Super Store Sales Data Analysis

- Data analysis is an essential aspect of any business model as it helps to identify patterns and provide insights about the data in a concise and clear manner.
- It enables individuals to understand problems and make informed decisions based on the analysis.
- Without data analysis, it becomes challenging to understand the underlying trends and patterns in the data.

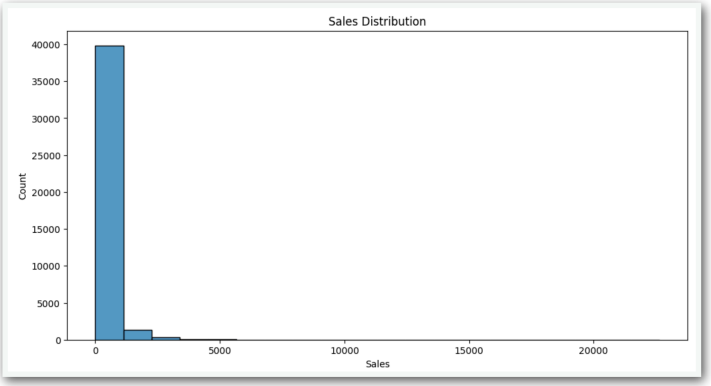
Here is the data, that we had for the sales analysis

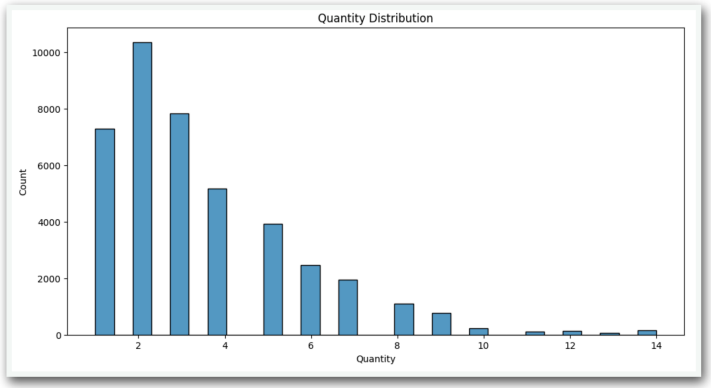
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority
0	42433	AG-2011-2040	1/1/2011	6/1/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	Algeria	NaN	Africa	Africa	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	Australia	NaN	APAC	Oceania	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	Hungary	NaN	EMEA	EMEA	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	Sweden	NaN	EU	North	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	Australia	NaN	APAC	Oceania	FUR-FU-10003447	Furniture	Furnishings	Eldon Light Bulb, Duo Pack	113.670	5	0.1	37.770	4.70	Medium

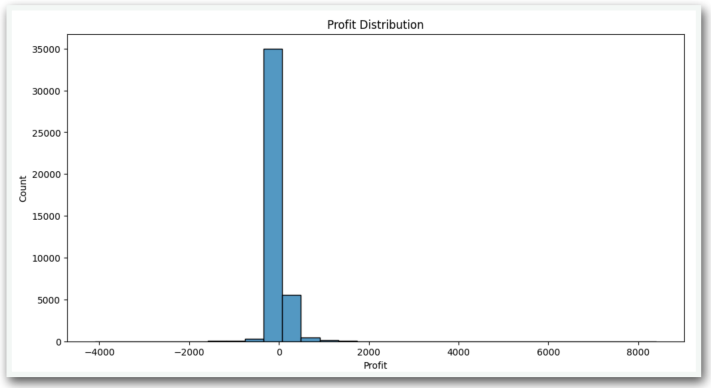
Data Description

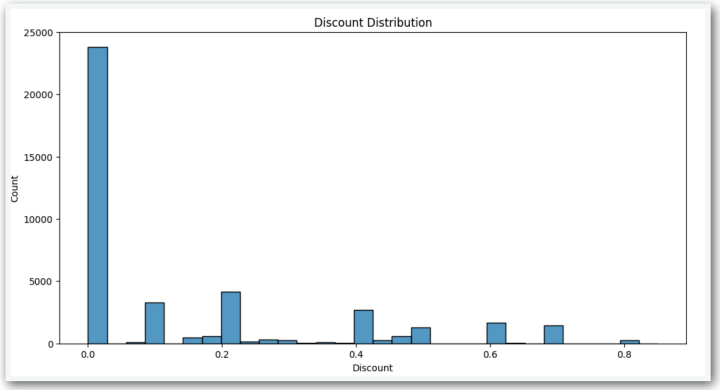
- Row ID: the unique identifier of each row
- Order ID: the unique identifier of each order
- Order Date: the date on which the order was placed
- Ship Date: the date on which the order was shipped
- Ship Mode: the shipping mode for the order
- Customer ID: the unique identifier of each customer
- Customer 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)

