**M6 – FINAL PROJECT**

**Name: Sunil Raj Thota**

**NUID: 001099670**

**Date: 10/21/2020**

**Title: M6 Project 6 Report**

**ALY 6000 - 71618 (CRN Number)**

**Prof. Dee Chiluiza Reyes**

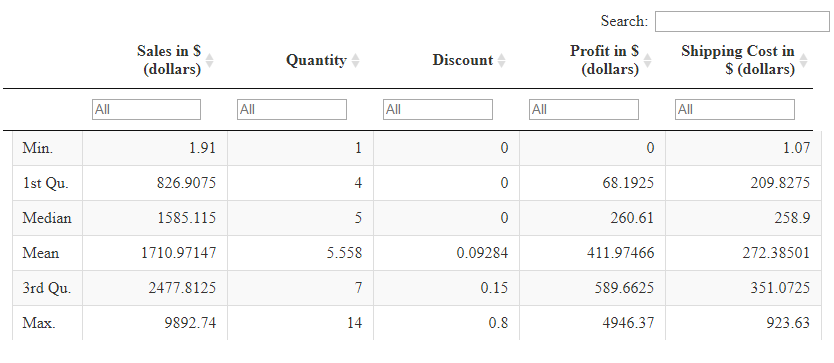
**Introduction:**

The Global Sales in 2016 of a Super Store Company (a.k.a M6Set1) is a data set which has around 1000 records in the form CSV (comma-separated values). This data revolves around the customers and sales so this is crucial for a company to get customer retention. This has the data of all the orders that have been placed through different modes , starting from the year 2012 till 2015. Data Analytics can predict evolving trends, target the precise customer at the right time, the decline in marketing expenses, and raise the standard of customer service. Usually, retailers are customer-focused and customer-driven. Data Analytics domain shows significant impact on customer interactions and enhancements particularly in the retail sector. This data will improve in subsequent applications of retail industry and provides various insights which defines the goals and sales of a company. They know what their customers require. They're stimulated to get profound product knowledge so they'll aid customer requirements. They're qualified enough to be focused and well-mannered. Proper feedback from customers helps outline upcoming marketing plans and store actions. The massive Data and Industry revolution have shown that leading companies are using data to run their total businesses and operations. Tt's become so widespread and available that more retail brands than ever are trusting on data-driven insights to enhance pricing, update operations, and ameliorate the full customer experience.

This dataset consists around 24 columns with different values and filled with data (few records have NA, empty, and NULL). It has RowID, OrderID, OrderDate, ShipDate, CustomerID, CustomerName, PostalCode, City, State, Country, Region, Market, Segment, ProductID, Category, SubCategory, ProductName, OrderPriority, ShipMode, Sales, Quantity, Discount, Profit, and ShippingCost. I have thoroughly analysed and took 5 numerical and 5 categorical varaibles based on the data. There are:

* **Numerical Data:** Sales, Quantity, Discount, Profit, and ShippingCost
* **Categorical Data:** Region, Segment, Category, SubCategory, and ShipMode

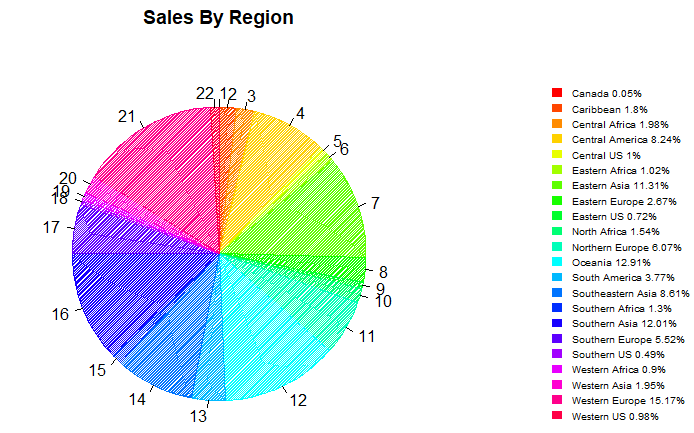
When analysing the data, I got to see that these attributes had played an important role in defining the sales and revenue of that company.



