

**A REPORT**

**ON**

**WALMART SUPERSTORE SALES ANALYSIS**

**BY**

**STUDENT NAME: Akhila.R**

**ZEN COURSE: BUSINESS ANALYICS AND DIGITAL MARKETING**

**BATCH CODE: MBE7**

**AT**

**GUVI**

**CHENNAI**

**(APRIL 2024)**

**E-LEARNING PLATFORM: GUVI**

**Location:** Chennai

**Duration:** 23 days

**Date of Start:** 01-04-2024

**Date of Submission:** 23-04-2024

**Title of the Project:** WALMART SUPERSTORE SALES ANALYSIS

**Name of the student:** Akhila.R

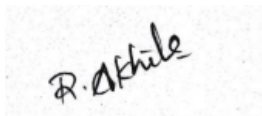
**Zen course name:** Business Analytics and Digital Marketing

**Batch code:** MBE7

**Key Words:** Power BI dashboards, Quick Measures, Sales and revenue analysis, Pareto chart, Profit Forecast, Inventory Management.

**Project Areas:** Excel data cleaning, Dashboard creation for analysis, Root cause analysis

**Project Description:** This project uses Walmart superstore dataset in Power BI to analyze various factors like sales, revenue, customer demography, inventory management, store management to gain hands-on experience on Power BI analysis techniques.



Signature of Student

Date: 23-04-2024

Thank You Team

ZEN

GUVI

## Table of Contents:

1. INTRODUCTION: .....	4
2. CLEANING AND IMPORTING DATA: .....	4
3. DATA INSIGHTS: .....	5
4. PROFIT FORECAST:.....	6
5. SALES : CUSTOMER RATIO ANALYSIS:.....	8
6. SALES AND REVENUE DASHBOARD: .....	10
7. CUSTOMER ANALYSIS DASHBOARD: .....	11
8. INVENTORY AND STORES ANALYSIS: .....	12
9. ROOT CAUSE ANALYSIS: .....	13
9.1. Problem statement:.....	13
9.3. Dashboard Creation: .....	14
9.4. Factors causing loss: .....	15
9.5. Solution: .....	16
10. CONCLUSION:.....	17

## Table of Figures:

Figure 1: Profit Trendline/ Forecast by year .....	6
Figure 2: Profit Forecast by year and quarter-Line Chart .....	7
Figure 3: Profit forecast by year, quarter and month- Line chart.....	7
Figure 4: Profit analysis dashboard.....	8
Figure 5: Sales : Customers Pareto Chart .....	9
Figure 6: Sales : Customers Dashboard .....	9
Figure 7: Sales and Profit Analysis Dashboard.....	10
Figure 8: Customer Analysis Dashboard .....	11
Figure 9: Inventory and Store Analysis Dashboard .....	12
Figure 10: Profit Trendline by year, Quarter and Month .....	13
Figure 11:Profit Correlation dashboard .....	14
Figure 12: Profit Root Cause Analysis Dashboard .....	14
Figure 13: Profit Root Cause Analysis Dashboard for April 2014.....	16

## List of Tables:

Table 1: Column Data Types .....	5
Table 2: Top 5 Loss generating Product Details. ....	15

## **1. INTRODUCTION:**

This project is about analyzing the Walmart superstore dataset and gathering powerful insights on sales, profit, customer demographics, etc., using power BI visualizations.

Walmart is one of the most successful American multinational retail corporation. It operates both offline and online with wide variety of products. They are known for their everyday low-price strategy which makes them differentiated and successful.

Some of their main competitors are amazon and target.

## **2. CLEANING AND IMPORTING DATA:**

- Once the data set is downloaded convert it into excel file from csv.
- When it was a csv file some of the date values were not in the correct format and excel couldn't identify it as date data type. After saving it as a excel workbook file, the date data types are converted automatically to correct format.
- Check if all the columns are displaying the correct data type.
- Discount column was not categorised into percentage. Convert the discount column to percentage just by clicking the column and changing the format of the column display. This step can be done both in excel or in power BI table view space/ power query editor.
- In customer names some of unknown characters like ¶, Ã, ¢ were present. Use find and replace option in excel or replace values under transform in power query editor to remove or replace these symbols.
  - ¶ is erased or replaced by no value.
  - Ã is replaced with a
  - ¢ is erased or replaced by no value.
- Once cleaning is done, open Power BI and use the get data option in home tab and select file format as excel and click the Walmart superstore excel file.

### 3. DATA INSIGHTS:

The Walmart superstore data has 21 columns and 3203 data entries. The following is the table indicating the data types of the columns and the number of unique values.