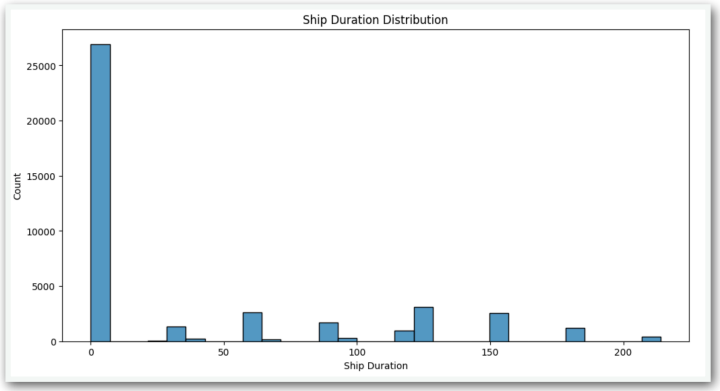
Exploratory Data Analysis



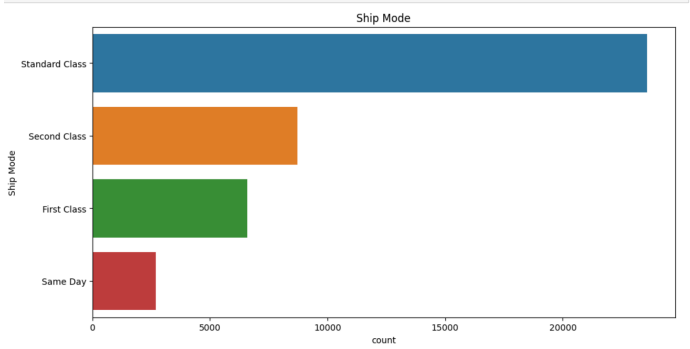




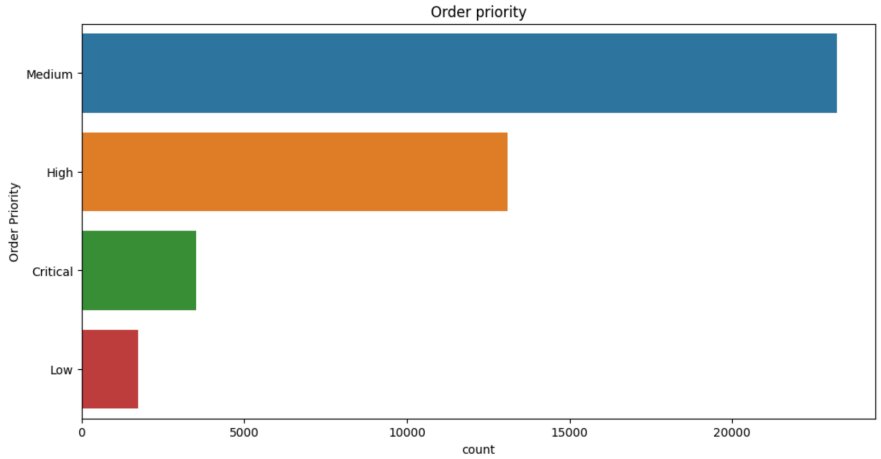




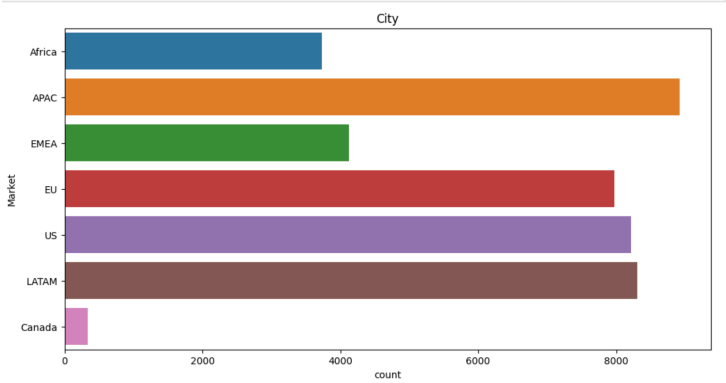
- we can see that most of the products (20000+) are standard class delivery.
- Following Second Class delivery (8000+)



