**Supermarket Data Analysis**

A close up of a toy store

Description automatically generated

By:

**Nanjesh Ramesh**

**CIN: 307402152**

**SUBMITTED TO**

**PROFESSOR SHILPA BALAN**

**Introduction**

## *****US Supermarket overview*****

Supermarkets, account for the largest share of food store sales in the US. In 2013, grocery stores accounted for 90% of the country’s food and beverage store sales, while supermarkets accounted for 95% of the total grocery store sales. There has been an increase in market concentration in the US Supermarket segment. The top four players, which accounted for 17% of the total industry sales in 1992, had a larger piece of the pie—around 36%—in 2013. Walmart Stores ([WMT](https://marketrealist.com/quote-page/wmt)), The Kroger Co. ([KR](https://marketrealist.com/quote-page/kr)), Safeway, and Publix Super Markets were the top four players in 2013. This increase in market share was due to the continuous consolidation of the supermarket segment.[1]

1. **Dataset URL:**

I have downloaded the dataset from Kaggle.com and data.world . These data sets display the situation in terms of sales, profit and the current market situation in the United States.

<https://www.kaggle.com/shobanama/superstore>

<https://data.world/stanke/sample-superstore-2018>

**Dataset Description:**

* The dataset contains the information of the superstore such as:
  + - Products (Name, Category and Product ID)
    - Sales
    - Profits
    - Quantity
    - Region (Country, State and City)
    - Customer
    - Shipment
    - Order Details
* Using these categories of data, we can come up with an analysis that can help the Superstore to make better decisions.
* By using this analysis, the Superstore can identify certain trends and figure out the possible solutions for the same.
* The Superstore can monitor the regions where they are earning more profit and concentrate on the regions where the profit is comparatively low.
* The Superstore can identify which product is best-selling and increase of those products in its inventory.
* It is also possible to identify the regular Customers by looking into the Customer Demographics.

1. **Data Cleaning:**

Data Cleaning or Data Cleansing is a process required to identify and correct the data which is to be analyzed. It is important to make sure of the quality of the data. Quality of data depends on following factors, the data should be consistent, complete, uniform, valid and accurate. Data can get corrupted because the data is sourced from a different source, and hence it might have different definitions in the source systems. It is important to have standard definitions for the data before we can analyze the data. If data is not cleaned, then there are chances your results will also get corrupted. It is also important to understand that data cleaning is not data validation which makes sure that no corrupt data can enter the system, whereas data cleaning helps in cleaning the data already in the system. For this project, I have used the following techniques to clean the data I sourced from Kaggle.

1. **Illegal Values:** The data in the Column Order Date consisted of Illegal values which were exceeding the bound.

**Before:**

**A screenshot of a cell phone

Description automatically generated**

**After:**

****

1. **Uniqueness Violation**

**Before:** In my datasets, I realized multiple values are in one column. Which are embedded values. I corrected them by removing the unneeded details.

**A close up of a newspaper

Description automatically generated**

**After:**

****

1. **Missfield Values:** In one of my table city names are in state columns and state names are in city columns. So, I corrected them by updating them with appropriate values.

