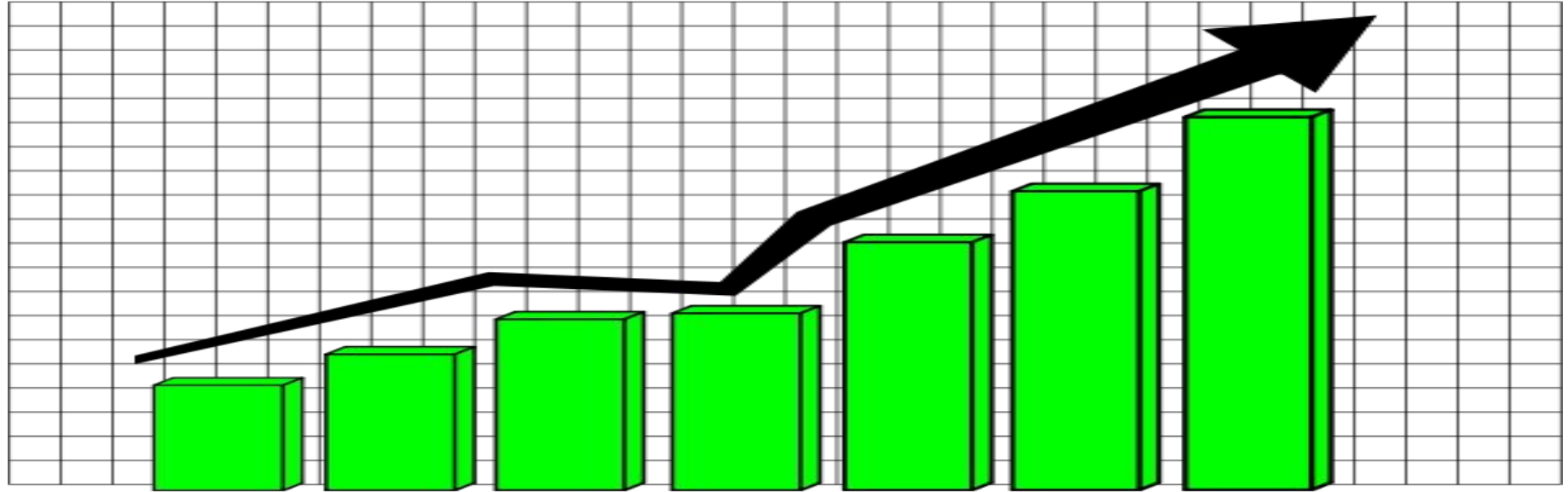


Analyzing and visualizing Superstore sales data



Name - Kartik Dhiman

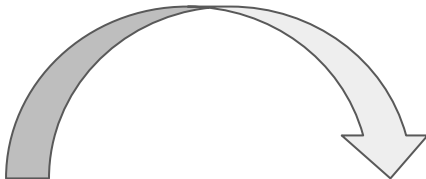
[Follow Me!](#)

Data Set -

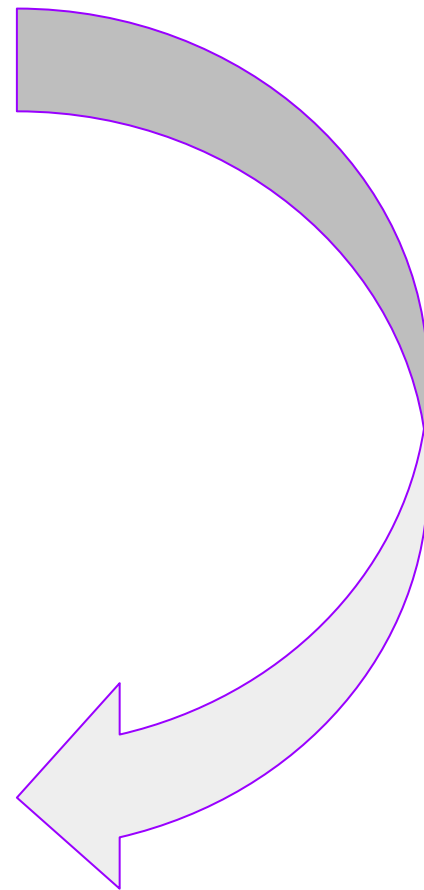
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	...	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority
0	42433	AG-2011-2040	1/1/2011	6/1/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	...	OFF-TEN-10000025	Office Supplies	Storage	Tenex Lockers, Blue	408.300	2	0.0	106.140	35.46	Medium
1	22253	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	...	OFF-SU-10000618	Office Supplies	Supplies	Acme Trimmer, High Speed	120.366	3	0.1	36.036	9.72	Medium
2	48883	HU-2011-1220	1/1/2011	5/1/2011	Second Class	AT-735	Annie Thurman	Consumer	Budapest	Budapest	...	OFF-TEN-10001585	Office Supplies	Storage	Tenex Box, Single Width	66.120	4	0.0	29.640	8.17	High
3	11731	IT-2011-3647632	1/1/2011	5/1/2011	Second Class	EM-14140	Eugene Moren	Home Office	Stockholm	Stockholm	...	OFF-PA-10001492	Office Supplies	Paper	Enermax Note Cards, Premium	44.865	3	0.5	-26.055	4.82	High
4	22255	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	...	FUR-FU-10003447	Furniture	Furnishings	Eldon Light Bulb, Duo Pack	113.670	5	0.1	37.770	4.70	Medium

```
df.shape
✓ 0.0s
(51290, 24)
```

This Data Set Have 51290 Rows And 24 Columns.