The above is the interactive table summary of the 5 Numerical Variables selected for further analysis to deep dive more. It consists of Min., 1st Quartile, Median, Mean, 3rd Quartile, and Max. values for each of the Sales, Quantity, Discount, Profit, and Shipping Costs Data. The median amount of Profits and Shipping Cost is almost similar where the organisation has to increase its Profits by developing new strategies.

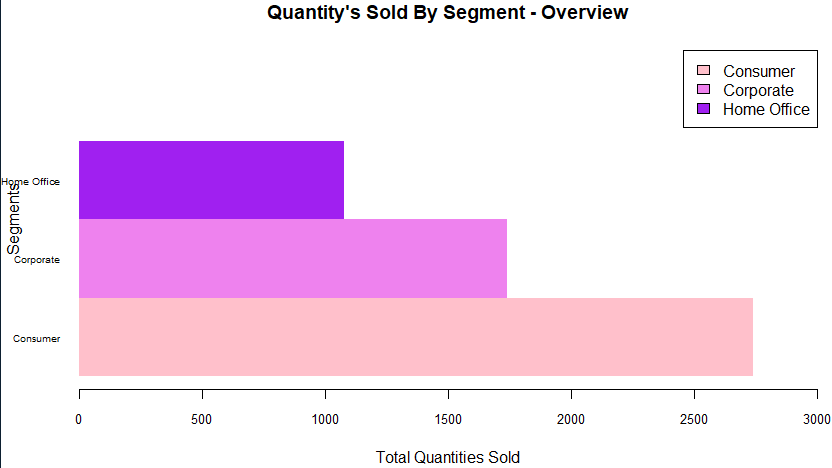
**I would like to focus on the customers and their needs, so that the company’s market size will increase as with the sales and in return it can make a huge profit out of it and can also minimize shipping costs to be competitive and compliance to what matters actually…**

**To Calculate Overall Sales by Region**

****

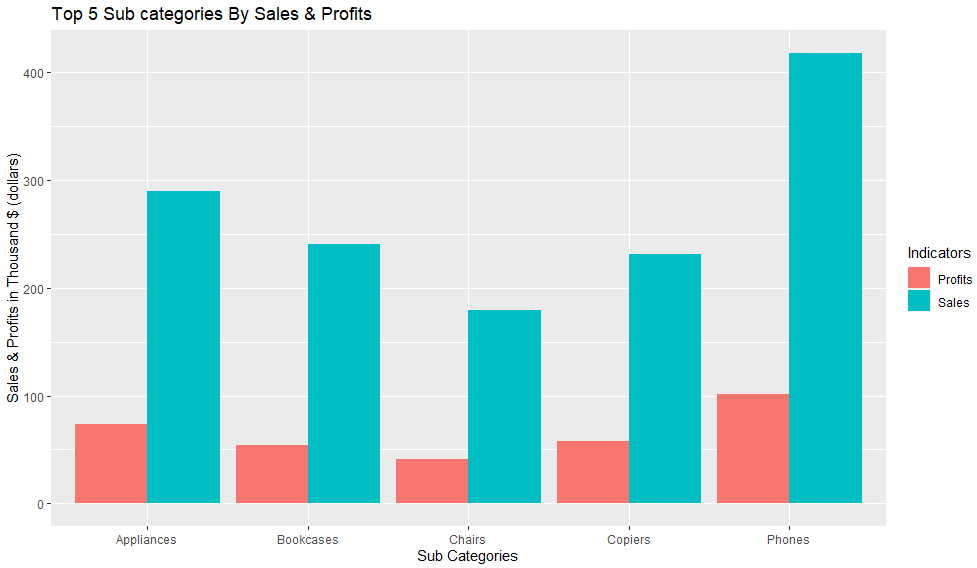
I have calculated the percentages of Sales by Region to analyse where it doing well and got to observe that it’s doing great in Eastern Asia, Central America, Oceania, Southern Asia, South-eastern Asia making around 50% of the sales. So, in this market we can get the customers stronger and better by applying personalization recommendations, offers, and special discounts.

