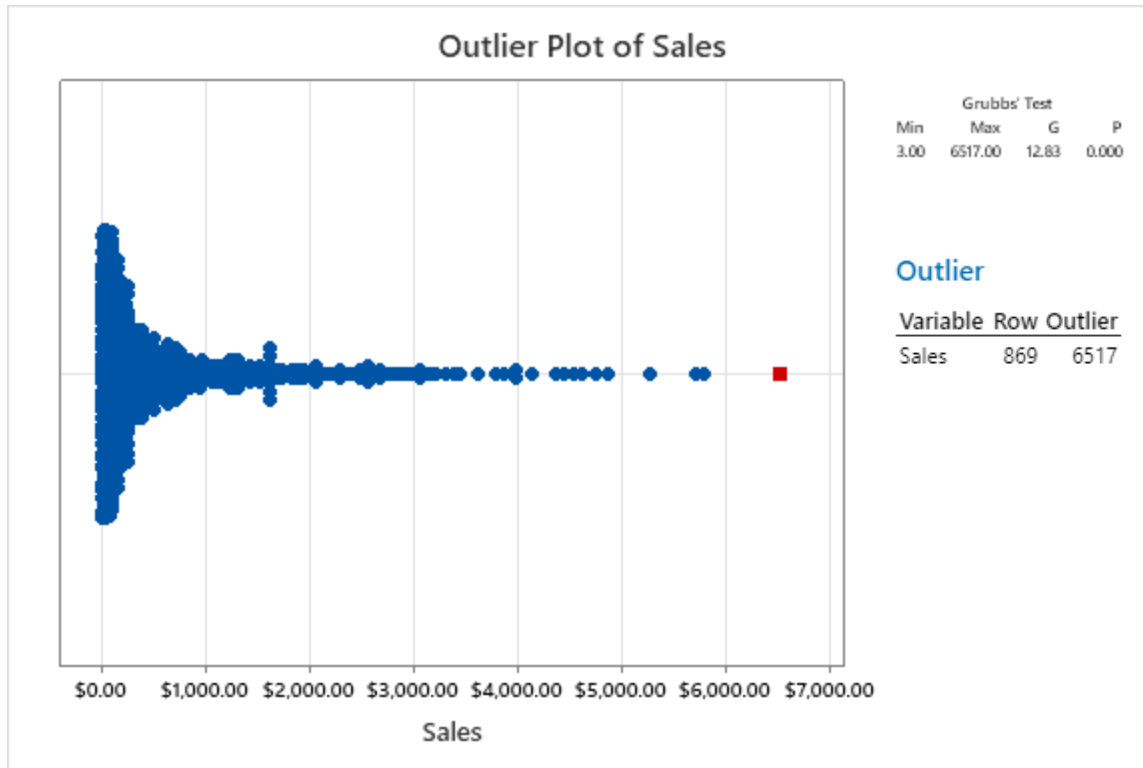




## Outlier Test

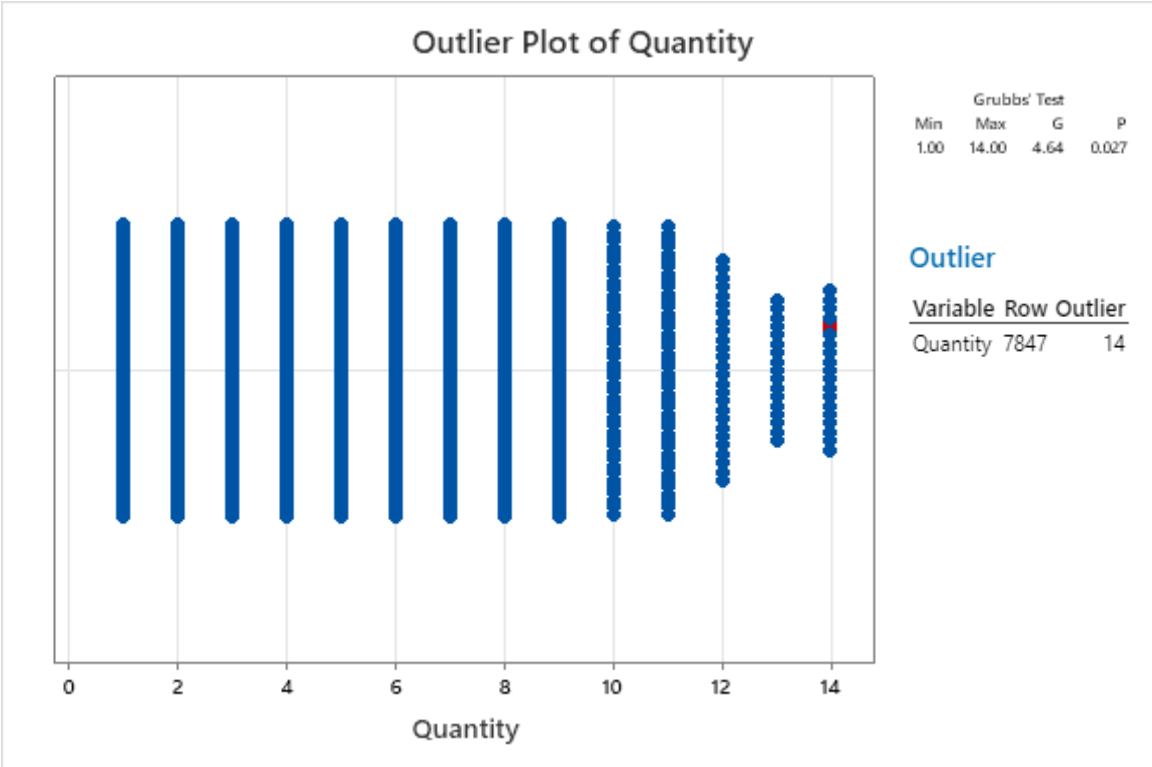
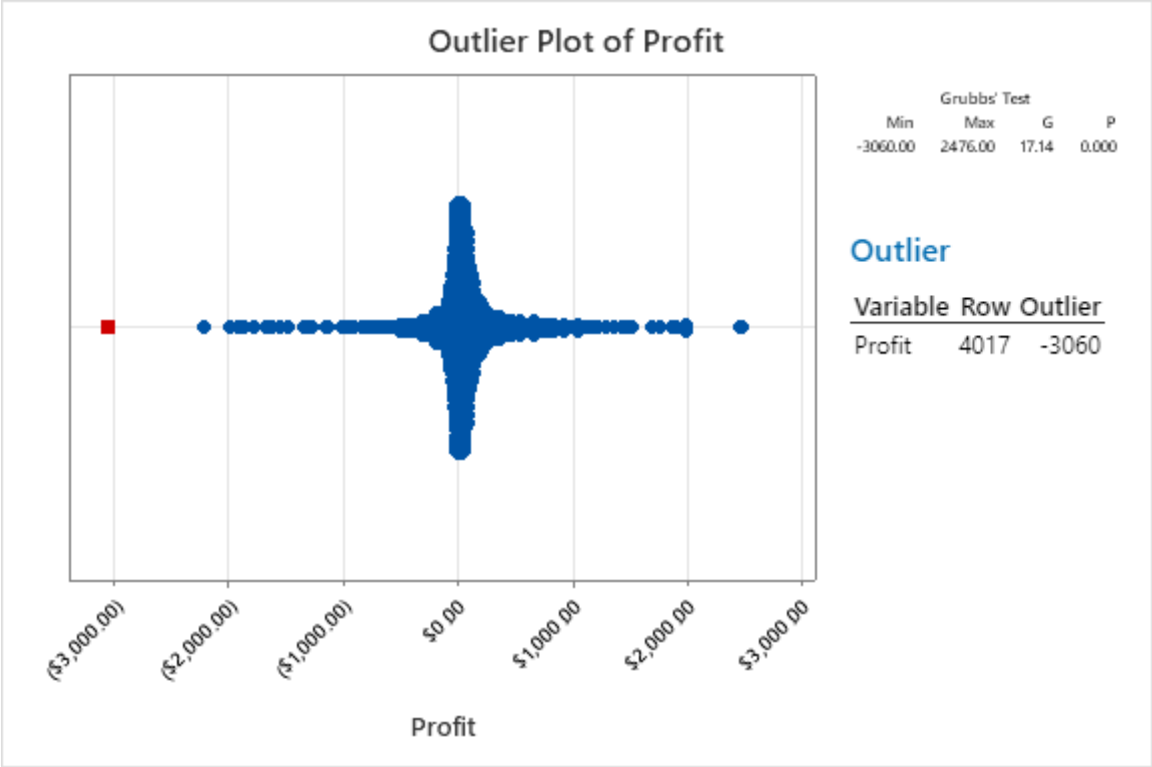