- Row ID: the unique identifier of each row
- Order ID: the unique identifier of each order
- Order Date: the date on which the order was placed
- Ship Date: the date on which the order was shipped
- Ship Mode: the shipping mode for the order
- Customer ID: the unique identifier of each customer
- Customer Name: the name of the customer who placed the order
- Segment: the customer segment (Consumer, Corporate, or Home Office)
- City: the city where the customer is located
- State: the state where the customer is located
- Country: the country where the customer is located
- Postal Code: the postal code where the customer is located
- Market: the market (US or APAC) where the customer is located
- Region: the region where the customer is located
- Product ID: the unique identifier of the product in the order
- Category: the category of the product (Furniture, Office Supplies, or Technology)
- Sub-Category: the sub-category of the product
- Product Name: the name of the product
- Sales: the sales amount for the product in the order
- Quantity: the quantity of the product in the order
- Discount: the discount applied to the product in the order
- Profit: the profit earned from the product in the order
- Shipping Cost: the shipping cost for the order
- Order Priority: the priority level of the order (Low, Medium, High, or Critical)



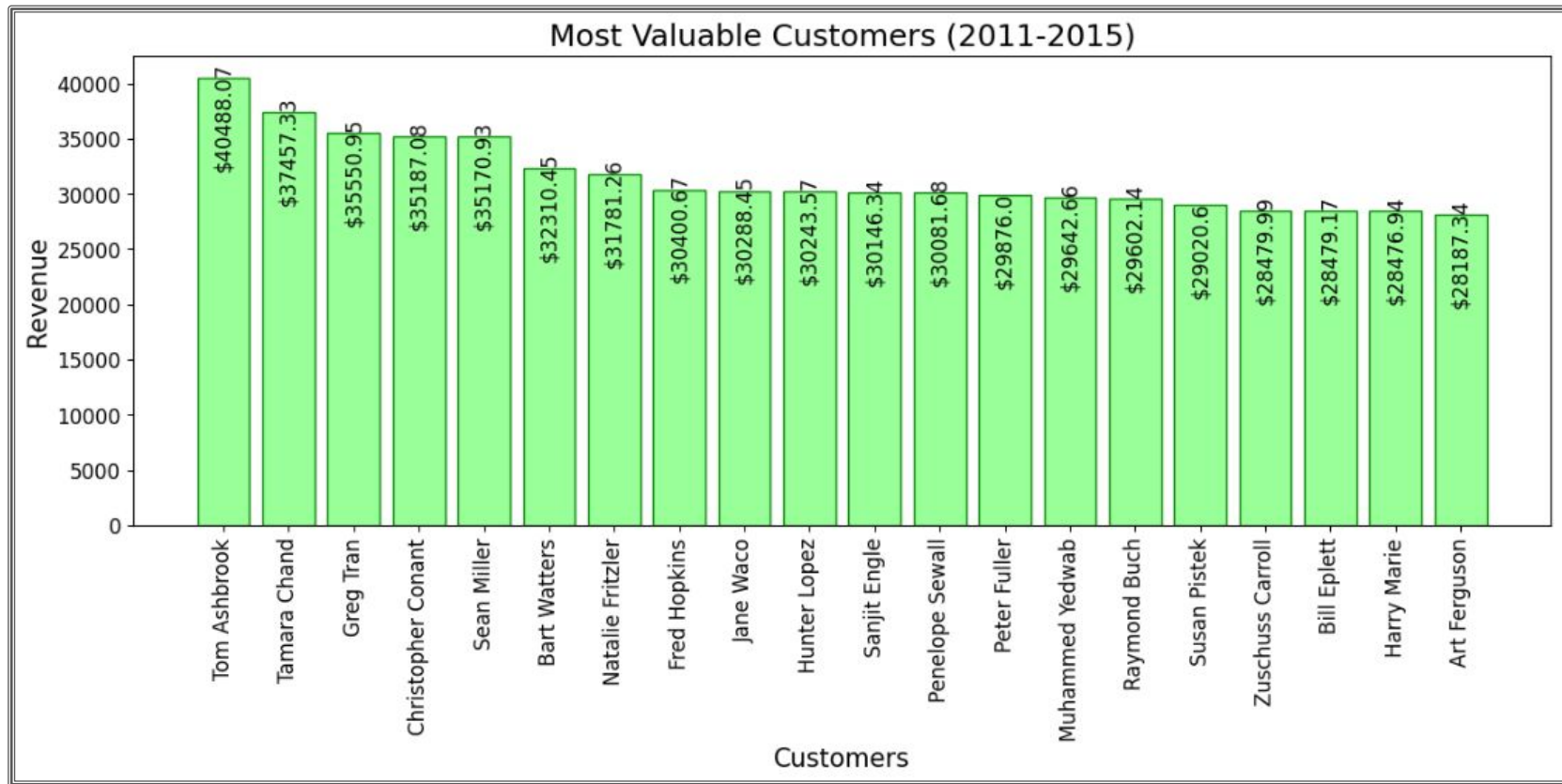
Descriptive Statistics:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit	Shipping Cost
count	51290.00000	9994.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	25645.50000	55190.379428	246.490581	3.476545	0.142908	28.610982	26.375915
std	14806.29199	32063.693350	487.565361	2.278766	0.212280	174.340972	57.296804
min	1.00000	1040.000000	0.444000	1.000000	0.000000	-6599.978000	0.000000
25%	12823.25000	23223.000000	30.758625	2.000000	0.000000	0.000000	2.610000
50%	25645.50000	56430.500000	85.053000	3.000000	0.000000	9.240000	7.790000
75%	38467.75000	90008.000000	251.053200	5.000000	0.200000	36.810000	24.450000
max	51290.00000	99301.000000	22638.480000	14.000000	0.850000	8399.976000	933.570000