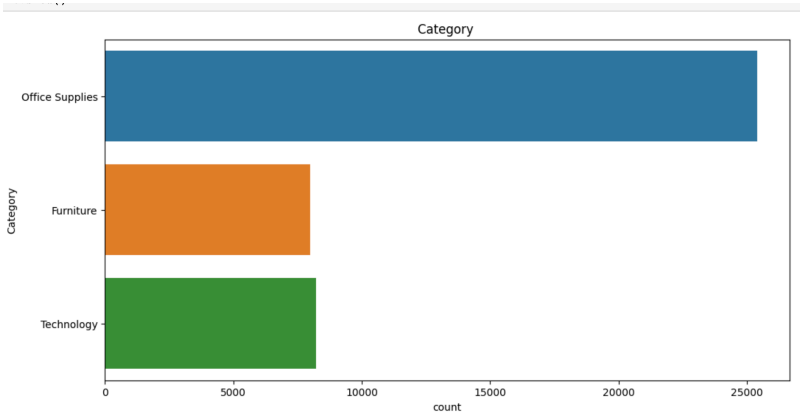
- we can see that most of the product priority is in medium and folowed by high.
- there are only few cases or counts where the priroity is critcal and low



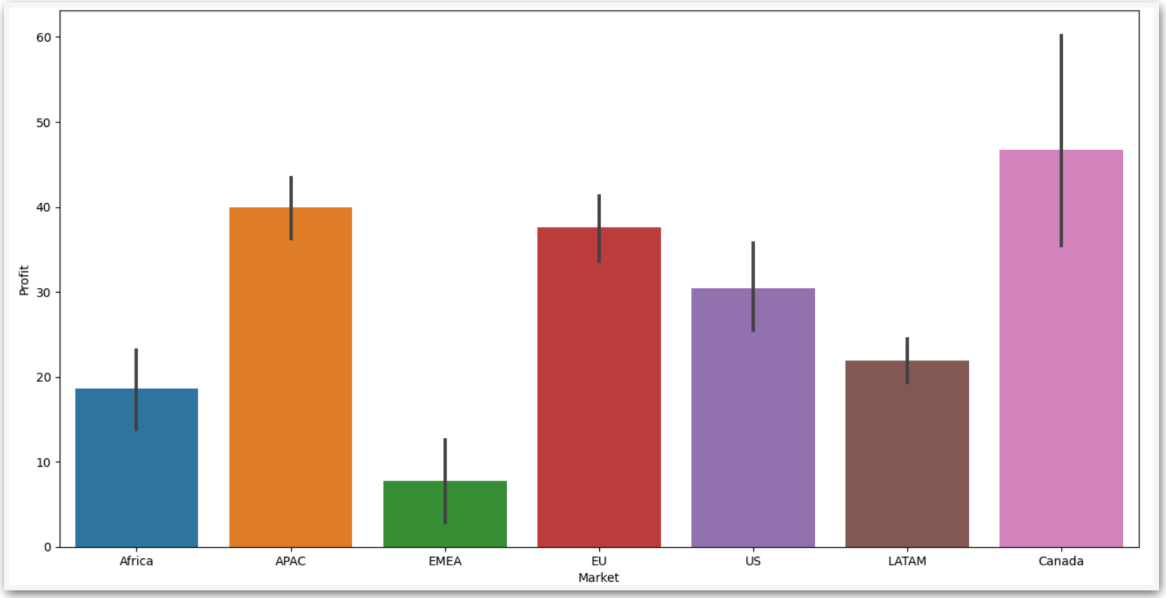
- we can see that the dominating market is APAC, followed by LATAM, US and EU.
- The least count in terms of market is Canada