There are no outliers in the discount data.



From the outlier plot, the sales data has one outlier. Upon further investigation, this data is the result of the type of product (high-priced products) and the quantity (large) purchased. This data is actually valid.





### Check for Missing Values

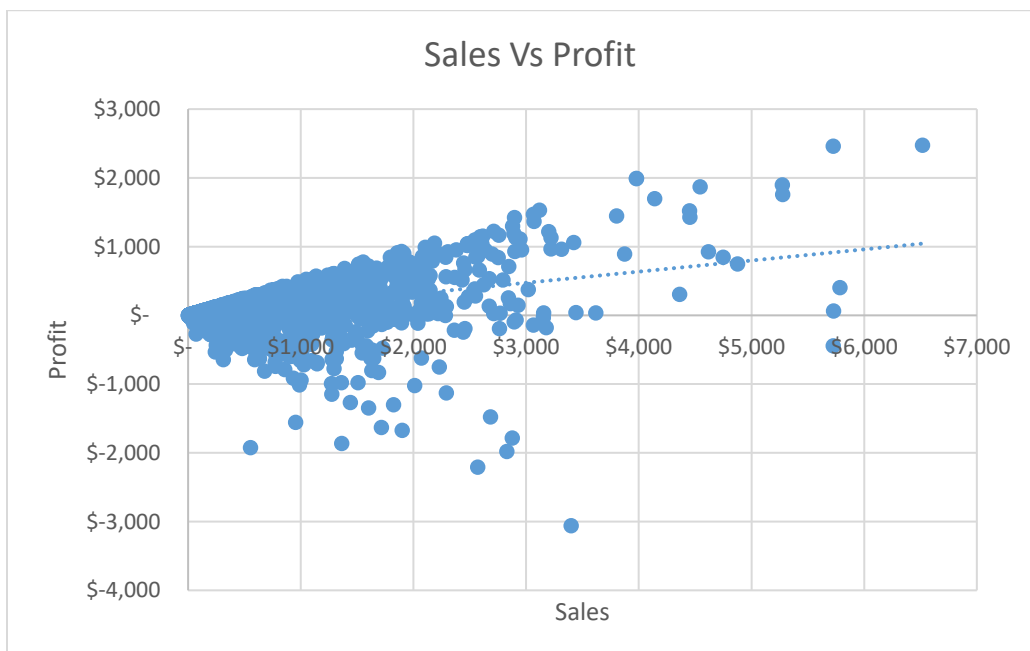
There are no missing values in the dataset. The numerical variables that have the value “\$ -” indicate that the value is 0 which makes sense. It is possible to give \$0 discount, make \$0 sales if quantity is 0 and it’s also possible to make \$0 profit. Thus, there’s no imputations to be done in this dataset.

Upon thorough review of the dataset, I have decided to proceed with my analysis using the values provided in the original data (not removing the outliers as the data seem valid).

### Relationship Between Variables

	<i><b>Discount</b></i>	<i><b>Sales</b></i>	<i><b>Profit</b></i>	<i><b>Quantity</b></i>
<b>Discount</b>	1			
<b>Sales</b>	-0.02635	1		
<b>Profit</b>	-0.35133	0.441217	1	
<b>Quantity</b>	0.003686	0.353441	0.09867	1

There’s no high correlation between any 2 variables.



The scatter plot above between Sales and Profit. We can see that in some cases, sales value have corresponding negative profit, and in other cases, high sales value shows high profit, This shows that there's no clear correlation between sales and profit. High sales don't necessarily mean high profits and low sales don't mean low profits.