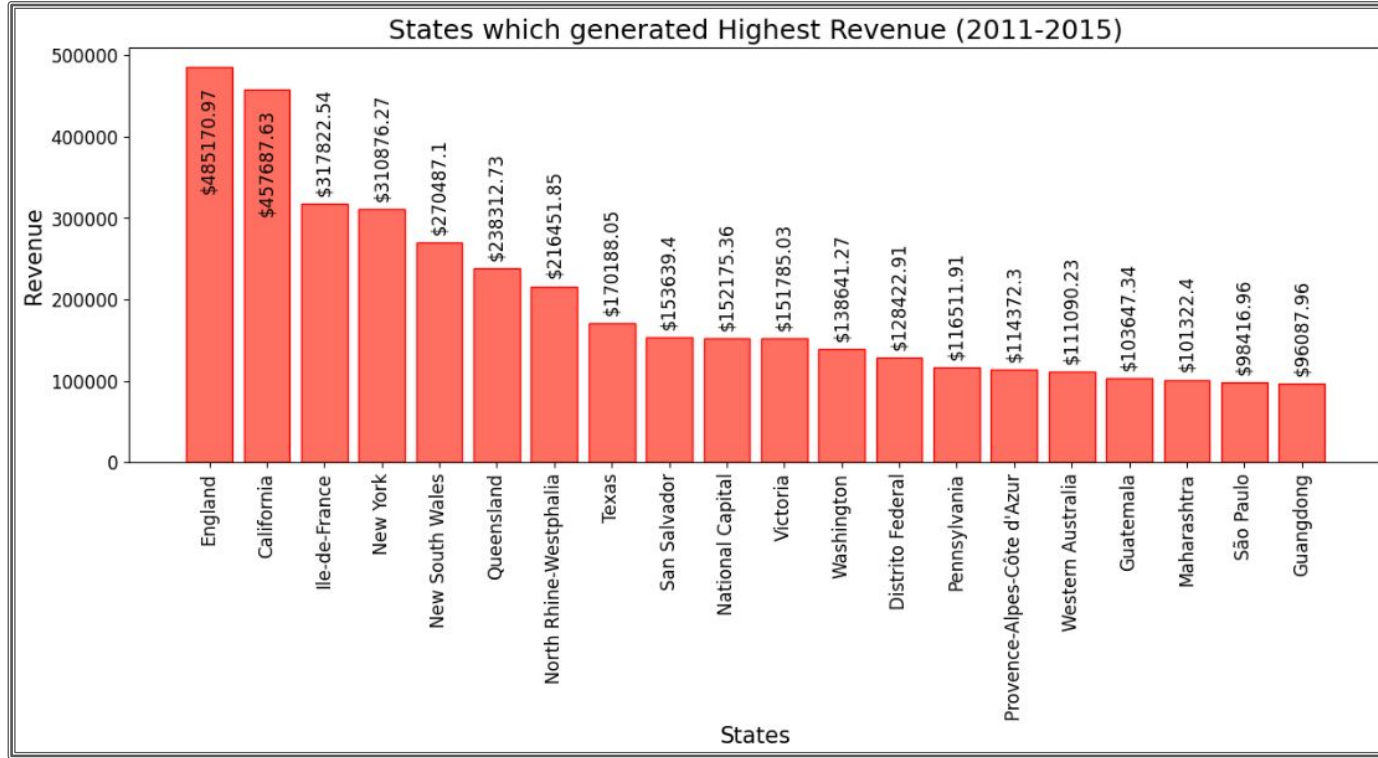
The average sales, profit, and shipping cost are all relatively low, indicating that the company may be operating with relatively low profit margins and the maximum values for each variable are very high, indicating that there may be some sales or products that are generating significant revenue and profit for the company.

The standard deviations for sales, quantity, and profit are relatively high, indicating that there is a wide range of values for these variables. This suggests that there may be some products or regions that are performing better than others. The majority of discounts and shipping costs are relatively low, indicating that the company may be prioritizing cost control in these areas

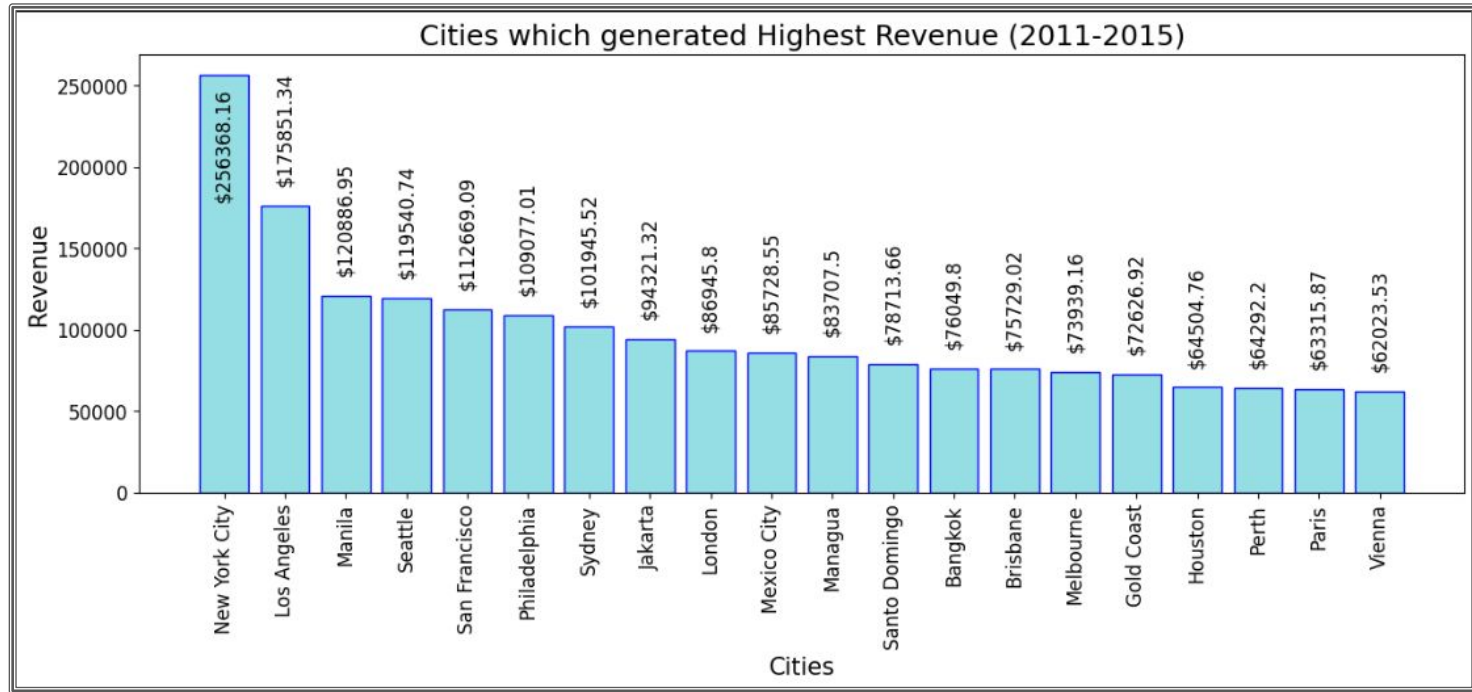
Exploratory Data Analysis:



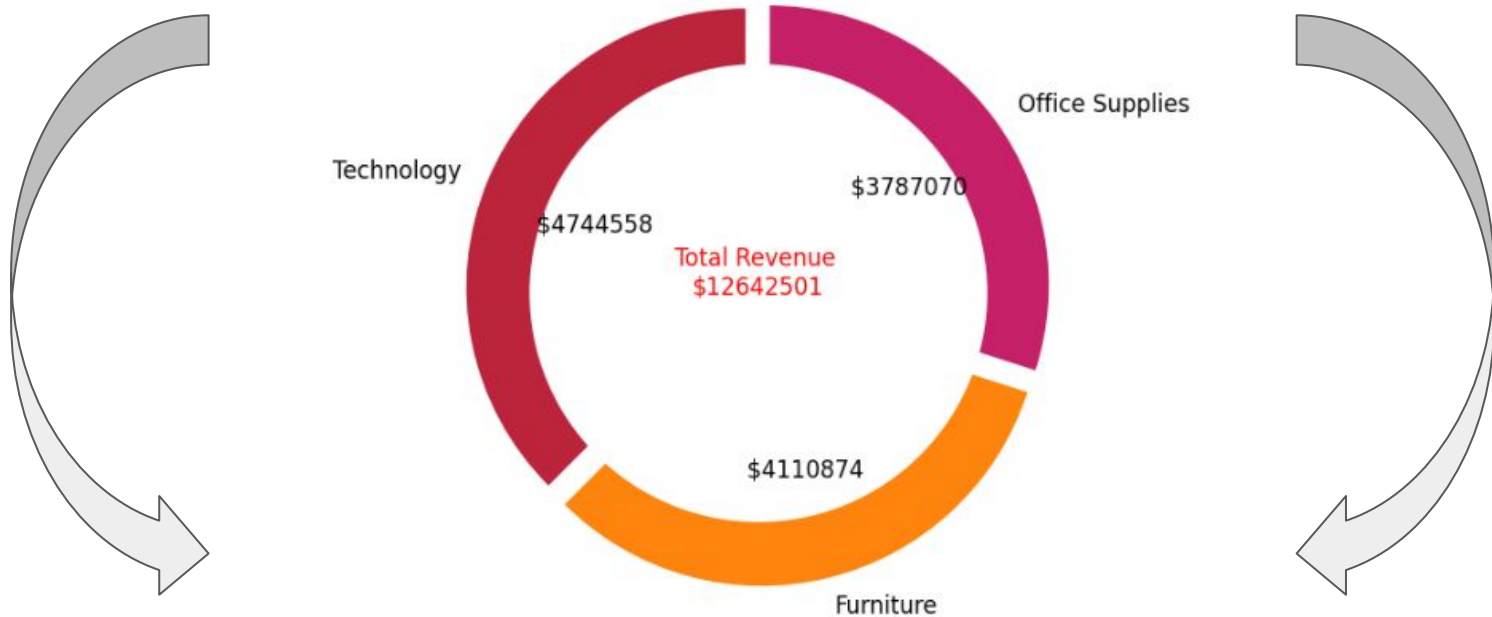
Our Top 10 Most valuable Customers with total revenue.



Our TOP 10 States with Highest Revenue.