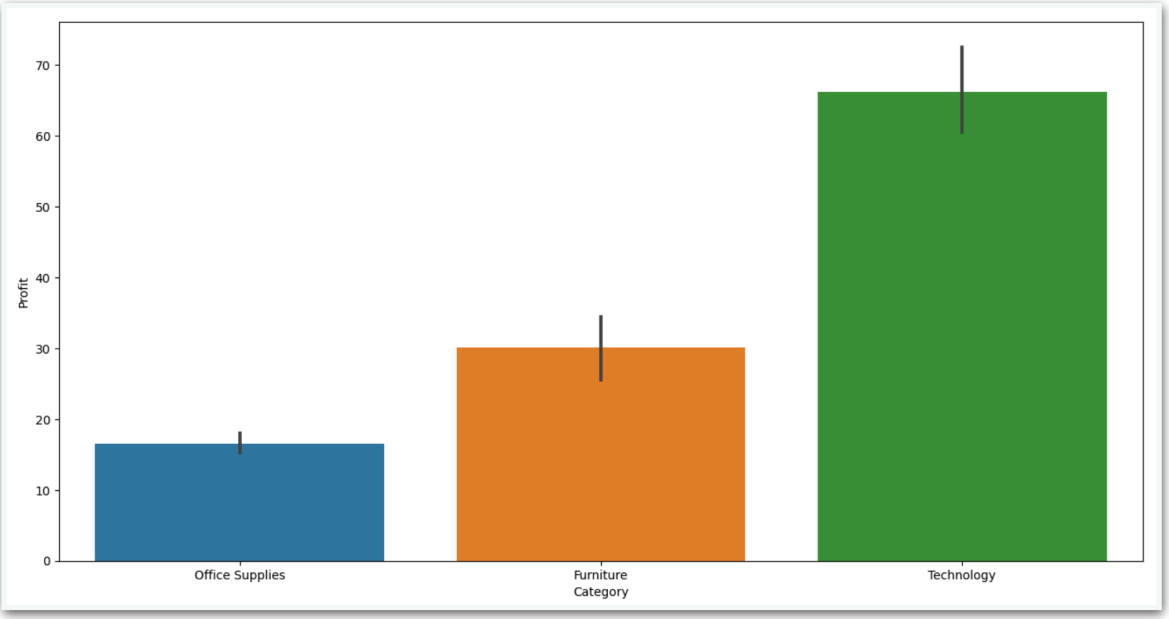


we can see that the majority of the products in the dataset belong to the Office Supplies category, followed by Technology and Furniture