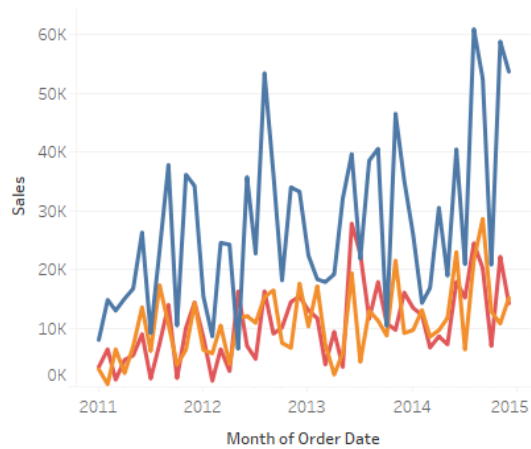
The same explanation for Sale Price Vs Profit applies to this second plot of Sales and Quantity. We can see that high quantity does not necessarily mean high sales. This shows that there's no clear correlation between sales and quantity. In some cases, it is, in other cases, it is not.

There are a number of variables (categorical and numerical), if when combined, influence the sale price such as a combination of subcategory, quantity, category etc), but we won't be going into that now.

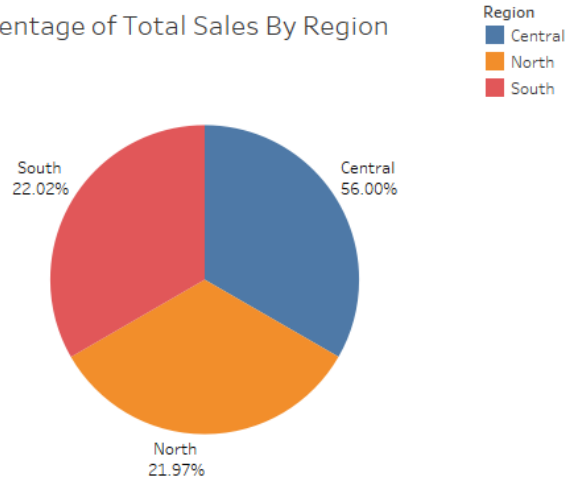
## Dashboard

I have made the dashboard interactive in that when you select a region, only the details corresponding to that region will be highlighted.