**To Calculate Quantity's Sold by Segment**

****

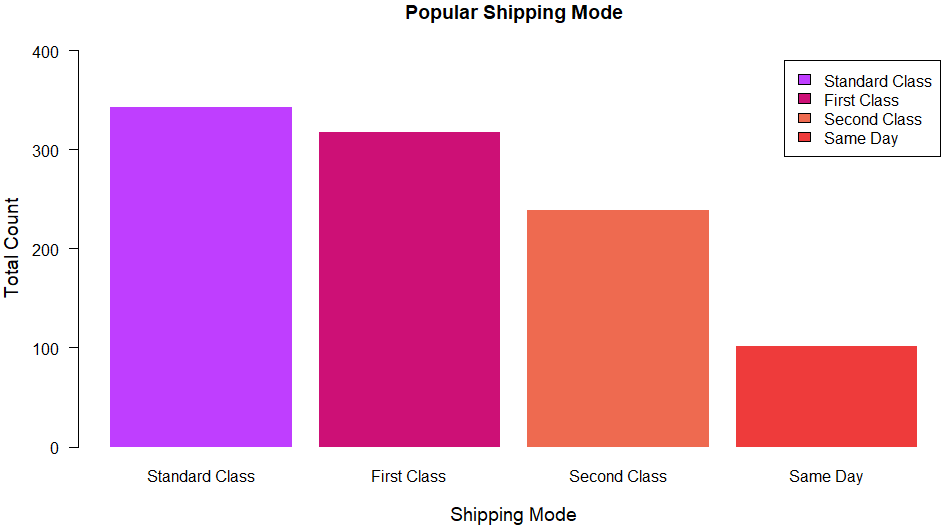
The above chart depicts that the number of quantities sold by various Segments like Home Office, Corporate, and Consumer. We can see that the Consumer Segment alone sold 70% of the Products. It is known that this market is accelerating and booming like never before after the advent of AI, Data Analytics, and ML

**To Calculate Top 5 Sub Category's by Sales & Profits**

****

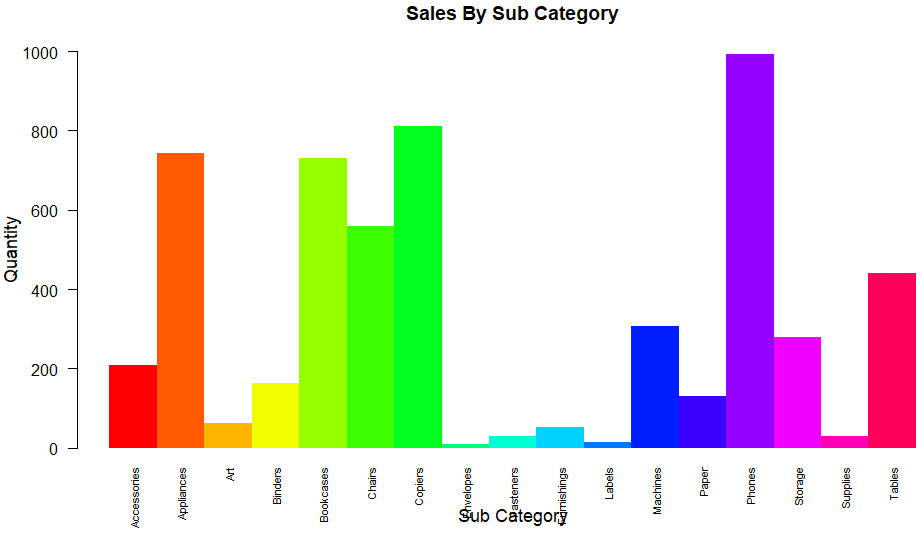
The above chart shows top 5 Sub category’s by Sales and Profits in thousands (dollars). In this I have shown the comparison between the Sales and Profits of that company and indicated them in a plot. By this we can draw the inference like the Phones, Appliances, and Bookcases are the top 3 Sub categories in Sales and their Profits. They account around 750+ products in sales and profits from them which is huge in a particular segment and market.