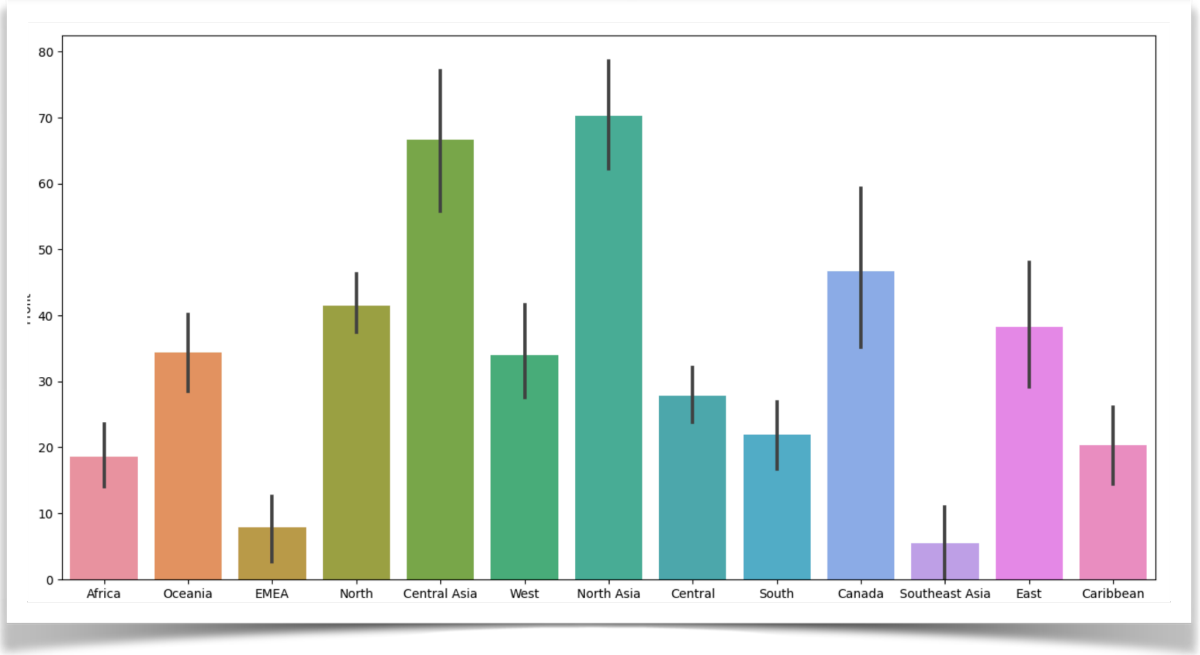
In terms of profit





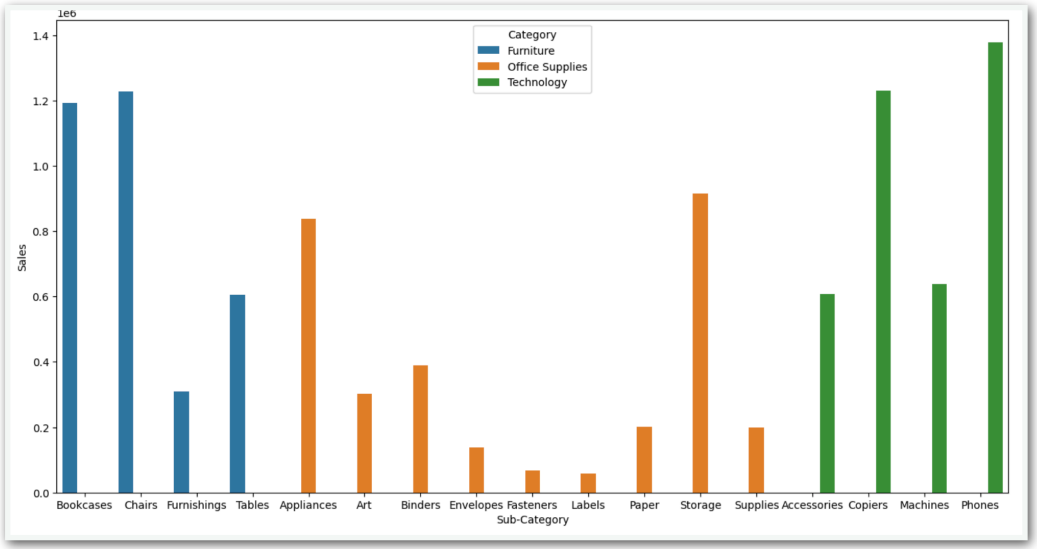
Which region has the highest sales and profits?

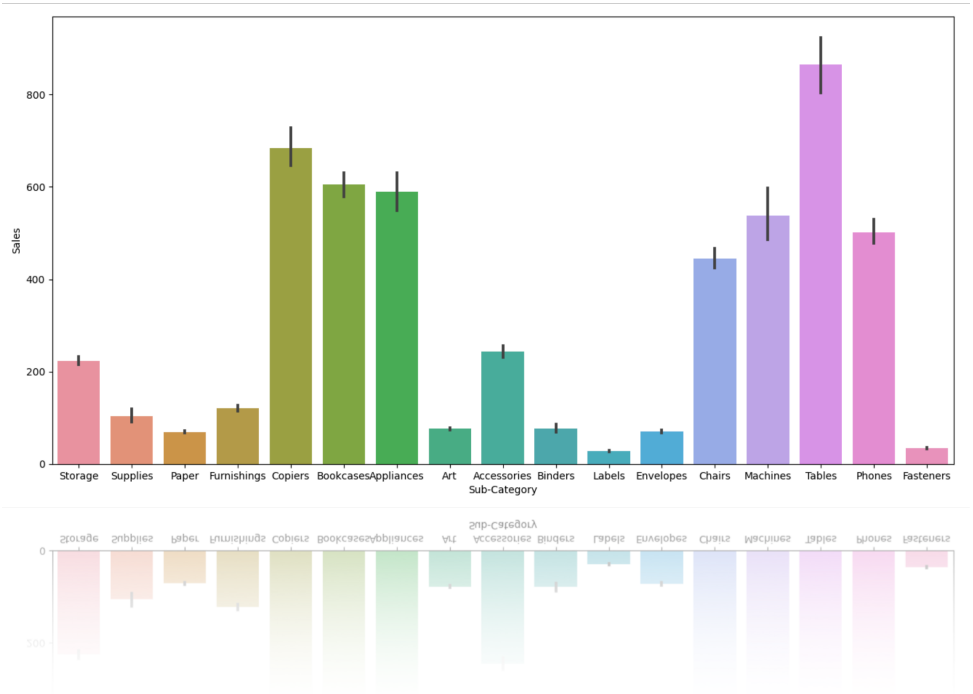
The North Asia Region and Central Asia followed by canada are the regions that having the more profits.