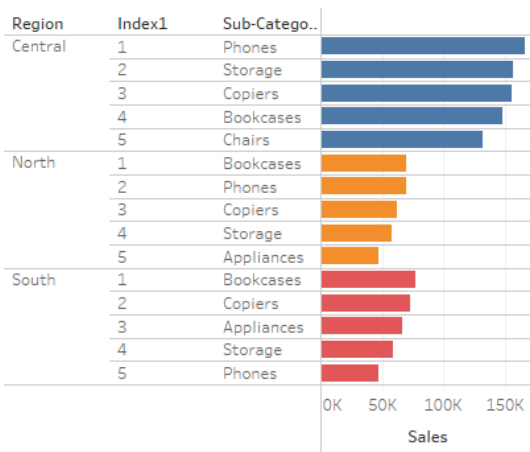
Sales Trend from 2011 to 2015 By Region



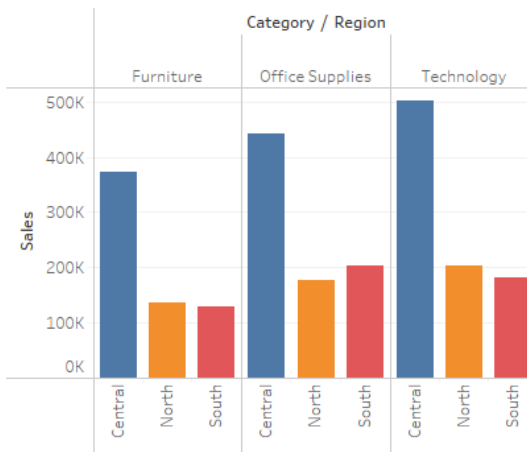
Percentage of Total Sales By Region



Top 5 Selling Subcategory by Region

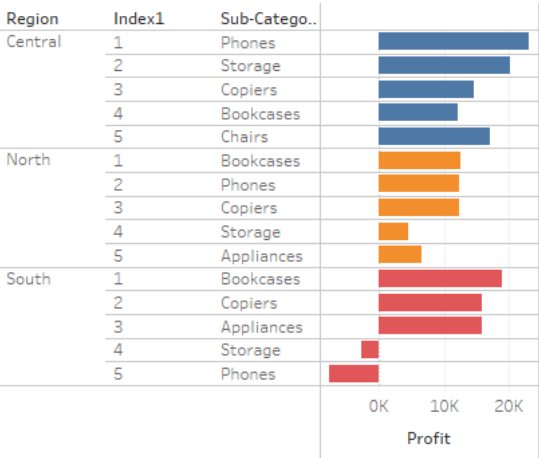


Sale by Region and Category

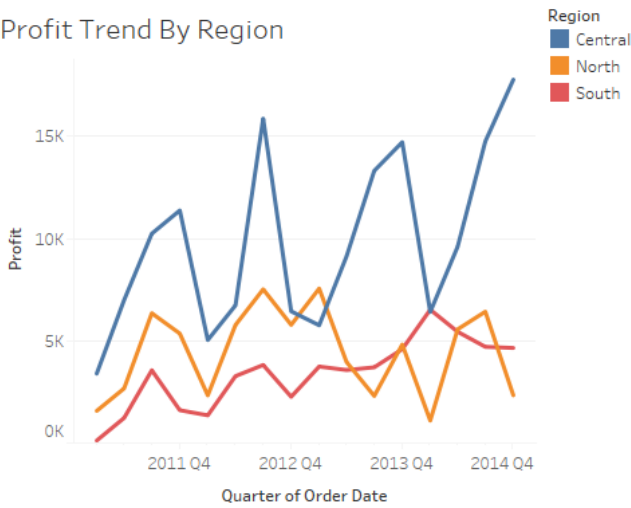


Dashboard 2

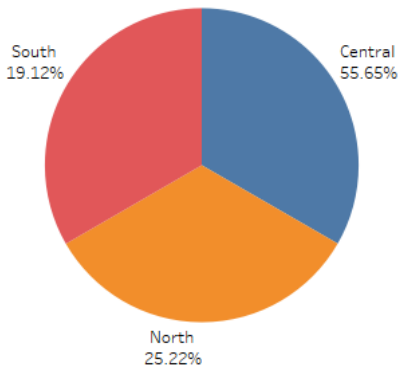
Top 5 Profitable Subcategory By Region



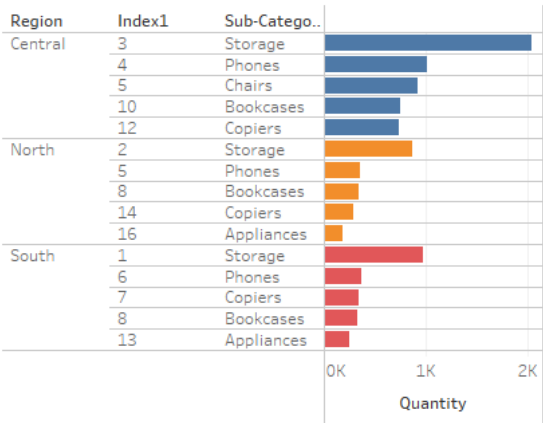
Profit Trend By Region



Percentage of Total Profit By Region



Top 5 Subcategory Products Quantity Sold By Region



# Superstore Exam

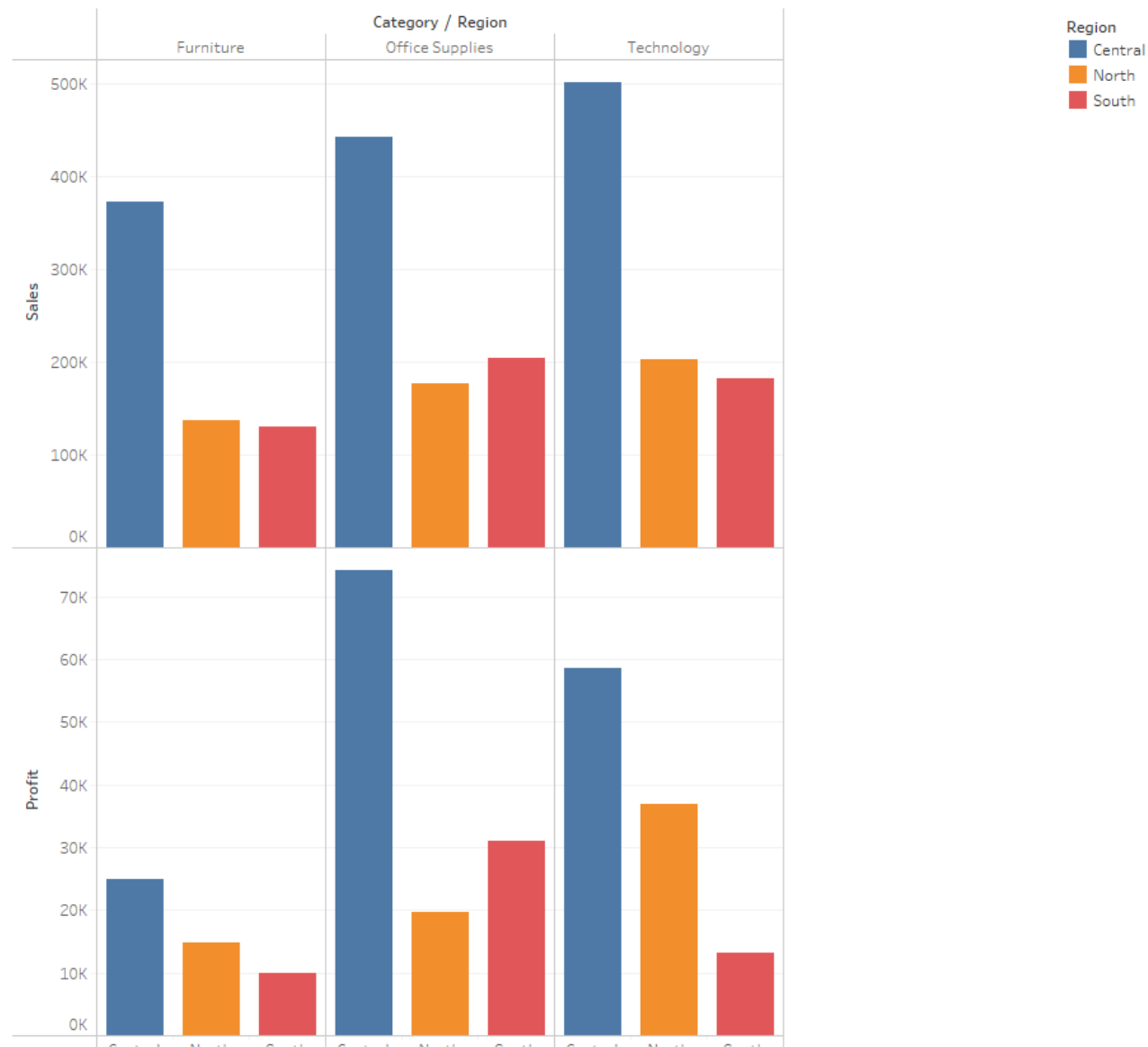
File created on: 8/21/2022 12:39:11 AM

Quantity Sold by Region and Product Name	Sale and Profit by Region and Category	Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from Januar...
--	--	------------------	-----------------	------------------------------	--	------------------------------------