**To Calculate Popular Shipping Mode by Count**

****

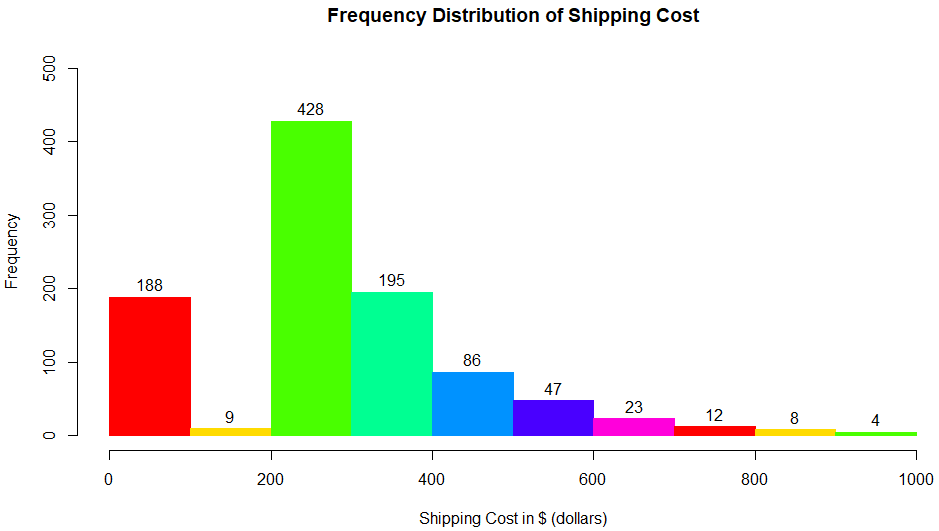
We mostly use data models that are built by diverse algorithms and models to pursuit for outlines and relations in the data by means of numerous data mining methods. Now-a-days, customers need speedy delivery at minimum costs. And, many of them can’t offer for premiums unless and until it brings some value to their purchasing. This creates a probability for new customer relations.

**To Calculate Overall Sales by Sub Category**

****

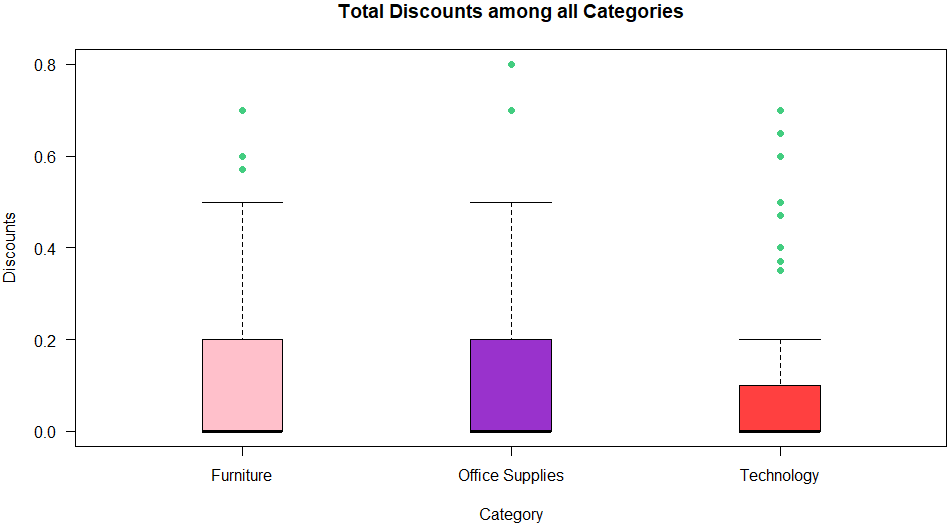
In this, I have calculated the overall quantities sold by the Sub Category and it clearly shows that Phones is the highest and Envelopes are the least sold. So, with this data one can utilize What to sell? How many quantities sold and when needs to be refilled in the inventory to pile up the stocks so that customers can come back? How to market properly and find the target audience? What Customers do require is necessary to know!!