S.NO	DATA TYPE	COLUMN NAMES	NO. OF UNIQUE VALUES
1	INTEGER	Row ID	3203
		Postal Code	180
		Quantity	14
2	FLOAT/ DECIMAL NUMBER	Sales	2348
		Discount	5
		Profit	2614
3	DATE	Order Date	845
		Ship Date	911
4	TEXT	Order ID	1611
		Ship Mode	4
		Customer ID	686
		Customer Name	686
		Segment	3
		Country	1
		City	169
		State	11
		Region	1
		Product ID	1509
		Category	3
		Sub-Category	17
		Product Name	1494

Table 1: Column Data Types

- Row ID: contains an ID or serial number for each entry.
- Order ID: contains two alphabets (CA/US) followed by the year of order and 6 numbers. These three elements are separated by a hyphen (-). Example: CA-2011-143336.

**NOTE:** Multiple entries contain same Order ID.

Example: In case a customer has placed an order containing three different products, there will be three entries with same Order ID.

**IMPORTANT INSIGHTH:** Order ID doesn't change with products, only the Row ID changes.

- Order Date and Ship Date consists the dates of order and shipping of the order placed.
- Ship Mode: indicates the shipping mode, and has four unique values
  - First Class: order takes a shipping time of 1 to 3 days.
  - Same Day: maximum orders get shipped on the same day or on the next day.
  - Second Class: orders get shipped within 2 to 5 days.
  - Standard Class: orders get shipped within 3 to 7 days.

**IMPORTANT INSIGHTHS:** Orders take a minimum of 0 to maximum one week to get shipped.

- Customer ID and Customer Name: Each customer has a unique ID, containing Two Alphabets with the first character of their first and second name followed by a 6-digit numeric value separated by a "-". Example: RA-19885. There are 686 unique customers.
- Segment: This indicates the three customer segments: Corporate, Consumer, Home Office.

- Country and Region: All entries belong to United States, West Region.
- City: indicates the city where the product/ order is to be shipped. Has 169 unique values.
- State: indicates which state the order is being shipped. Has 11 unique values.
- Postal Code: Has 180 unique values.
- Category and Sub- Category: Indicates which category and sub- category the product belongs to. Category is divided into: Furniture, Office Supplies and Technology. There are total of 17 sub-categories.
- Product ID and Product Name: Each product has first three letters of the category followed by first two letters of the sub category and an 8-digit numeric value separated by “-”. Example: OFF-BI-10002215.
- Sales, Profit and Quantity: These columns contain the sales, profit (contains both positive-profit and negative-loss values) and the quantity of products ordered.
- Discount: Indicates the discount value of the ordered product. Discounts are either 0%, 15%,20%,50%,70%.

#### 4. PROFIT FORECAST:

Profit forecast can be done with line chart in Power BI using the power analysis option under visualization. On the x-axis select Order Date and y-axis select Profit.

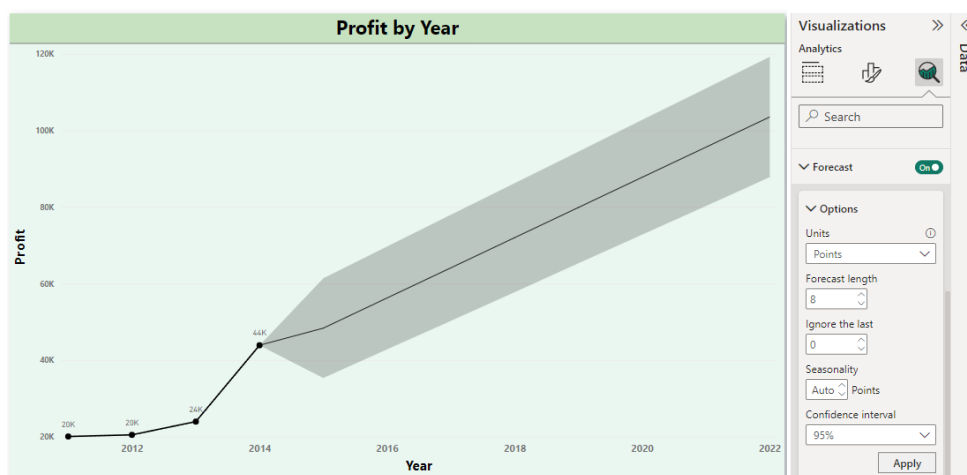


Figure 1: Profit Trendline/ Forecast by year

- We have predicted profit for the next eight years with 95% confidence interval.
- The profit of next few years are predicted to be **2015:** \$48368.64, **2016:** \$56249.39, **2017:** \$64130.15, **2018:** \$72010.91 and so on.
- Each point has a upper and lower bound value according to the confidence interval and the forecast line passes through the center of this region.
- We can represent the line graph in different combination of hierarchies of order date with drill down and up settings.
  - Like profit by year and quarter to show data point of four quarters in each year. And the forecast will be calculated for the next 8 quarters (2 years).
  - Profit by year, quarter and month: there will be 12 data points for each year having the month and quarter. Profit will be predicted for the next 8 months.
  - Similarly we can see the profit by year or month or quarter or days individually.