Product Name	Region	
Acco 3-Hole Punch, Clear	Central	19
	North	2
	South	10
Acco 3-Hole Punch, Durable	Central	26
	North	9
	South	11
Acco 3-Hole Punch, Economy	Central	25
	North	15
	South	19
Acco 3-Hole Punch, Recycled	Central	14
	North	4
	South	12
Acco Binder Covers, Clear	Central	3
	North	13
	South	14
Acco Binder Covers, Durable	Central	16
	North	7
	South	3
Acco Binder Covers, Economy	Central	22
	North	12
	South	16
Acco Binder Covers, Recycled	Central	21
	North	12
	South	3
Acco Binder, Clear	Central	18
	North	4
	South	3
Acco Binder, Durable	Central	25
	North	4
	South	9
Acco Binder, Economy	Central	12
	North	7
	South	14
Acco Binder, Recycled	Central	3
	South	8
Acco Binding Machine, Clear	Central	20
	North	2
	South	3
Acco Binding Machine, Durable	Central	22
	North	12
	South	6

## Summary of Products sold by Region from 2011 to 2014

Quantity Sold by Region and Product Name	Sale and Profit by Region and Category	Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from Januar..
--	--	------------------	-----------------	------------------------------	--	-----------------------------------



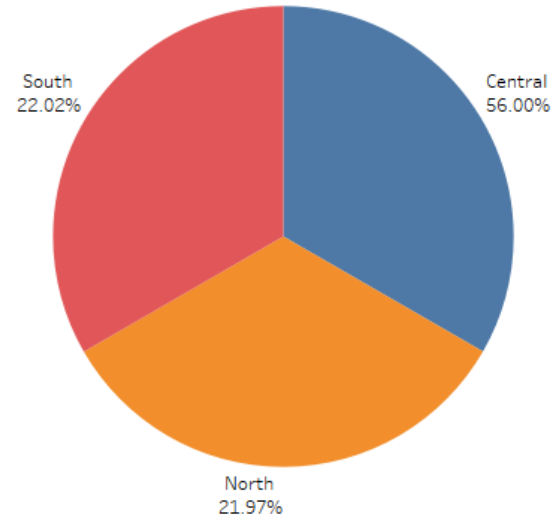
## Which Category is the Top Selling and Most Profitable?

Out of the three Product Category, Technology has the highest sales revenue in all the regions with Central region taking the lead, followed by Office Supplies, then Furniture.

Office Supplies is the most profitable in all the regions, with the highest profit being Central region.



Quantity Sold by Region and Product Name	Sale and Profit by Region and Category	Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from Januar..
--	--	------------------	-----------------	------------------------------	--	-----------------------------------



## Which Region is the most profitable?

The most profitable region is the Central region with 56% of the total profits.

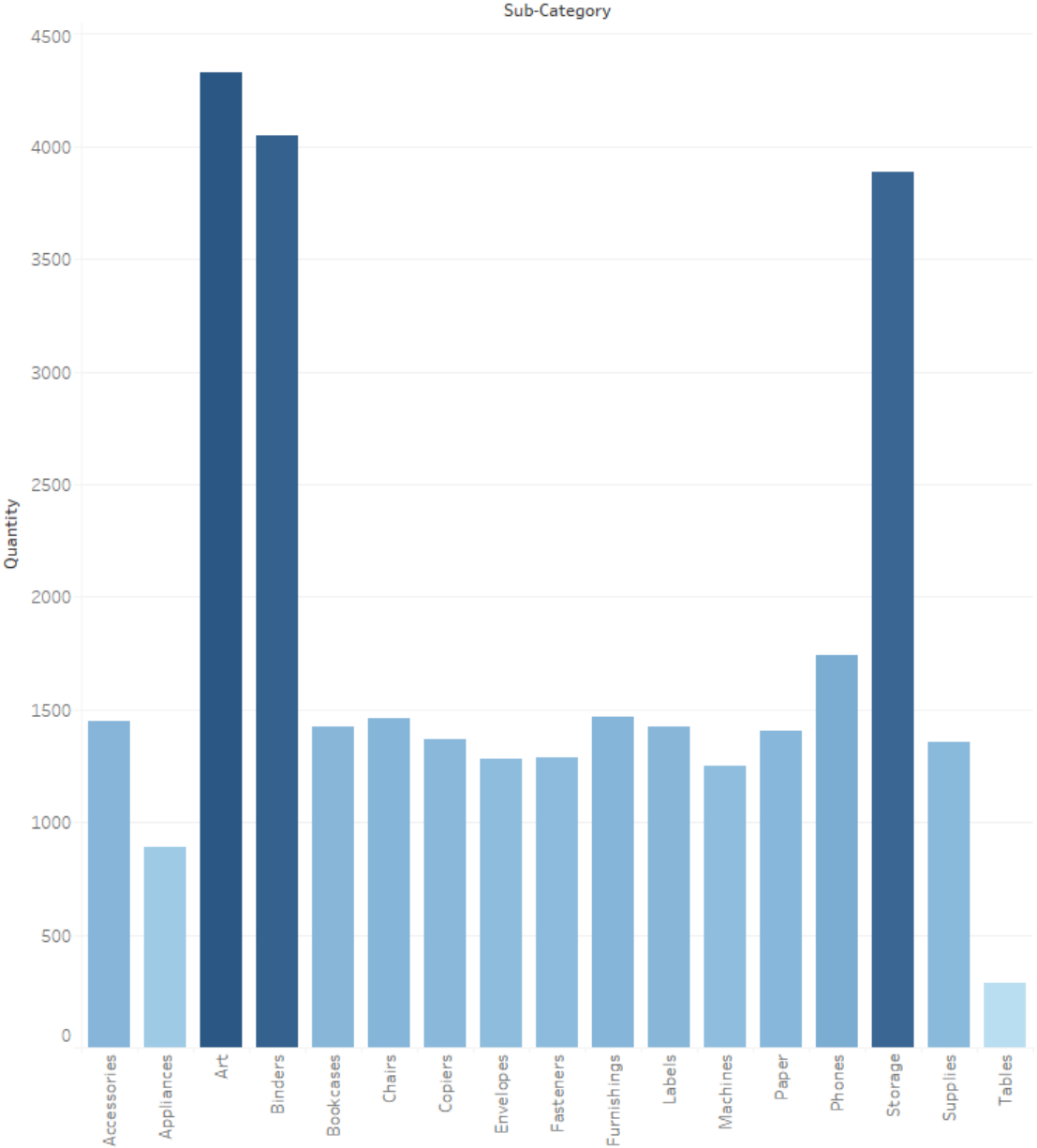
Quantity Sold by Region and Product Name	Sale and Profit by Region and Category	Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 ..
--	--	------------------	-----------------	------------------------------	--	---



## Which Region is the Top Selling?

Central region is the top selling region, just as it is the most profitable region. This region is a “star” region and the management should look out for opportunities for expansion and growth in this region.