**Before:**

**A close up of text on a white background

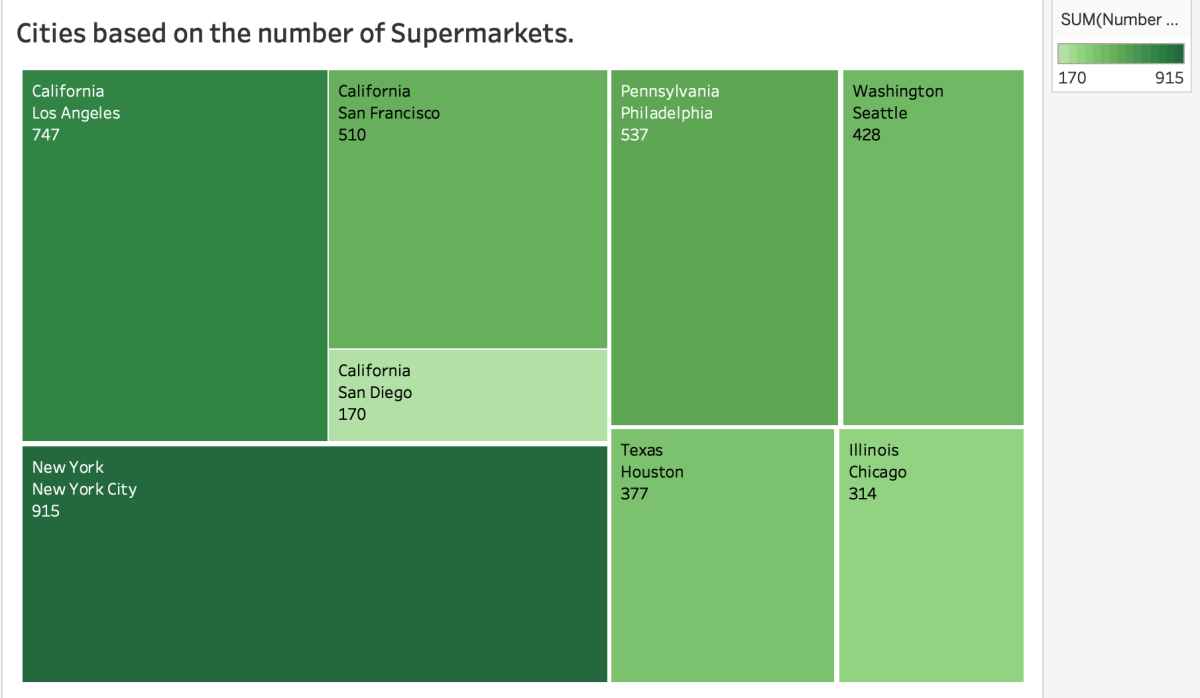
Description automatically generated**

**After:**

****

1. **Visualizations:**

**Question 1: Which are the cities with a greater number of Supermarkets?**

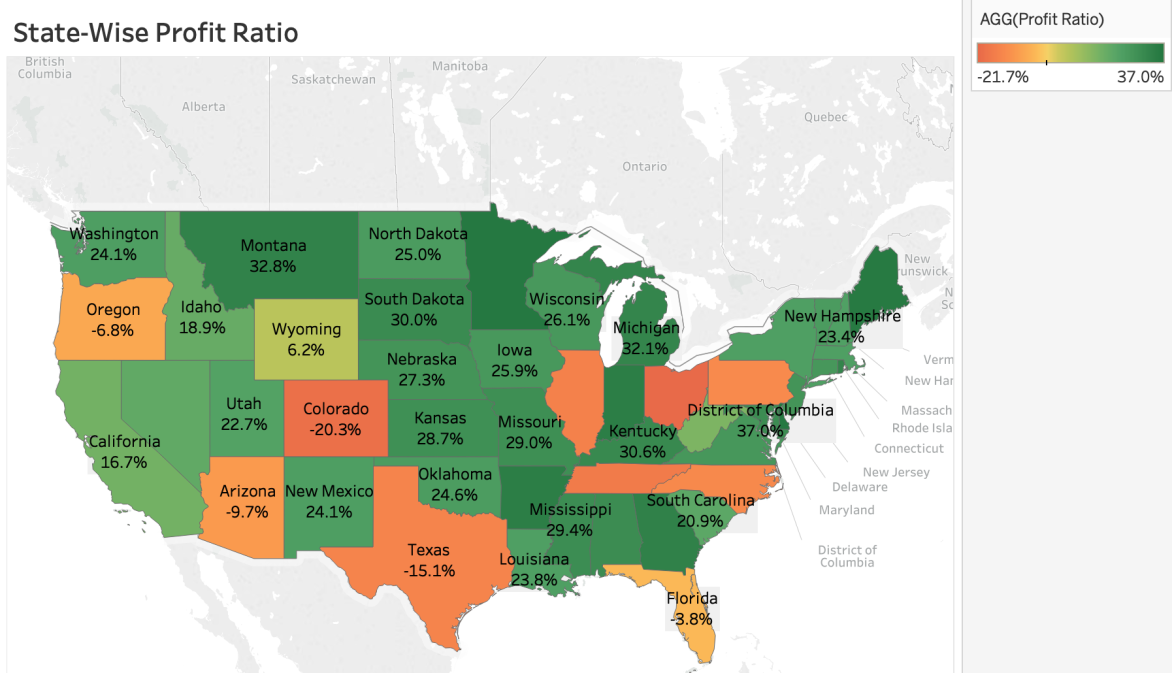


**Fig 1: (Feature: Tree map)**

In this visualization, I am showing the cities that are having a greater number of Supermarkets. I have used Tree map representation because the above feature can be depicted very precisely. From the visualization, we can observe that the City with the Darker colour and the larger size consists of more number of Supermarkets. In the above Tree map representation, the city of New York is having the greatest number of Supermarkets which is followed by Los Angeles. And the city with the least number of supermarkets is depicted by a lighter colour and smaller size that is, San Diego. [2]

Thereby, we conclude that in the United States we can find a greater number of Supermarkets per square mile in the City of New York and we can plan on increasing the number of supermarkets in the cities where the count of supermarkets is less.

**Question 2: What is the Profit Ratio for each State?**

****

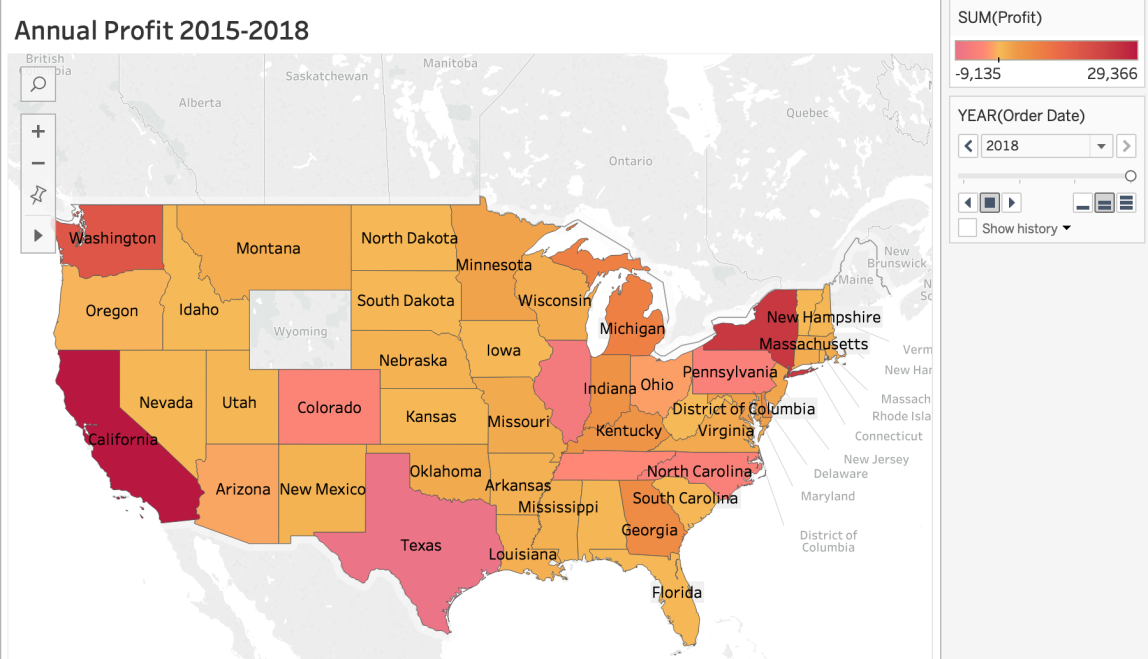
**Fig 2: (Feature: Geographic map with Calculated Field)**

The simple definition of profit margin, also known as gross margin, whether for a supermarket or any other business, is the price they sell an item for, minus how much the store paid to make or buy that item. [3]

I have demonstrated the profit ratio for the Supermarkets in each state. In the above visualization I have used a calculated field called Profit Ratio which computes our purpose. We can also observe from the Geographic Map Representation that the Green colour represents high profit ratio and the red colour represents low profit ratio. From the Fig 2, it is evident that District of Columbia has more profit per margin for a product that is 37% and the least profit per margin is found in Ohio State that is -21.7%.

Thereby, we can conclude that the profit-margin is less in Ohio and it is better for the Customers to shop in Ohio compared to other states.