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

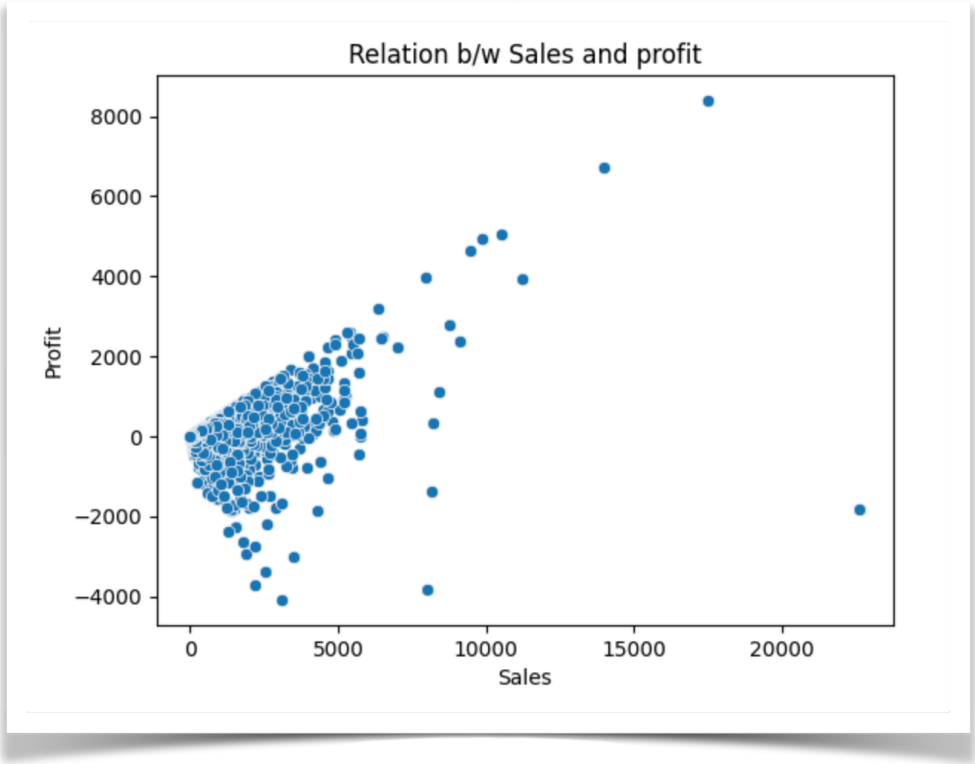
- most of the sales are from the technology having two peaks.
- the second most having sales are the Furniture also having the two peaks.
- When we compare with these two, the office supplies sales are less, and some are performing very low.