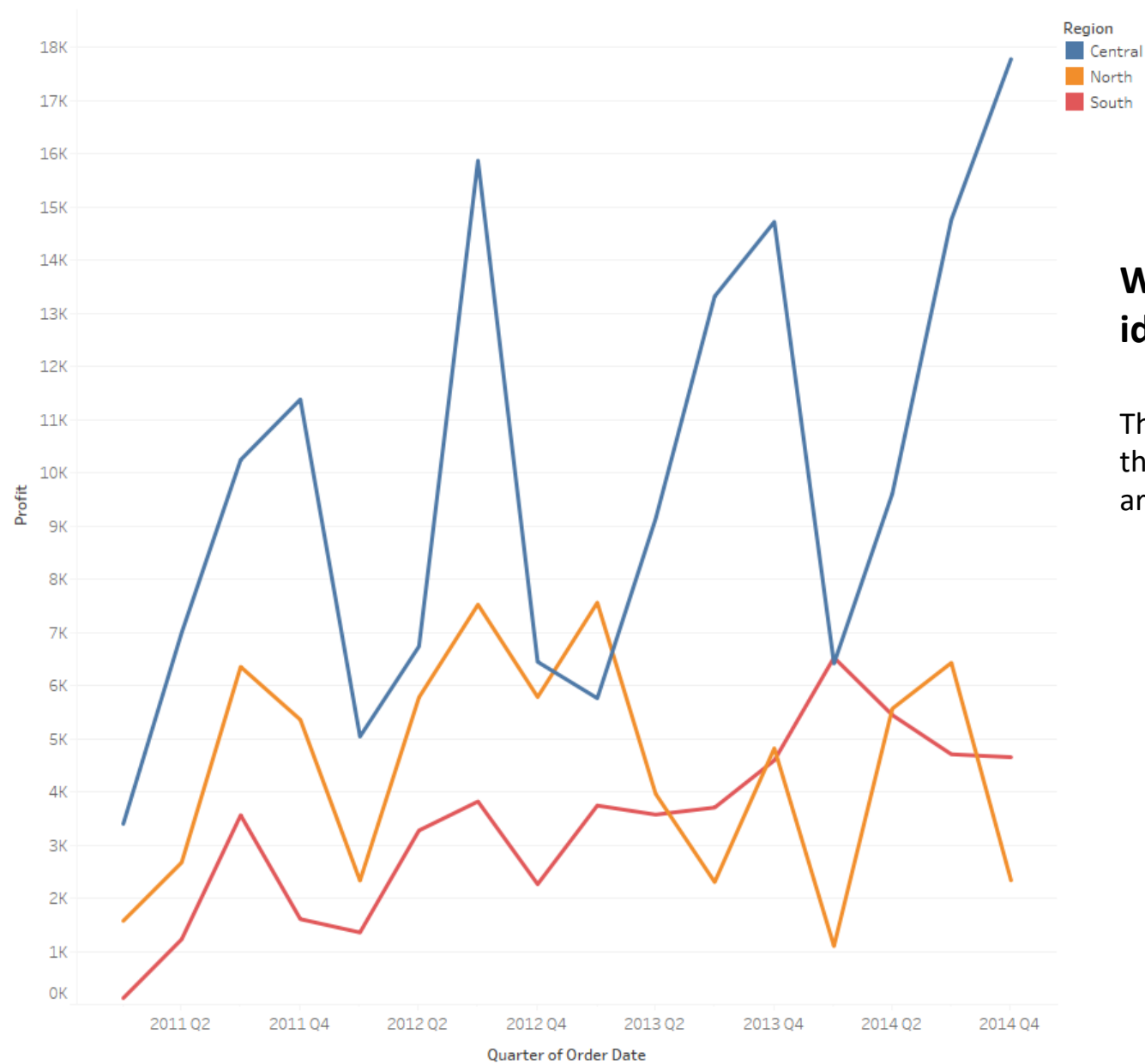
Sale and Profit by Region and Category	Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 By Re..	Sales and Profit By Customer Segment
--	------------------	-----------------	------------------------------	--	--	--------------------------------------



### Which 3 Subcategory are the Top Selling?

The top 3 selling Subcategory are Art, Binders and Storage, in the order of highest to lowest. This implies that the demand for these 3 subcategories is very high compared to other subcategories. Management should make sure to stock up inventory and never run out of stock.