**Question 3: What is the Annual Profit throughout the United States from 2015 to 2018?**

****

**Fig 3: (Feature: Geographic map consisting of Parameter Control with Play Button Option)**

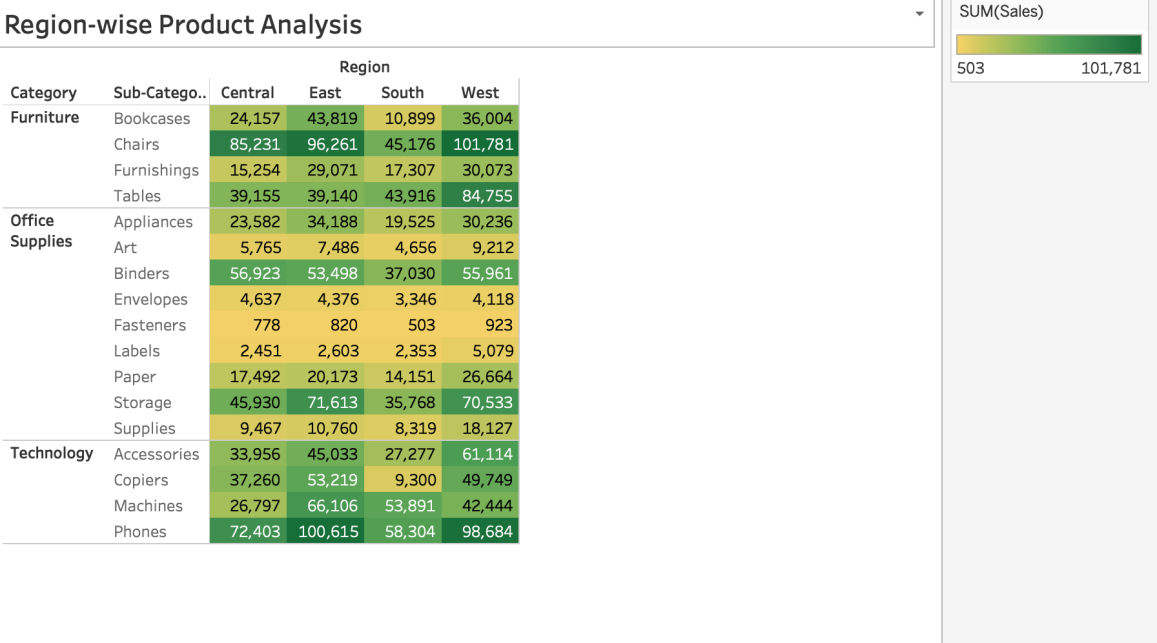
Any supermarket would monitor its progress at the end of the day or a week and so on. It is very important for the supermarket company to look into the yearly profits and make necessary decisions to improvise on the areas where they need to do well and focus on regions where they can improve their profits. [2]

The above visualization depicts the Geographic Map which represents the Annual Profit throughout the United States. I have used Play option in the Parameter Control which allows the slideshow view of the annual profit. In the above visualization the dark red colour indicates profit and the pink colour indicates the loss. For the year 2018 we can say that the state with the highest profit is California and the state which suffered a great loss is Texas.

Thereby, we can conclude that the supermarket company can work on areas in Texas to enhance and improvise their profit. They may have to take alternative methods and follow strategies to improve the performance of the supermarkets for the upcoming years. [2]

**Question 4: What Category of Product is in demand in each Region?**

Market demand describes the demand for a given product and who wants to purchase it. This is determined by how willing consumers are to spend a [certain price on a particular good or service](https://blog.blackcurve.com/all-about-the-abcs-always-be-changing-your-pricing). As market demand increases, so does price. When the demand decreases, price will go down as well. Market demand is the total of what everyone within a specific industry desires and can help guide merchants when [building an ecommerce site](https://www.bigcommerce.com/essentials/features/). [5]



**Fig 4: (Feature: Tabular Representation)**

In the above tabular representation I have used the columns such as Region, Category and Sub-category so as to analyse the demand of a particular commodity in each region. The darker the colour means that the product is in more demand which means that it is frequently sold. Whereas the lighter colour means that the product is in less demand. From the above visualization we can conclude that the regions sold majority of these Sub-categories:

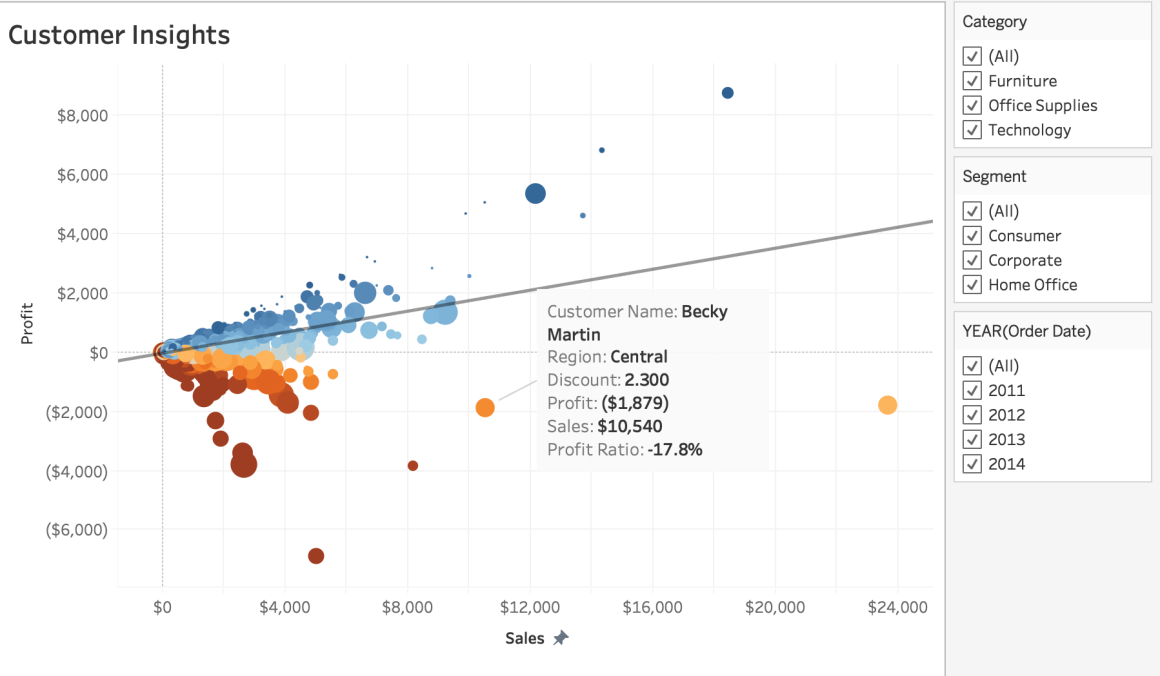
Central: Chairs

East: Phones

South: Phones

West: Chairs

**Question 5: What is the possible Customer Analysis that could be performed?**



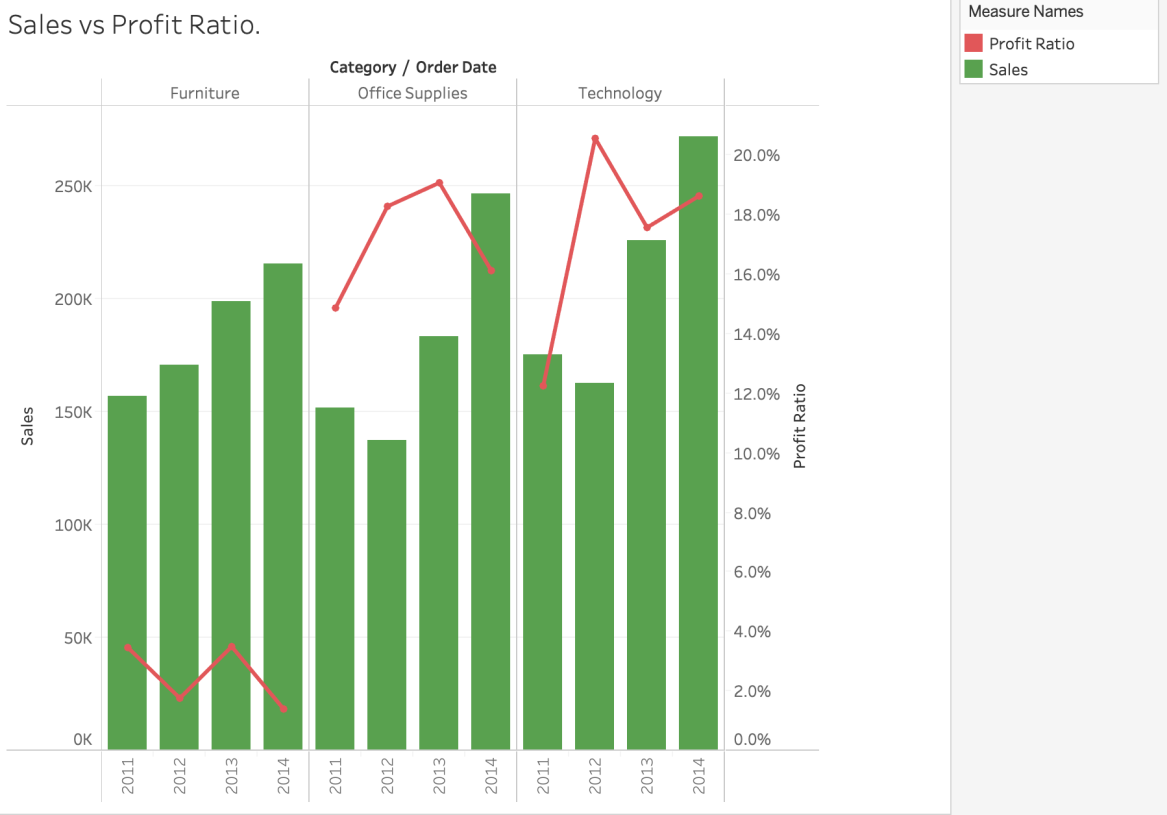
**Fig 5: (Feature: Scatter Plot with Trend Line, Calculated Field and Point Marking/Annotation )**

Customer analysis is vital to any effective business strategy. If a business doesn’t know who its customers are or what its customers want, it can’t meet customers’ needs. A customer analysis will do three main things:

* Identify the target customer
* Understand the needs of the customer
* Show how the company’s product or service meets the customer’s needs. [6]

In the above visualization I have used Scatter Plot representation along with Trend Line, Calculated field (profit ratio) and point marking/annotation. I have marked a customer called Becky Martin for whom the details such as discount, profit, sales and the profit ratio is displayed.[5][6] With these types of visualizations, a company can have better understanding of its customers and aim to facilitate their demands and needs.

**Question 6: In which year was the Supermarket Company most Successful and in which Category?**

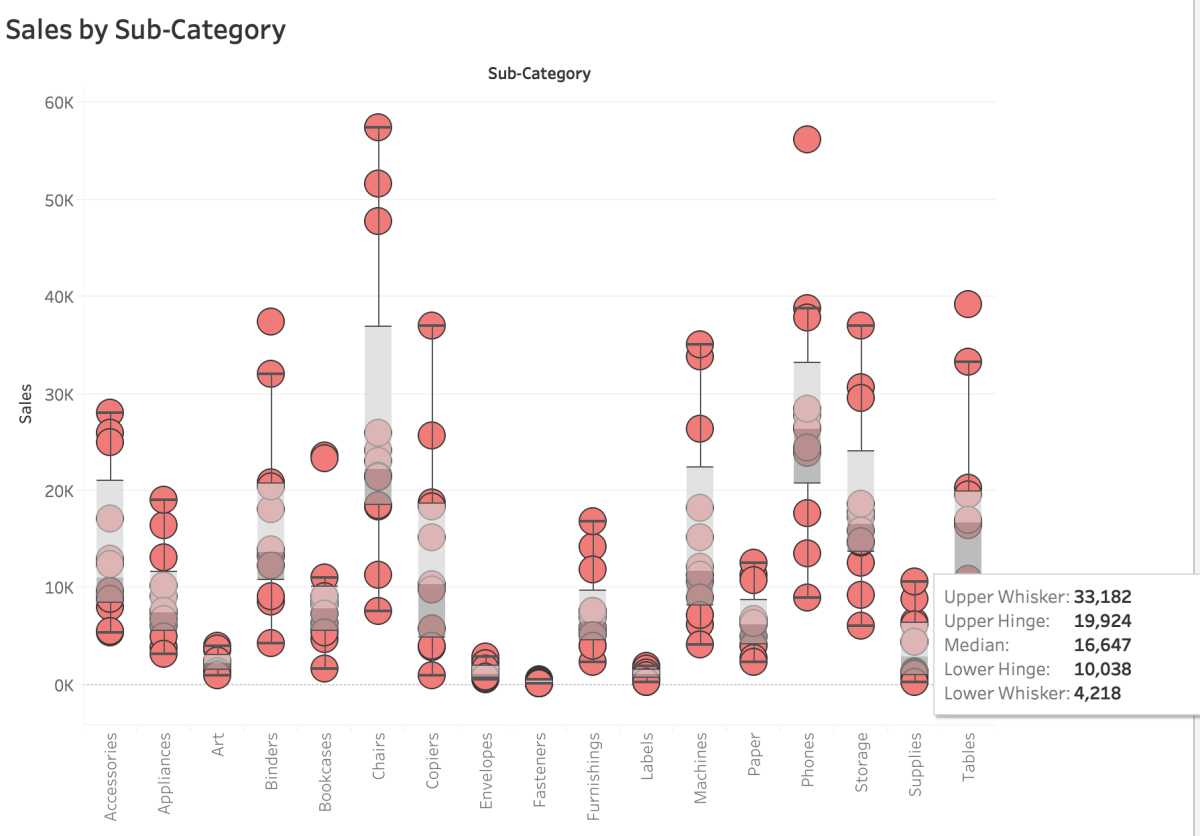
****

**Fig 6: (Feature: Dual Combination with Calculated Field and Dual Axis)**

The two most important characteristics which are required to analyse the company’s growth is the sales and profit ratio. This graph represents that profit-per-margin and the number of products sold. [3] In the above visualization, I have represented the sales and profit ratio in the Dual Combination with a Dual Axis of sales and profit ratio (calculated field). The green bar graph represents the sales and the red line represents the profit ratio for each subsequent year.