- We can use Decomposition tree with profit and order date to visualize heirarchy of profit by year, month and quarter.
- With the data of four years we can see there is a positive trend (profit increases with every year).
- There is no pattern found in profit by year and quarter. However in 2011 and 2014 Quarter 3 had the highest profit and in 2012 and 2013 Quarter 4 had the highest profit.
- **March 2014** marked the highest profit of all months, having profit value of 9.1k.

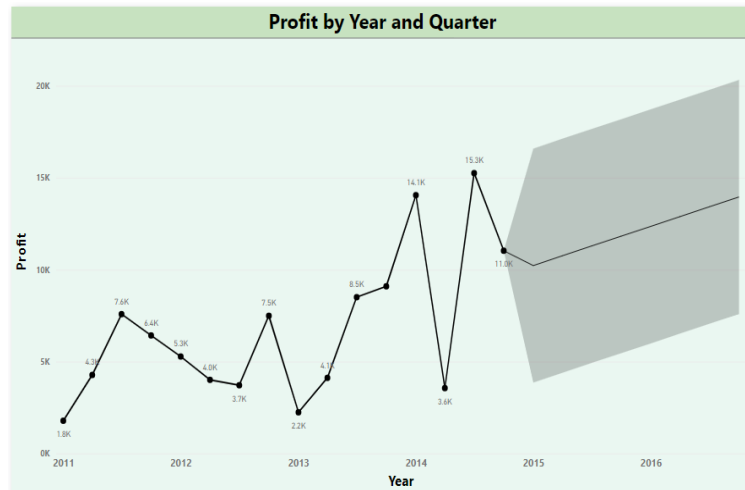


Figure 2: Profit Forecast by year and quarter-Line Chart

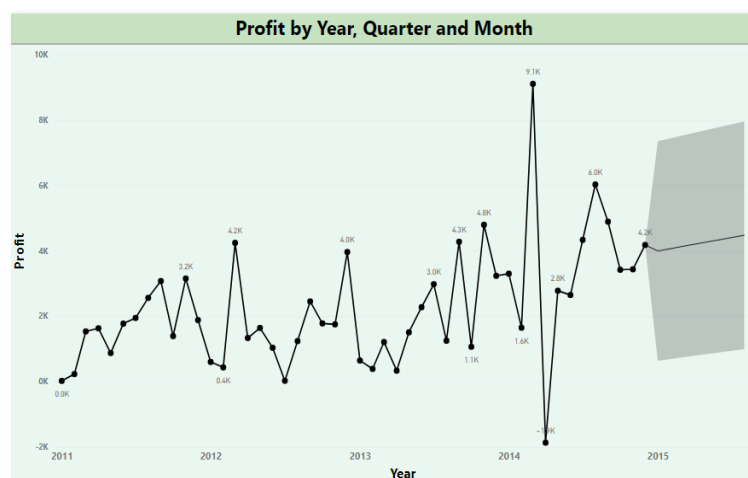


Figure 3: Profit forecast by year, quarter and month- Line chart

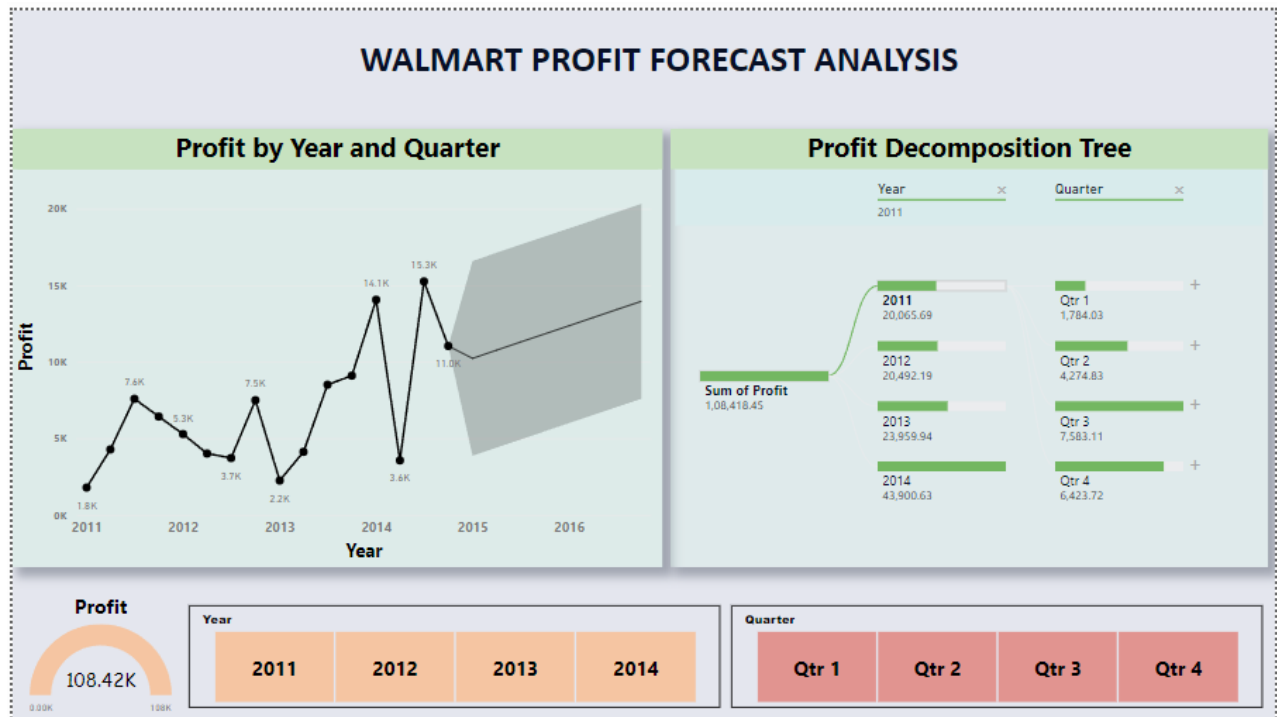


Figure 4: Profit analysis dashboard

## 5. SALES : CUSTOMER RATIO ANALYSIS:

To prove that 80% of sales is coming from 20% of customers we use a popular chart “Pareto Chart”. In order to visualise this chart, we need three values

1. **No. of Customers** in descending order of 1 to 686 with customer 1 having the highest sales value.
2. **Sales** from each customer in descending order.
3. **Cumulative Sales percentage**.

**Steps followed to calculate values in excel:**

- In power BI select customer name (distinct) and sum of sales in descending order, now save the table as excel file.
- Open the excel and create a column as customer count having customer with highest sales on the top. This will be the x-axis value to know how many customers has contributed to 80% of sales.
- Create a column for % of Total sales from each customer and the cumulative % of sales
  - Formula: % of total sales = Sum of Sales/Grand Total Sales\*100
  - Cumulative value of a customer x will consist the value of customers before them plus the value of the customer x.
- Upload this file to Power BI and use combination of column and line chart to visualize the pareto chart.
- Put customer count in x-axis and sum of sales in y-axis (descending order) to visualize a bar chart.
- Add cumulative % of sales in line y-axis section.
- We can see the line increasing to 100% and the bar chart in descending order.
- We can observe that 80% of sales is contributed by 260 out of 686 customers which is 37.9%.