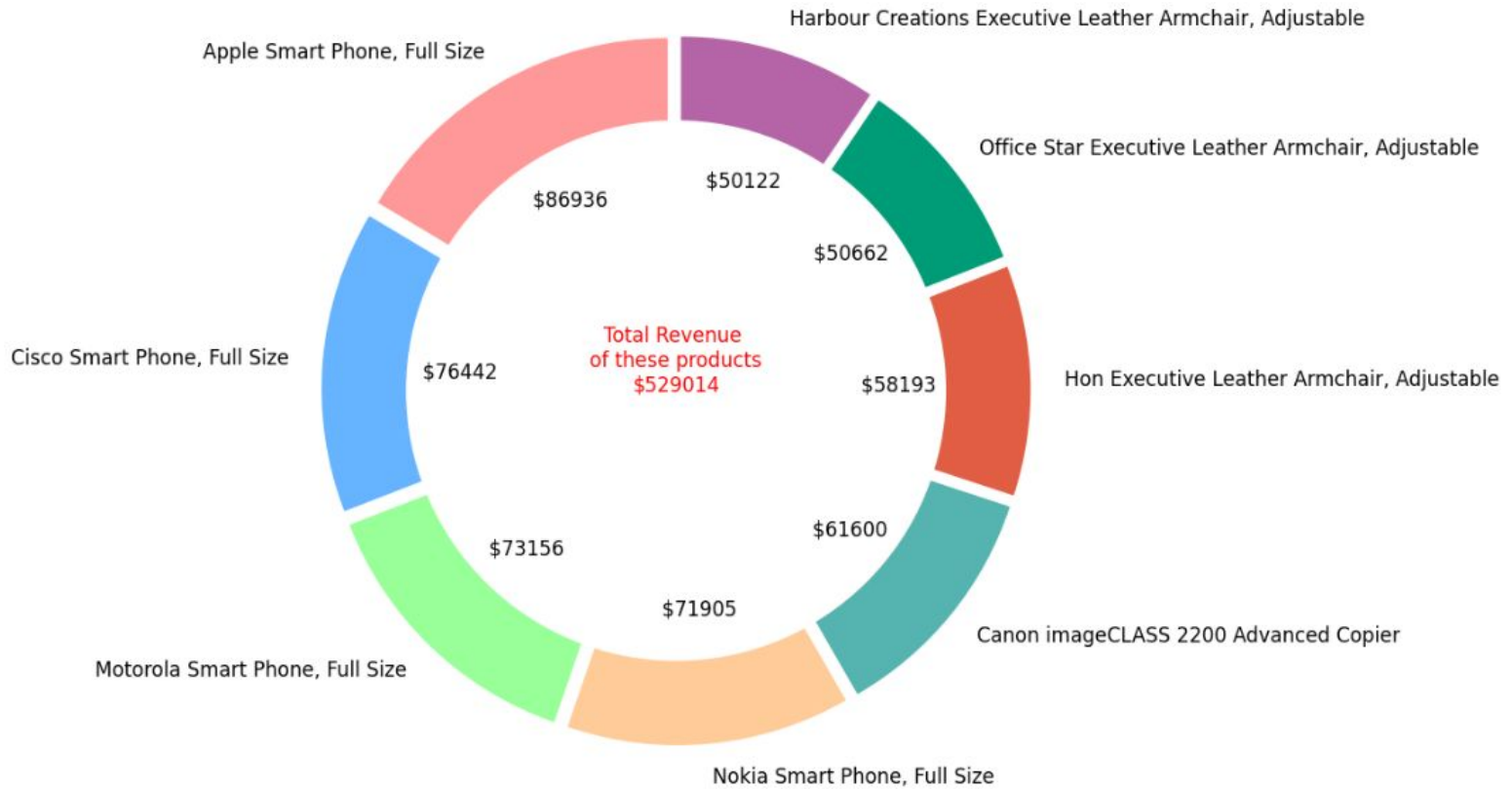


Top 10 Cities That Generated Highest Revenue.



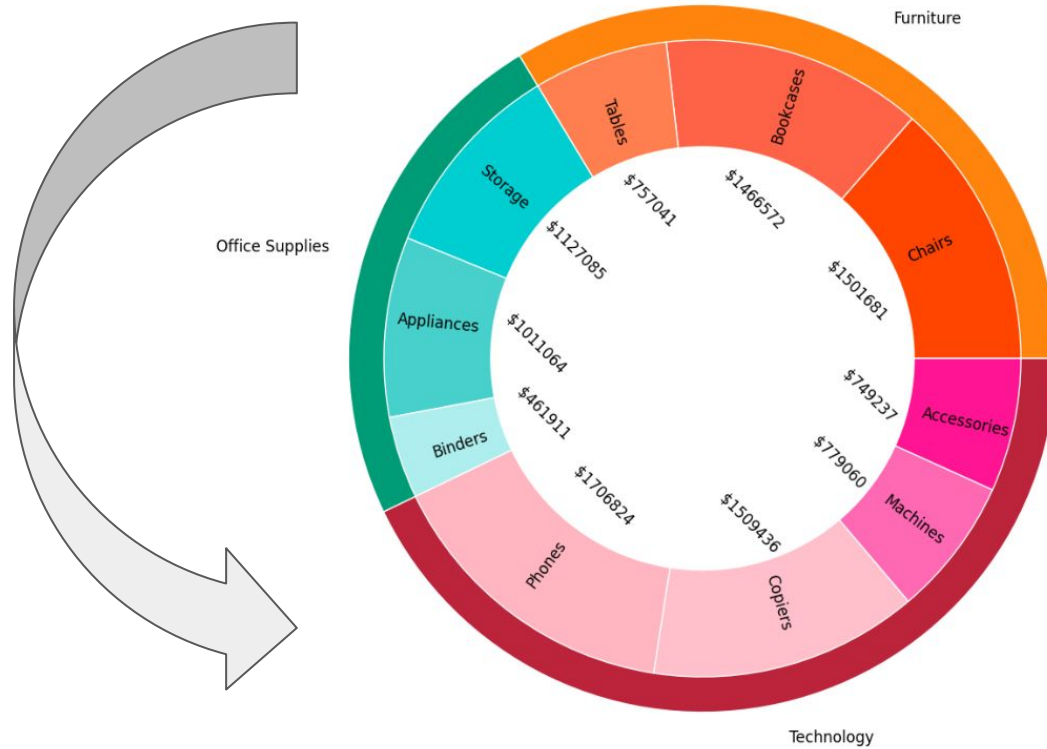
We can see that Category - Technology generated the highest revenue of about \$4744558!

The Total Revenue generated by all the categories - \$12642501!