We can observe for the order category Technology, in the year 2014 both the sales and profit ratio are high. This means the Superstore Company did well in the field of Technology for the year 2014. Using visualizations like this it is easier for the Supermarket Companies to analyse their performance.

**Question 7: Which sub-category has a wide margin of sales and which has the least sales margin?**

****

**Fig 7: (Feature: Box-and-Whisker Plot)**

Each line on the box-and-whisker plot provides a piece of statistical context. The most important line is the one right in the middle of each “box”, which represents median. With median displayed, you can quickly look across the dimension members and compare medians, regardless of how big or small the range of values is within each column.

In the above visualization I have represented the sales and sub-category in a Box and Whisker Plot. The representation in this format would give us a median value using which we can recognize if our product is sold well or not.

We observe that for the sub-category chairs the sales margin is maximum. The median is 22,270. Any product with sales greater than 22,270 means that the product is sold well and if it is less, then it implies that the product didn’t do well in the market.

With this type of analysis, we can categorize products based on the margin of sales and group them accordingly. If there are a group of products whose sales margin is less, the Supermarket company can use new methods to improvise the sales of that particular product. [2][5]

**Question 8: Which is the Sub-Category that is mostly in demand or What are shoppers spending the most money on ?**

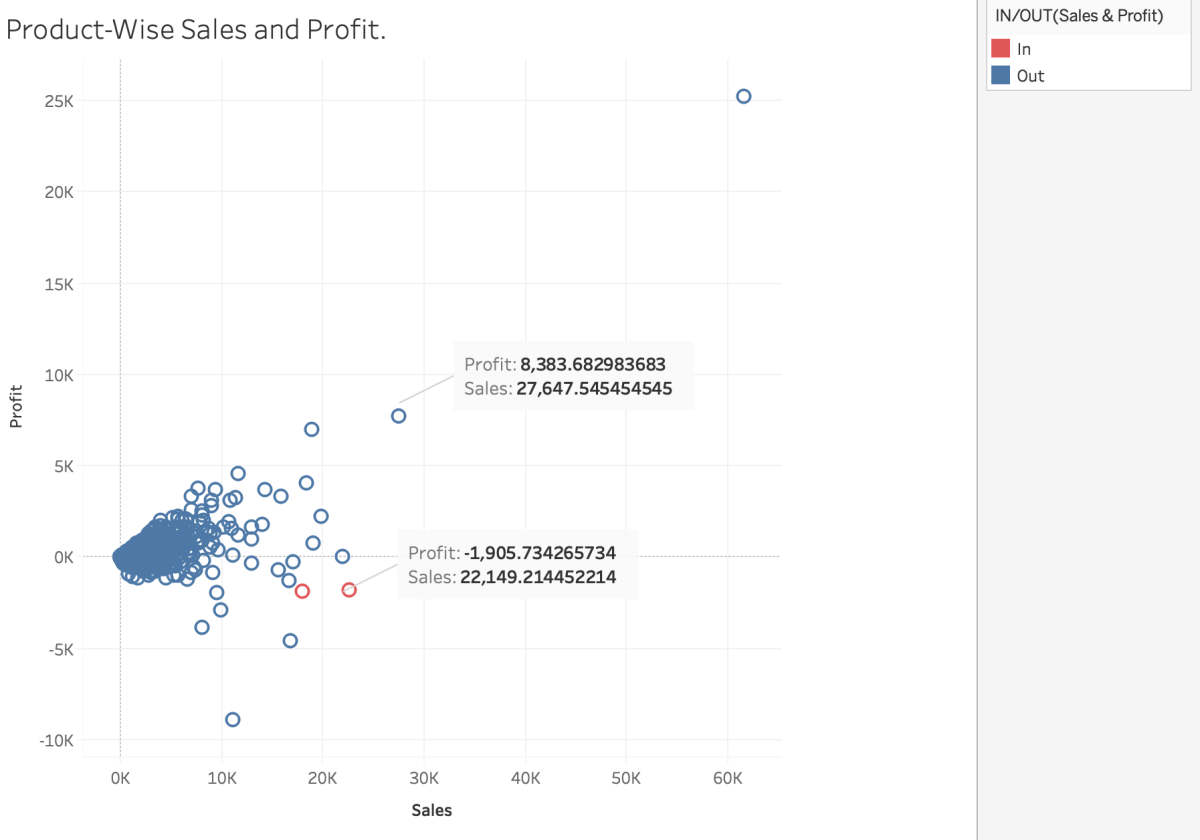
****

**Fig 8: (Feature: Dual Axis and Graph in the shape of Candy)**

In the above visualization I have used the Dual axis feature for the sum of Sales. We can observe from Fig 8 that the Sub-Category Binders has sold 84,266 items which means it is more in demand compared to the chairs, copiers, machines and supplies. The Sub-Category which is very less in demand is the office supplies, because it has only sold 17,030 items.

Based on the items which are in demand the Supermarket can stock their inventories and if the items are not in demand they could make sure that they first complete the current stock.

**Question 9: What Product-wise Sales and Profit ?**

****

**Fig 9 (Feature: Scatter Plot with Set and Point Annotation)**

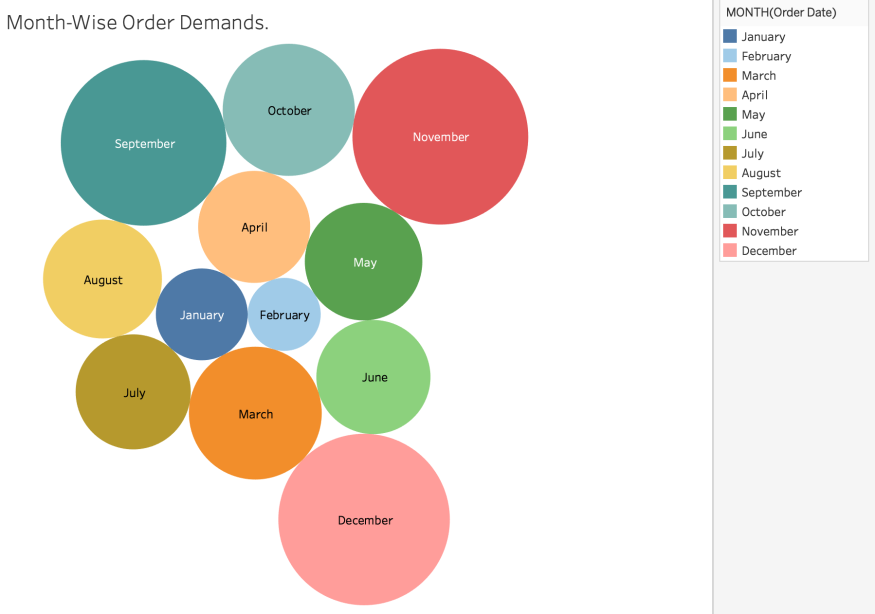
Performing sales trend analysis gives valuable insight into the inner-workings of a business. Merchants use their data to make informed decisions like when to raise or lower prices on your products.[4]