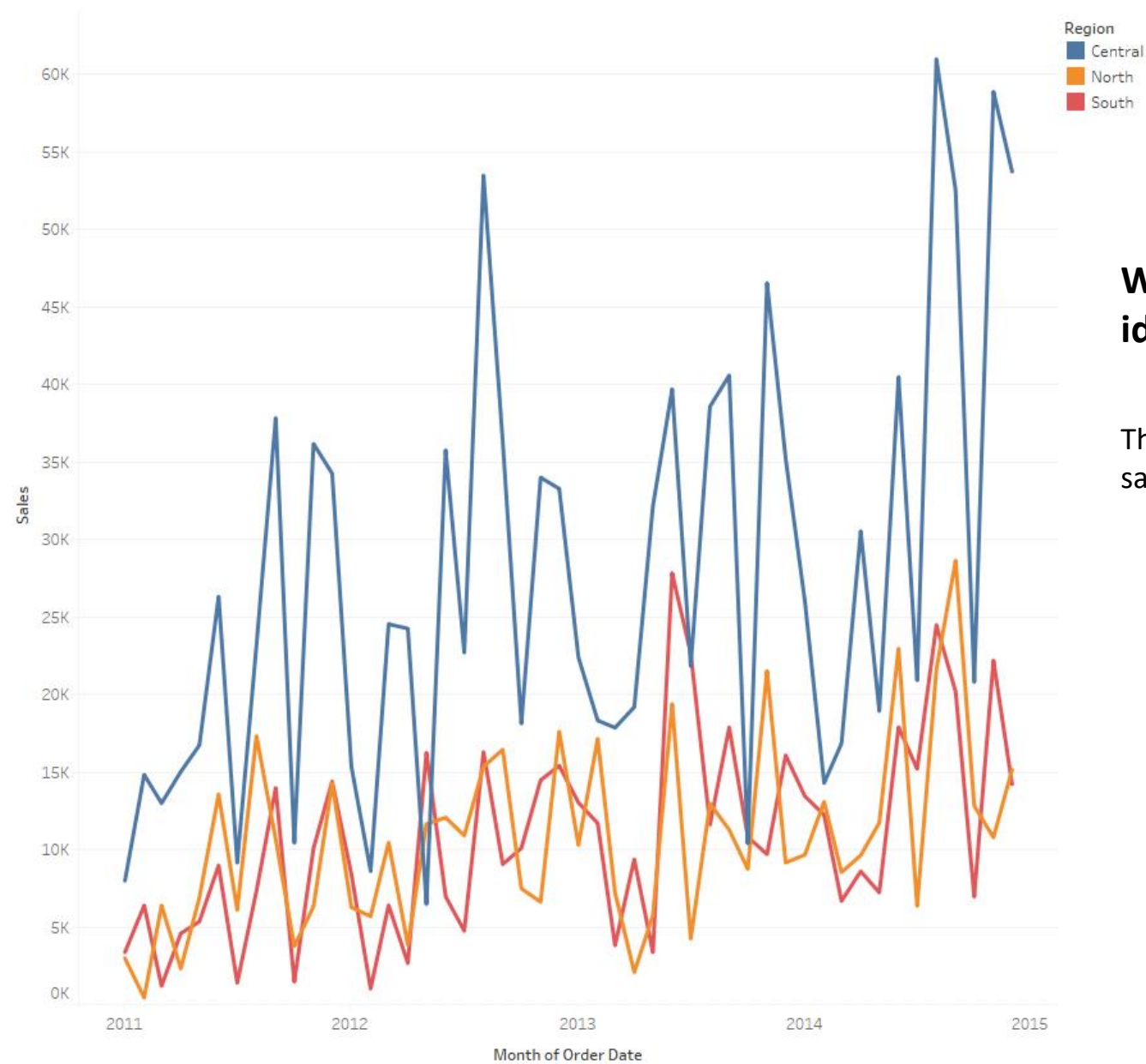
Profit By Region	Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 By Re...	Sales and Profit By Customer Segment	Top Subcategory in Demand
------------------	-----------------	------------------------------	--	---	--------------------------------------	---------------------------



**Which seasonal trend can you identify in the Profit time series?**

There is a seasonal trend to profit in all the regions up till Q2 2013. Then there are irregularities and fluctuations.

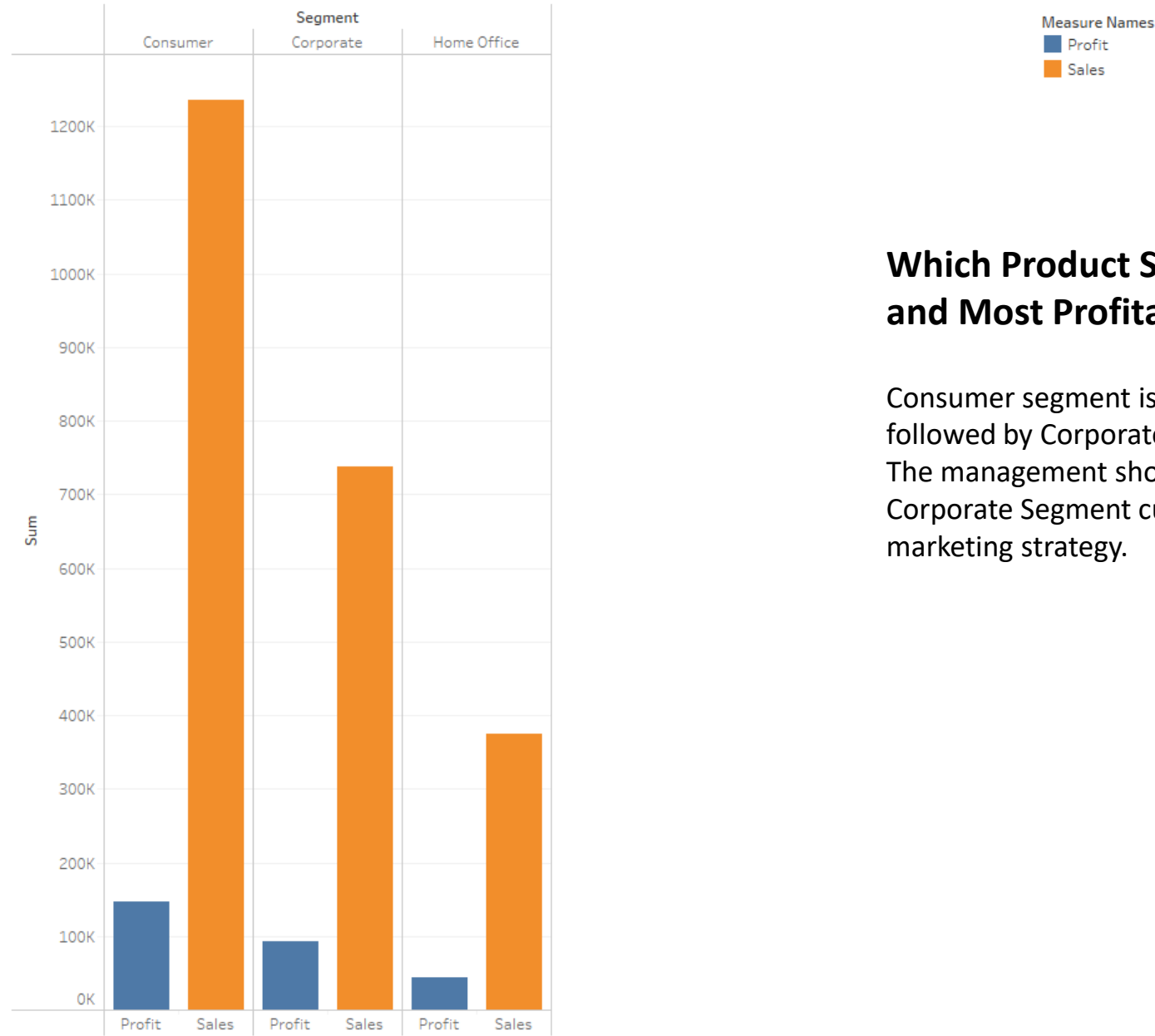
Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 By Re...	Sales and Profit By Customer Segment	Top Subcategory in Demand	Recommendation
-----------------	------------------------------	--	---	--------------------------------------	---------------------------	----------------