**To Calculate Shipping Cost in Dollars**

****

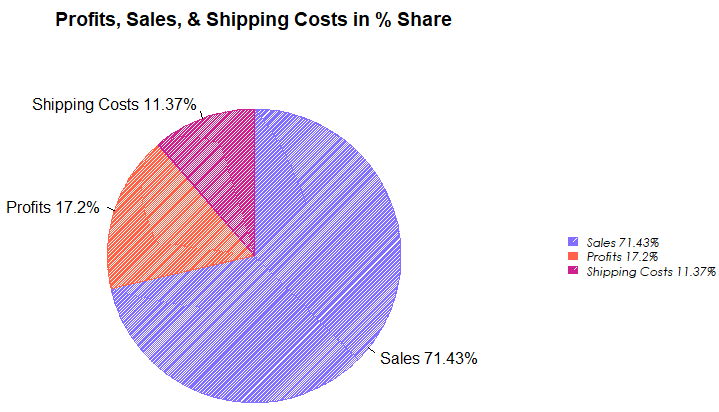
With the above chart, one can observe that the 42% of the orders have shipping cost from 200$ to 300$ and 38% of orders have 0 to 100$ and 300 to 400$. This can be minimized by making few strategic decisions and customer engagements to offer at lower price to retain them.

**To Calculate Total Discounts among all Categories**

****

This chart shows that the Discounts offered among all the categories and in which Office Supplies and Furniture offered more than the twice of what Technology has offered. The most influenced product categories are Technology. This showcases the people’s views on their purchasing interests.

**To Calculate Profits, Sales, & Shipping Costs in % Share**

****

Here, I have calculated the Overall Market Share in terms of Revenue and Sales. I observed and analysed that the company is getting most i.e., 70% its revenue from its direct sales and the rest is taken care by profits and shipping costs. Companies can involve in investing their time in various teams and stakeholders to work on models, recommendation systems, future aspects, developing algorithms, implementing strategies to make choices that in turn improves the businesses.

We can make use of these techniques to develop direct/ indirect marketing campaigns, advertisements, personalized customer discounts, ameliorating customer experience, customer retention, predicting future demand and supply, and level up the stocks, inventory, and warehouses with necessary resources.

**Conclusions:**

With the help of huge data collection from the customers the companies are planning to implement few advanced ways of buying their products. These large sets of shopper’s data can be applied to allow vendors to predict upcoming behaviour, and based on such forecasts, tailor the finest customer suggestions or behaviours to cooperate with customers or dealers. Customer favourites and customization transfers particularly great weightage in this industry. The big data revolution is going to set new standards and will become the future of this industry. An efficient supply chain, an enhanced drive of goods from dealers to silos to stores to the shopper, is very serious in every business. So, big data analytics is so essential of transforming the industry and excels in real-time, leveraging customer data.

I am happy to learn many new technologies from this coursework and will definitely work further on them to enhance my skills while working on real-time projects. This has helped me to gain some hands-on experience to analyse data a way better. This learning will not stop here and definitely Ill make sure to instil the tools, methodologies, techniques in every aspect.

**References:**

1. Big Data Analytics in Support of the Decision Making Process, by Nada Elgendy & Ahmed Elragal [www.sciencedirect.com/science/article/pii/S1877050916324206](http://www.sciencedirect.com/science/article/pii/S1877050916324206)
2. Big Data Analytic: New mode of retail intelligence for revenue growth by Amarjeet Singh, <https://www.irjet.net/archives/V3/i8/IRJET-V3I8357.pdf>
3. Axes and Text, <https://www.statmethods.net/advgraphs/axes.html>
4. Bar Chart Legend Position, <https://stackoverflow.com/questions/27688754/bar-chart-legend-position-avoiding-operlap-in-r>
5. Data camp’s R Tutorial, https://www.youtube.com/watch?v=9f2g7RN5N0I&list=PLjgj6kdf\_snYBkIsWQYcYtUZiDpam7ygg&index=12&ab\_channel=DataCamp
6. Books related to R, <https://www.r-project.org/doc/bib/R-books.html>
7. R Graph Gallery, <https://www.r-graph-gallery.com/>
8. Bar plot, <https://www.statmethods.net/graphs/bar.html>

**Appendix:**

An additional file containing the R codes has been attached to this report. The name of this file is **M6\_Projet 6\_Thota, Sunil Raj**