Sales trend analysis also helps to determine if the sales goals are met by providing an easy, measurable way to track the progress. You’ll actually know if you increased sales from last year and by what percentage. If you didn’t meet a goal, you can drill down to sales of a specific product or location to see what’s the factor for low sales.[4]

In the above visualization, I have used scatter plot representation with the combination of two sets to calculate highest sales and least profit. I have also used point annotation which gives the information of a point as shown in the above visualization.

For instance the above point in Fig 9, the Profit for a product is 8,383 and the Sales for the product is 27,647. Using these values, the Superstore company can track the progress of each product and carry on analysis accordingly.

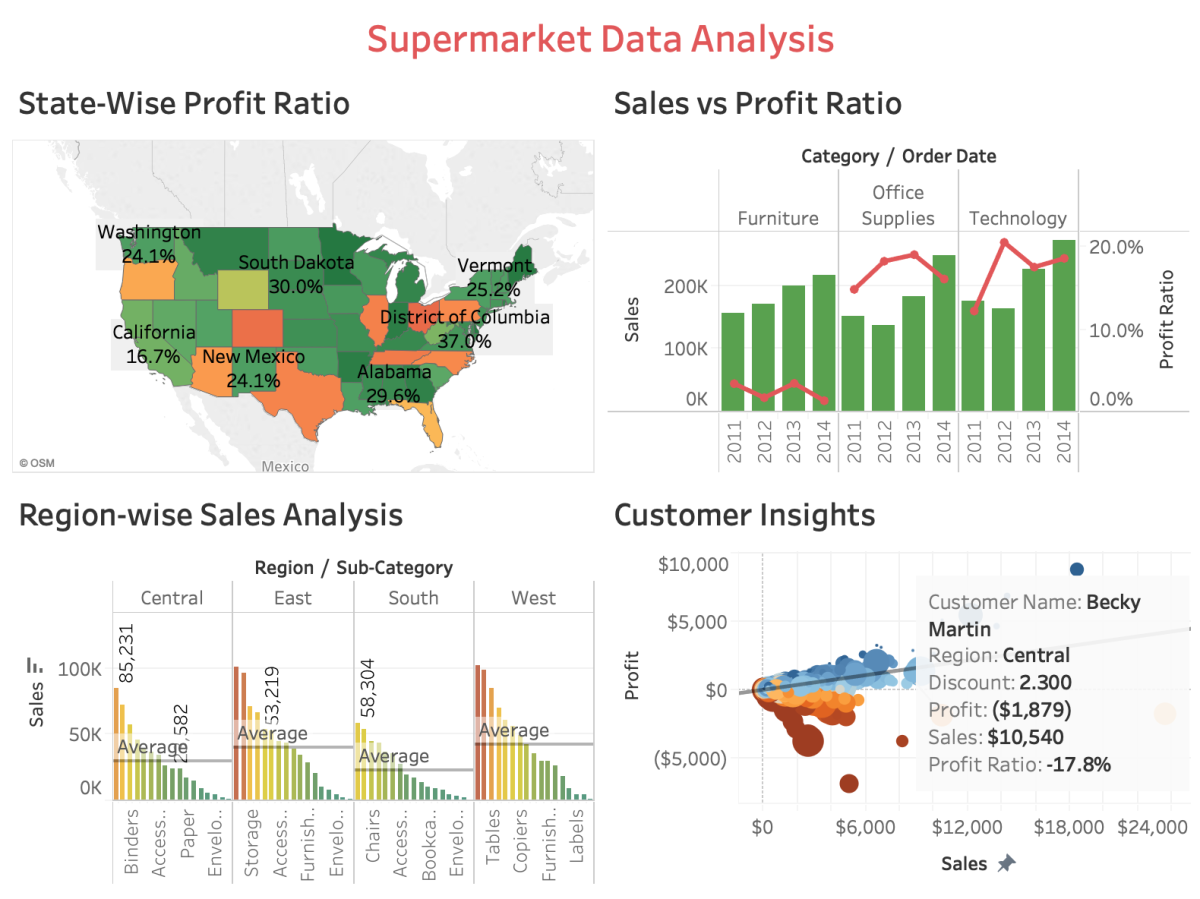
**Question 10: Which month of the year has the majority sales and orders?**

****

**Fig 10: (Feature: Packed Bubbles Representation)**

In the above visualization, I have shown the month wise sales analysis by calculating the sum of sales for each month and represented it using the Packed Bubbles representation. The larger the size of the bubble means greater the sales. And if the bubble size is small it means that the sales is lesser. In Fig 10, we observe that there are more sales and orders in the month of November and the least number of orders and sales were recorded in the month of February. Using visualizations like these, a supermarket company can steady their shipment methods in the months which have recorded greater sales.