**Which seasonal trend can you identify in the Sales time series?**

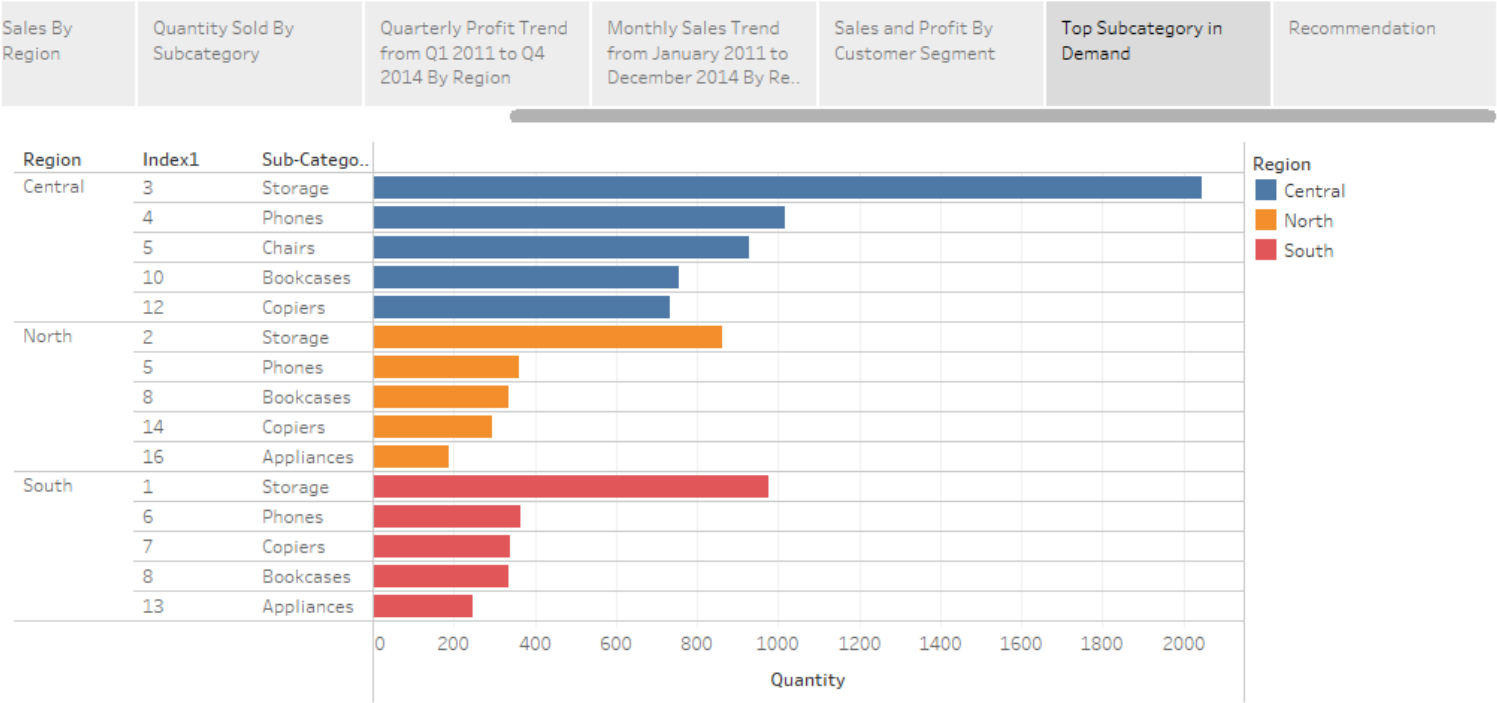
There is a clear monthly seasonality in the sales for each region.

Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 By Re..	Sales and Profit By Customer Segment	Top Subcategory in Demand	Recommendation
-----------------	------------------------------	--	--	--------------------------------------	---------------------------	----------------



### Which Product Segment is the Top Selling and Most Profitable?

Consumer segment is the top selling and most profitable, followed by Corporate Segment and then, Home Office. The management should work on retaining Consumer and Corporate Segment customers by improving their marketing strategy.



**Which top 5 Subcategory have the highest number of quantity sold?**

The top 5 subcategory with highest quantity sold in each region is Storage. In other words, it implies that there is high demand for Storage in all of the three regions. The management should therefore ensure that these products are never out of stock.

Sales By Region	Quantity Sold By Subcategory	Quarterly Profit Trend from Q1 2011 to Q4 2014 By Region	Monthly Sales Trend from January 2011 to December 2014 By Re..	Sales and Profit By Customer Segment	Top Subcategory in Demand	Recommendation
-----------------	------------------------------	--	--	--------------------------------------	---------------------------	----------------

### Recommendations

1. Storage and Phones should never be out of stock as they are the highest in demand as we can see in the plot of Top 5 Subcategory Quantity Sold by Region.
2. Focus on Technology Category and Office Supplies as they both are doing well in terms of sales and profit.
3. Overall, out of the three regions, Central region takes up about 56% of total sales and also 56% of total profit. This region is a star region and the management should look out for opportunities for expansion and growth in this region.
4. Follow up with your customers. You can offer loyalty program or points for customers. Solicit for feedback from your customers in return for points. Keep communication line open with your customers. You may want to offer incentives for the top 5 customers.
5. For products resulting in losses, the management should consider bundling those products together with high selling and profitable subcategory to reduce the impact of the loss.