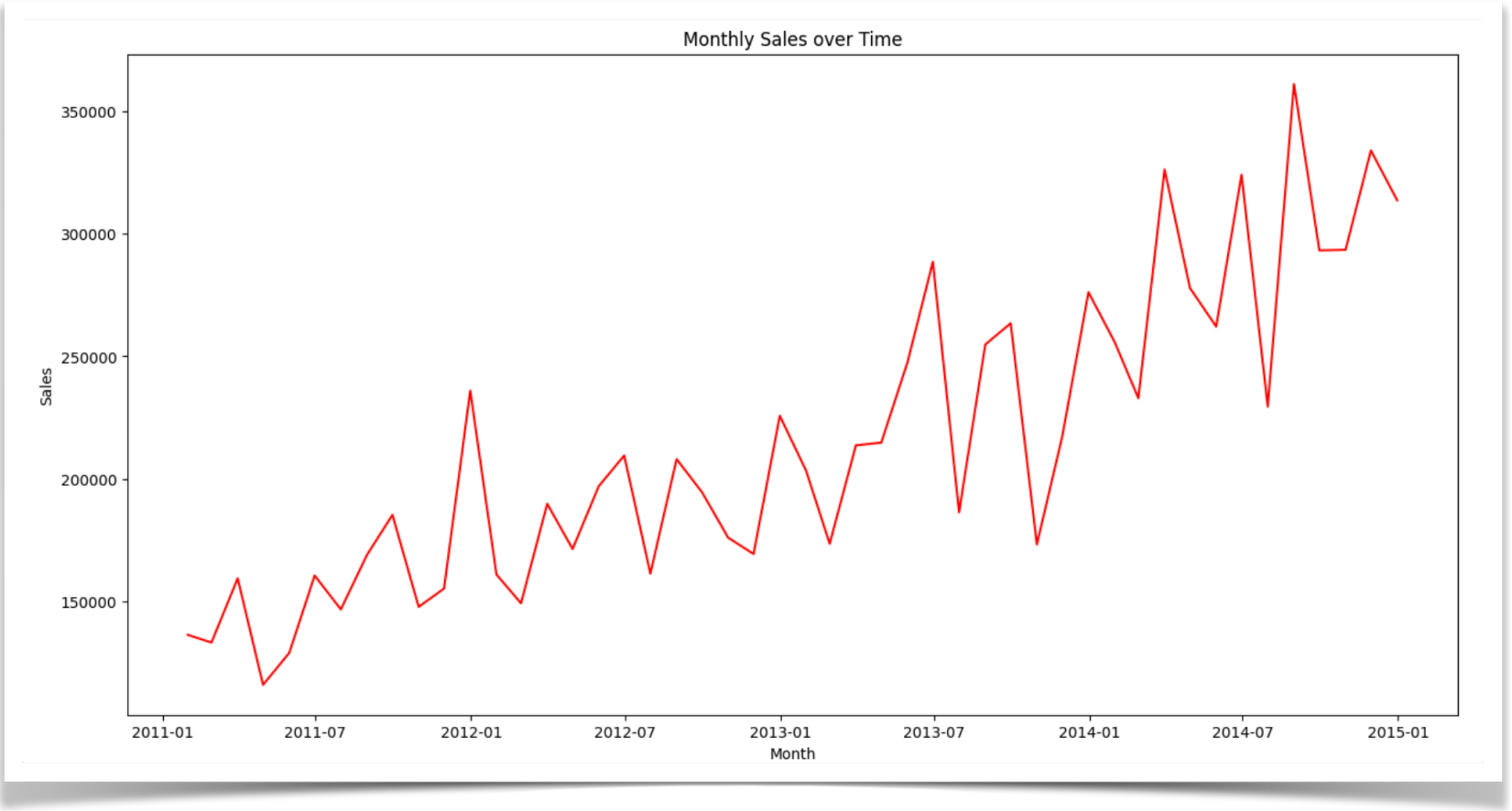
How is the relationship between sales and profit?

- As we already know, The sales are directly propotional to the profit
- Here as the sales increases, we can see the profit also increases and it has the positive Correlation
- There are some outliers, but we can ignore them, most of the data is linear, so we can say that the sales and the profit has a positive co rrealtion.



How have monthly sales changed over time?

- We can clearly see a trend over time, where the month increases the sales are also increasing.
- There are ups and downs in the trend, there might be some festival seasons where are sales are very high and remaining days are moderate.



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

visualizations are very important because they help us to understand the data easily regarding what background you are.

One picture is worth of thousand words

With the help of the visualizations we understood the data in a fast way, It is really hard for a non-technical person to understand the numbers and correlation etc, the visualizations like scatterplots, barplots, heatmaps, linemaps etc, are very helpful interms of understanding the data in a better way.

Regarding with these questions.

- Scatters plots are used to see the correlation between the features such as sales and the price.
- bar plots are used to understand the distribution in a better way, we have used for sales and profits.
- histograms are used to find the distribution of the data, and shows the outliers.
- Line plots are very useful and show trends over time or to compare multiple variables on the same scale.
-

Improvements in terns of data.

- the data has so-many outliers.
- we can add some more features, which may important for the analysis.
- we can remove some of the features, that are not useful for the analysis.