1. **Dashboard:**



**Fig 11: Dashboard with four Visualization.**

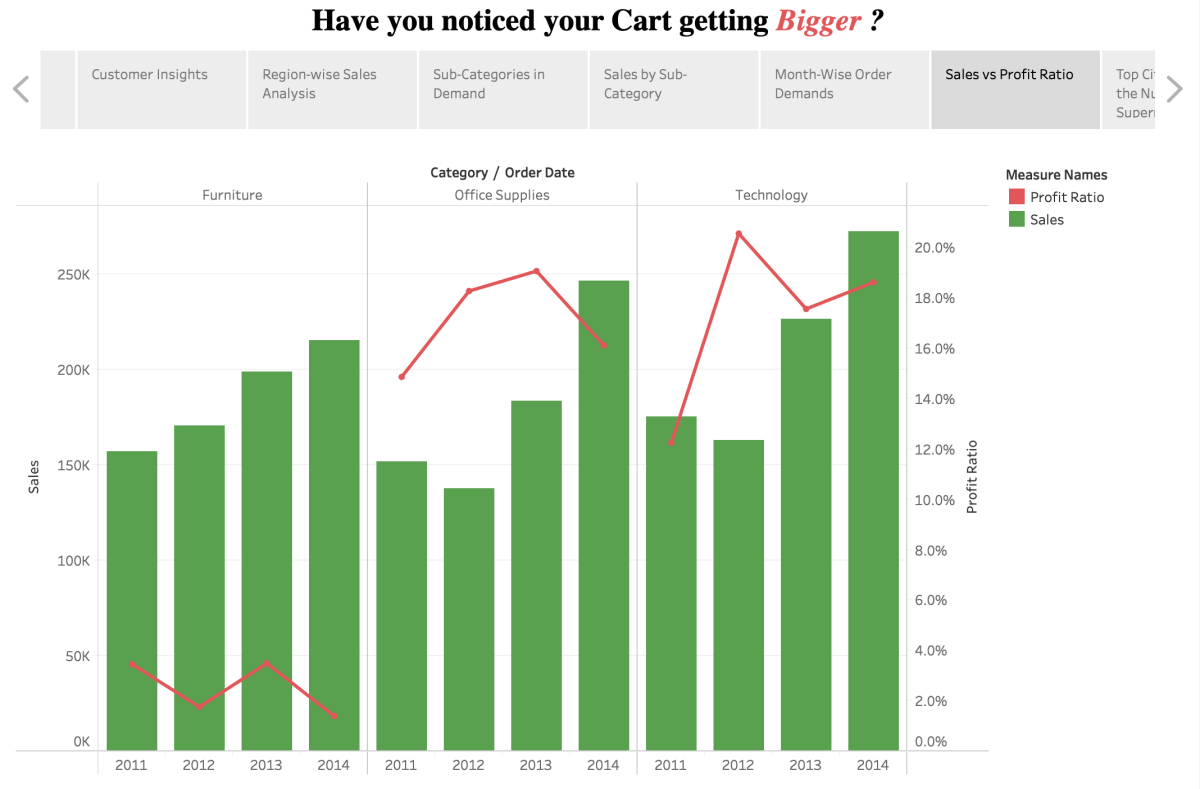
This is the Dashboard which gives us an idea of the Supermarket in the United Sates with the four visualizations. The first geographical map representation demonstrates the state-wise profit ratio for the Supermarkets in the United States. It is evident that District of Columbia has more profit per margin for a product that is 37% and the least profit per margin is found in Ohio State that is -21.7%. The second visualization shows the sales versus profit ratio and we can observe for the order category Technology, in the year 2014 both the sales and profit ratio are high. This means the Superstore Company did well in the field of Technology for the year 2014. The third visualization shows the region-wise sales analysis. Knowing the market sales demand can help inform future online businesses what industry is most profitable to enter into. Therefore, many business owners will have to conduct market demand research. Marketing research involves seeking out studies, data and general information about [an industry or sector](https://www.bigcommerce.com/blog/ecommerce-trends/). [5]

In the horizontal bar graph representation I have added a reference line which marks the average sum of sales for each sub-category of products. Any product which crosses or goes above this line is considered to be good selling and the ones below the line are considered to be not so good selling product in the market.

For instance, we can observe that for the East region the subcategories such as Chairs, Machines, Copies and Bookcases have sold well as they are above average sales. Whereas, the Appliances, Paper, Art and Labels have comparatively not sold well.

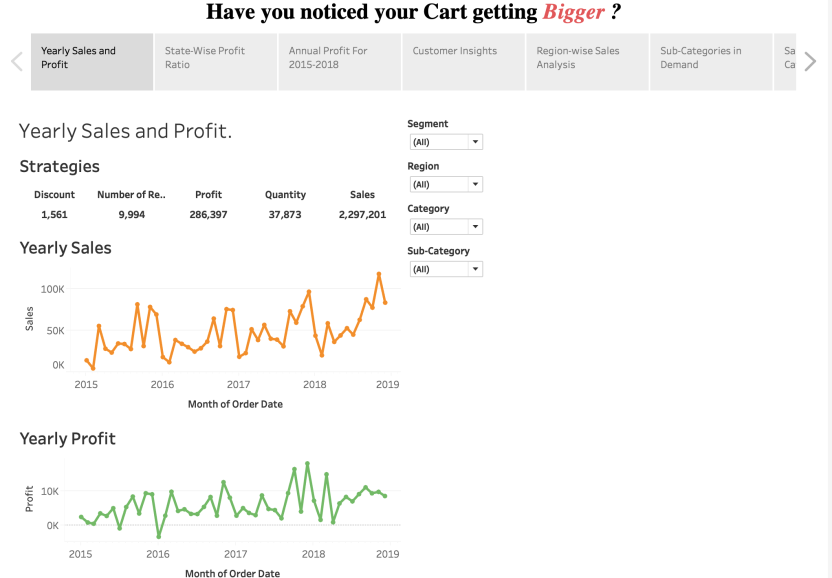
The fourth visualization helps in the process of Customer Analysis. The supermarket company can identify their regular customers using this type of visualization. For instance, in the above dashboard Tamara Chand is the most valuable customer or in other terms a regular customer for the Supermarkets.

1. **Storytelling:**



**Fig 12: Storytelling for Supermarket Data Analysis.**

When I first came to the United States in August 2018, me and my best friend Amogh went on a Grocery Shopping to Costco. We found that here the shopping experience was a little different compared to our home country, India. Before going to shop for groceries we used to make a list and we realised that each and every time our shopping cart got bigger as we ended up shopping few items that weren’t in the list as well. So, on the same day while we were coming back home me and Amogh had a great conversation about the Supermarkets in the United States.