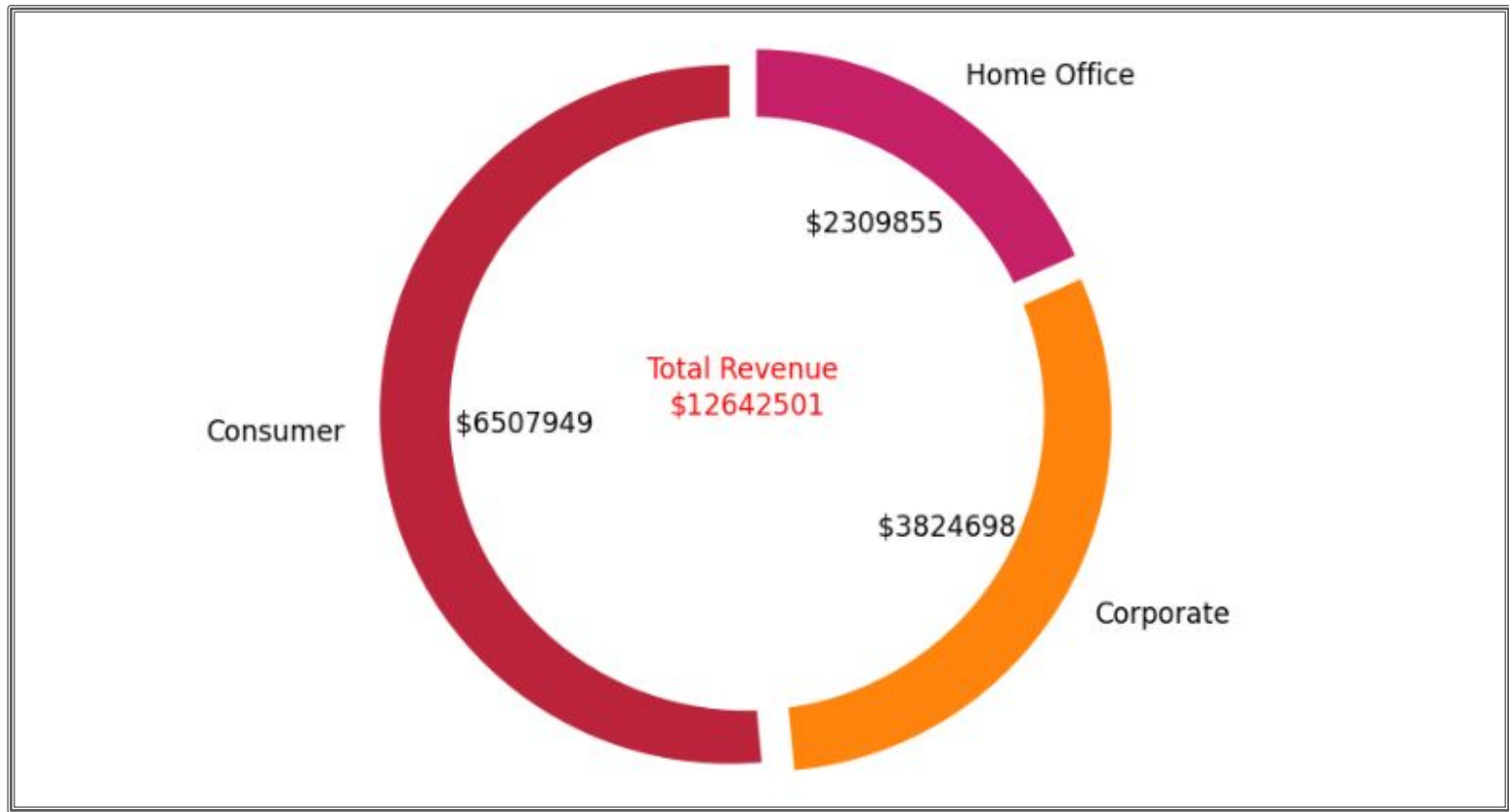


We can see that Product - Apple Smart Phone, Full Size generated the highest revenue of about \$86936!

The Total Revenue generated by all these products - \$529014!