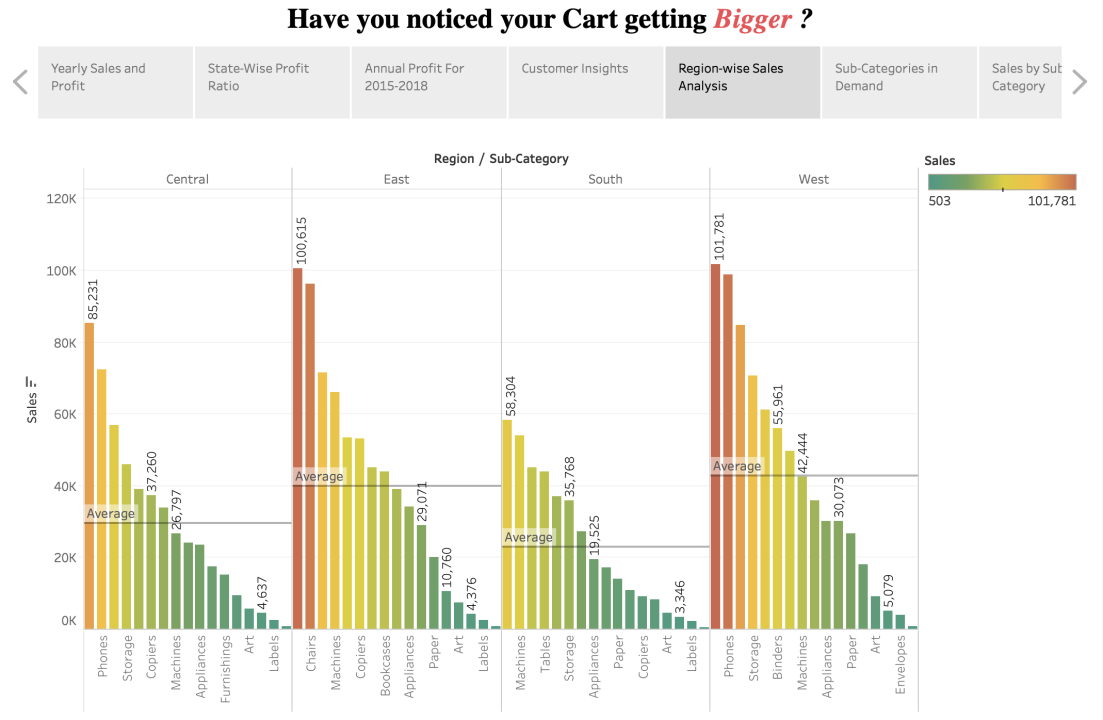
**Fig 13: Yearly Sales and Profit**

We started talking about the important aspects of the for a company’s growth and development, that is the sales and profit. We were curious to know how the supermarkets seem to be profitable after give us a great discount. The above Figure 13 which is my first page in the story, answered our questions regarding the strategies used by the supermarkets to enhance a very good sales of their project. The figure gave us an insight towards the ***yearly sales and profit*** for the Supermarket for all the Categories of Products sold by the same in all regions of the United States. Now we somewhat had an idea on how the framework for the Supermarket works.

Next, we were concerned with how the Supermarkets focus on each state when it comes to making Profit. We did speak about the profit-ratio which means, the price they sell an item for, minus how much the store paid to make or buy that item. So we understood, that with this margin called the profit margin is what really works for the Supermarkets. My second page in the story is related to the ***profit-ratio for all the states in the United States***. The Supermarkets have a certain amount of profit-margin for each commodity and based on this margin they decide the price for each product. We did conclude, that even when the products are sold for the discounted price it will still have a certain amount of profit-margin. I was amazed when my friend spoke about how they made a profit even during the times of Thanksgiving or Black Friday.

We went on talking about Walmart and Costco by keeping the annual profits as our perspective. With all the progress and performance of the Supermarkets, at the end of the year they record an annual profit. I knew about this and I explained to my friend on how the Supermarkets use strategies to improve their performance from the previous year and if they didn’t do well this year this will track back to their performance strategies that were used for the previous year. The third page in my story is related to the ***Annual Profit*** of the Supermarkets.

The next segment of our conversation was very interesting because it deals with how the Supermarkets see what their Customer needs and perform analysis on who are the regular customers and what are the type of discounts and rewards that could be made available to them. This is called Customer Analytics. The fourth page of my story gave us an idea towards ***Customer Insights***. I got to know about how company’s analyse the type of customer be it regular or periodic, how to effectively keep that loyalty of the customers by offering them good discounts and so on. It was really fascinating when my friend briefly told me how the Customer Analysis is performed.



**Fig 14: Region-wise Sales Analysis (Feature: Horizontal Bar Graph with Reference Line)**

Now we spoke about how the Supermarkets focus on different demands in the products in different regions. The fifth page in my story directed us to the ***Region-wise Sales Analysis***. We also understood that Supermarkets mainly focus on the region and then they decide on what type of products to deploy or sell in those regions based on the increasing demand. I gave an example to my friend about how in the East region the subcategories such as Chairs, Machines, Copies and Bookcases have sold well as they are above average sales. Whereas, the Appliances, Paper, Art and Labels have comparatively not sold well. He was surprised to know about this type of strategies used by the Supermarkets.

Moving on with our conversation I discussed about how companies focus on each subcategory when there are different varieties of the same. The sixth page in my story explains about the ***Sub-Categories in demand.*** The supermarkets have data analysts who are well versed in understanding what is the product that is currently in demand and how the Supermarkets have to cope with the rising demand, it may be something related to stocking their inventories, reducing the price on high selling products so that they increasing their sales. This topic extended to ***Sales by Sub-Category***, which is explained in the seventh page of my story.

Since it was the month of August my friend told me about how the Supermarkets record more sales in the year-end, majorly in the months such as November and December. We did realise that in November and December there are occasions such as Thanksgiving, Black Friday and Christmas which draw huge number of customers to shop because of the discounted sales. This contributes in a humongous scale to the Supermarkets profit. The eighth page in my story speaks about the ***Month-wise Order Demands.*** This when we understood that the Supermarkets make more profit at the year-end due to the discounted sales which attract a lot of customers.

We spoke about the two most important characteristics which are required to analyse the company’s growth that is, the sales and profit ratio. The profit-per-margin and the number of products sold. The ninth page of my story relates to the ***Sales versus Profit-Ratio***. We understood that using Data Visualization the Supermarket can monitor and understand their progress in a better way. In Fig 12, We can observe for the order category Technology, in the year 2014 both the sales and profit ratio are high. This means the Superstore Company did well in the field of Technology for the year 2014.

Since we live in Los Angeles, we wanted to know how many number of supermarkets are there in Los Angeles and what are the top cities with most number of Supermarkets. The tenth page of my story shows the ***top cities based on the number of Supermarkets.***  We concluded that the City of New York has the most number of Supermarkets, followed by Los Angeles.

This conversation with my friend helped me understand the Supermarkets from various perspectives, how data analytics has been used to shape the decision making process. This way the Supermarkets in the United States have proved prosperous with their relentless and persistent effort in catering to the needs of the people.

**References:**

[1] An analysis of the Supermarkets in the US.

<https://articles.marketrealist.com/2015/10/an-analysis-of-the-us-grocery-market/>

[2] The Strategic Management of the Supermarkets.

<https://www.governing.com/topics/mgmt/Grocery-Gap.html>

[3] The Profit-Margin of the Supermarkets.

<https://smallbusiness.chron.com/profit-margin-supermarket-22467.html>

[4] The role of Sales and Trend Analysis in the Supermarkets.

<https://www.nchannel.com/blog/how-to-perform-sales-trend-analysis/>

[5] The Research of the Market Demand

<https://www.bigcommerce.com/ecommerce-answers/how-research-market-demand/>

[6] Customer Analysis.

<https://cmgpartners.com/blog/what-is-customer-analysis/>