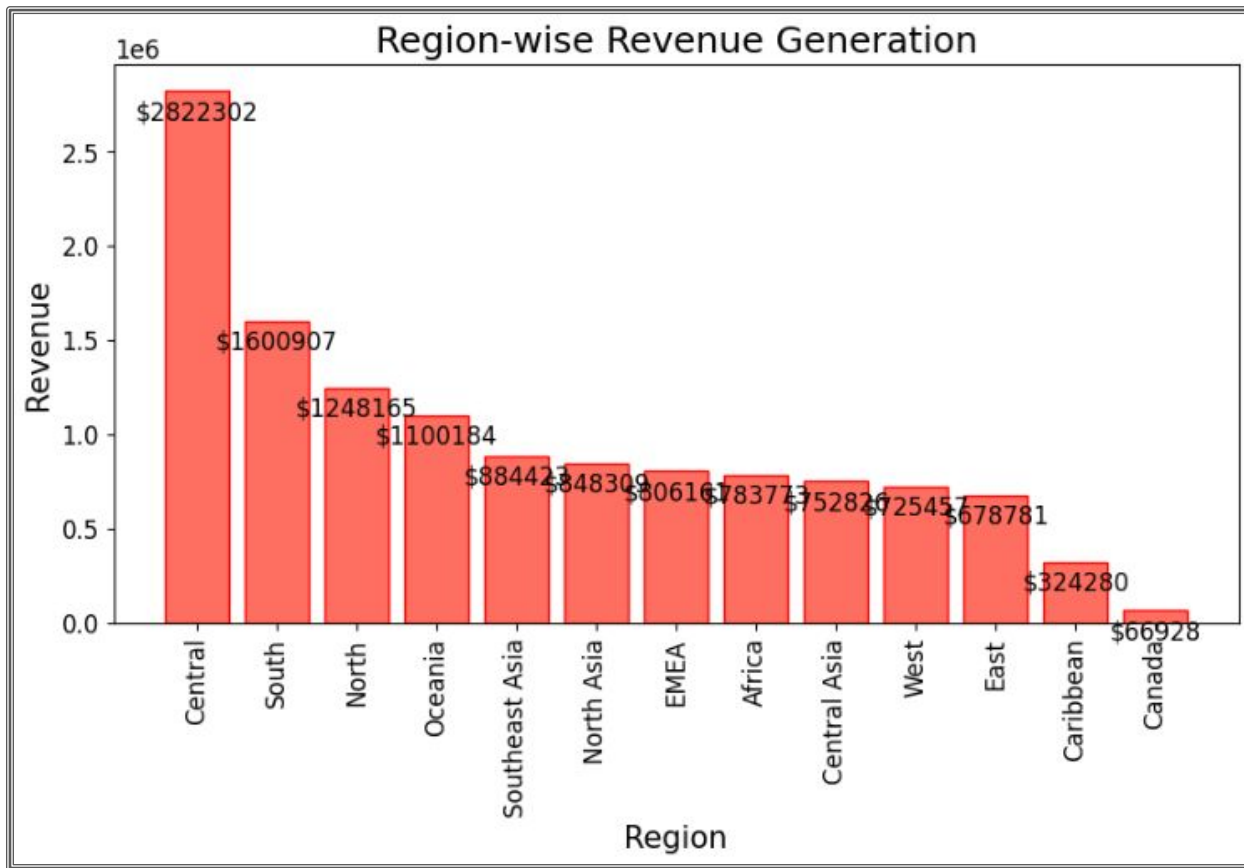


We can see that Sub-Category - Phones generated the highest revenue of about \$1706824!

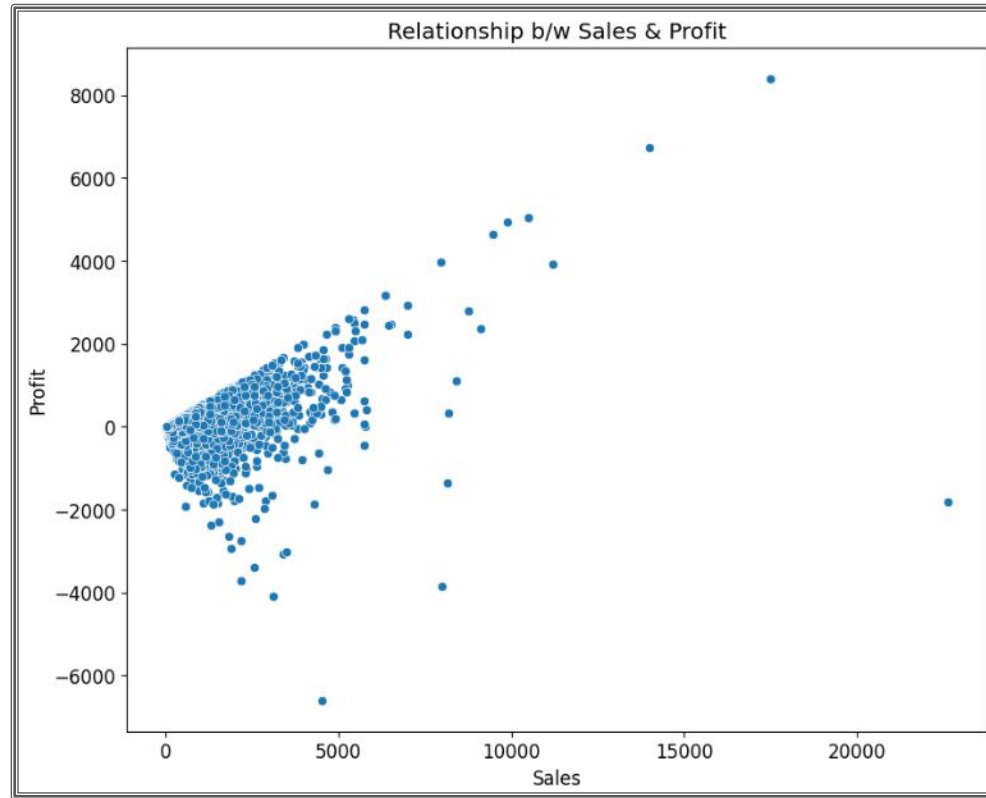


We can see that Segment - Consumer generated the highest revenue of about \$6507949!

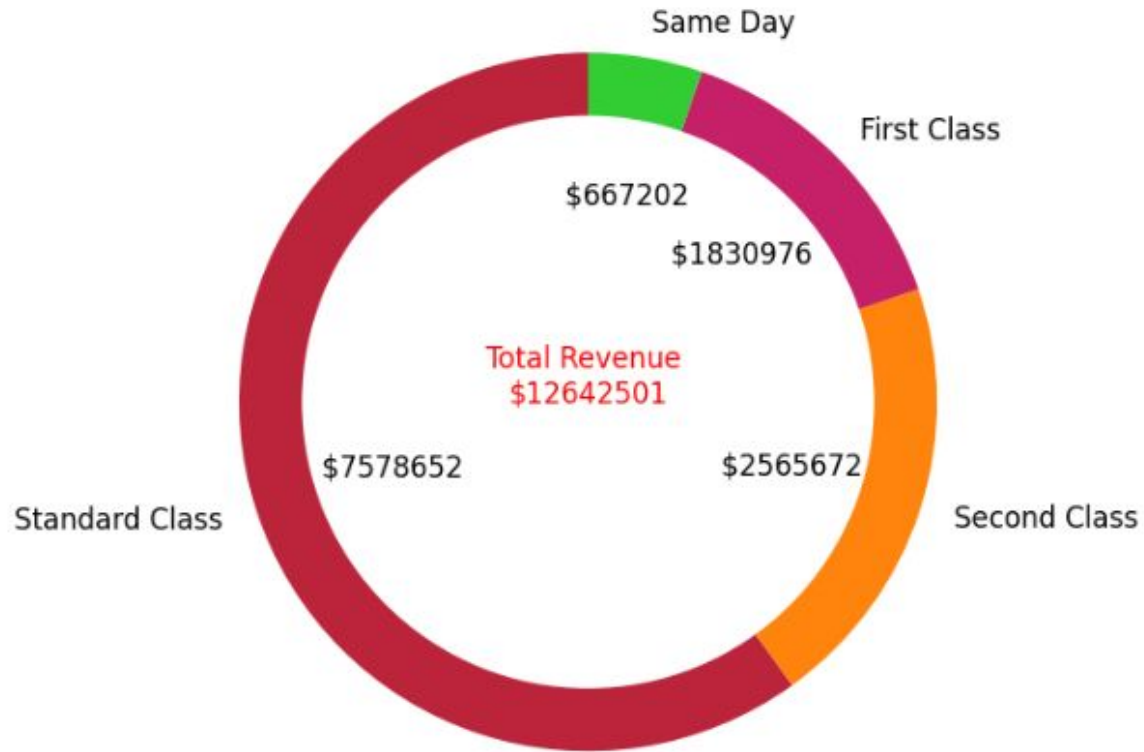
The Total Revenue generated by all the segments - \$12642501!



Region Central and South Have Generated The Most Revenue.



Relationship Between Sales And Profit As we Know Whenever Sales Increase Profit Also tend to increase.



We can see that Shipping mode - Standard Class generated the highest revenue of about \$7578652!

The Total Revenue generated by all the shipping modes - \$12642501!

1. Which region has the highest sales and profits?

Central Region and Central South are the regions that have the more profits.

2. What is the distribution of sales across product categories and sub-categories?

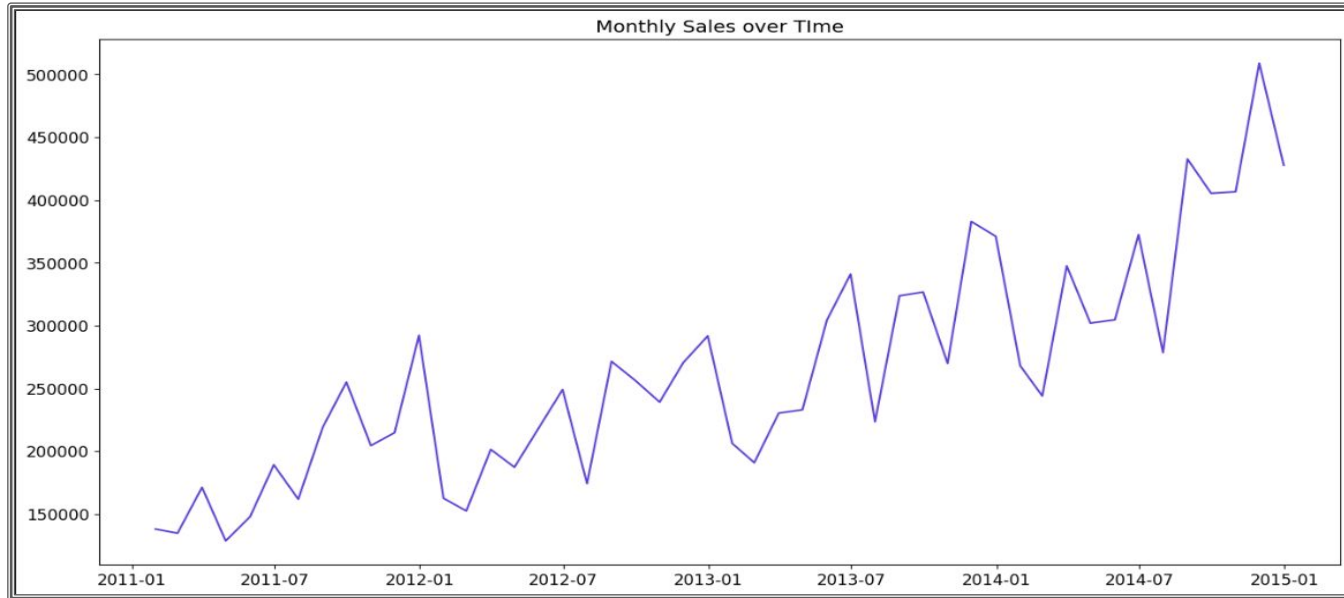
Sales of Technology performing very well and sub-category of technology also performing very well.

3. How is the relationship between sales and profit?

We can see as the sales increases, the profit also increases and it has the positive Correlation.

4. How have monthly sales changed over time?

From the analysis of the monthly sales trend, it is evident that sales have been consistently increasing with the passage of time. However, there are also periodic fluctuations in the trend which could be attributed to factors such as seasonal demand or promotional activities. It is important to note that sales are not uniformly distributed throughout the year, and certain months may see significantly higher sales than others. By identifying these patterns, the company can optimize their marketing and inventory strategies to maximize profits.



How effective are the visualizations in answering these questions, and what improvements can be made?

The visualizations used in this project are very effective in answering the questions easily, the data used in this project had some outliers that could potentially skew the results. In future analysis, removing outliers could improve the accuracy of the insights derived from the data. Additionally, including more features could provide a better understanding of the factors driving sales and profits. On the other hand, removing less useful features could streamline the analysis and make it more efficient.

Kartik Dhiman