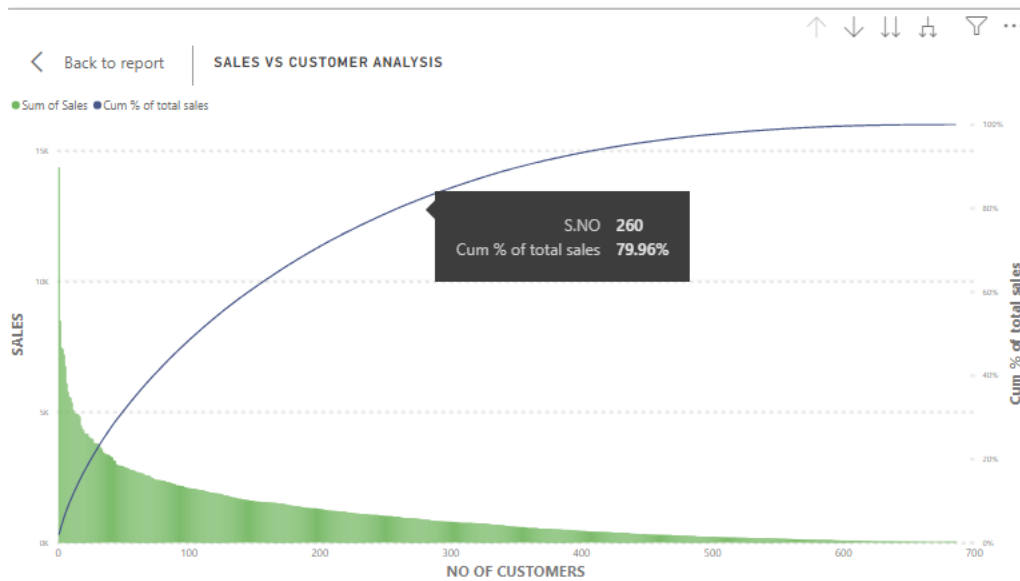


Figure 5: Sales : Customers Pareto Chart

With the help of slicers and cards, we can set the sales percent to 80% and see the percentage of customers or slide the number of customers and know how much they are contributing to sales.

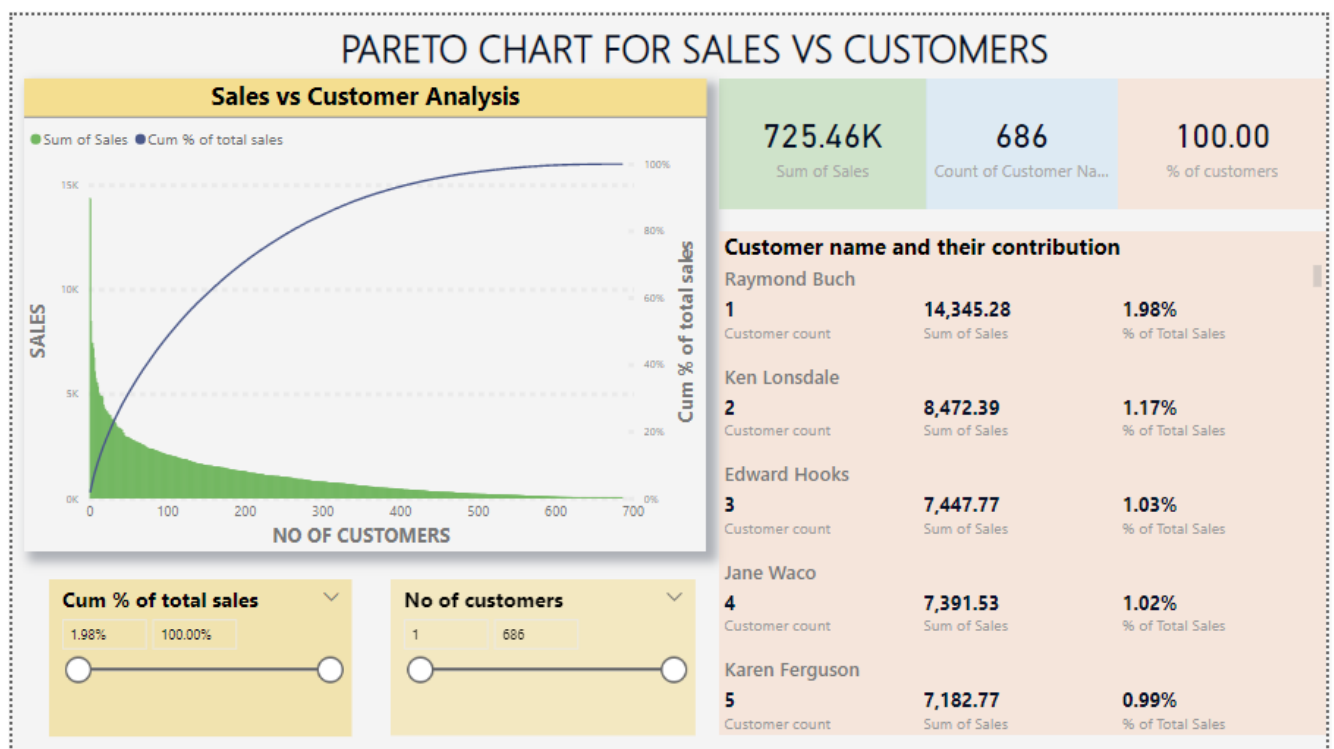


Figure 6: Sales : Customers Dashboard

## 6. SALES AND REVENUE DASHBOARD:

A Sales and Profit dashboard is created to track the store's sales and revenue over time, by category, region, and product.

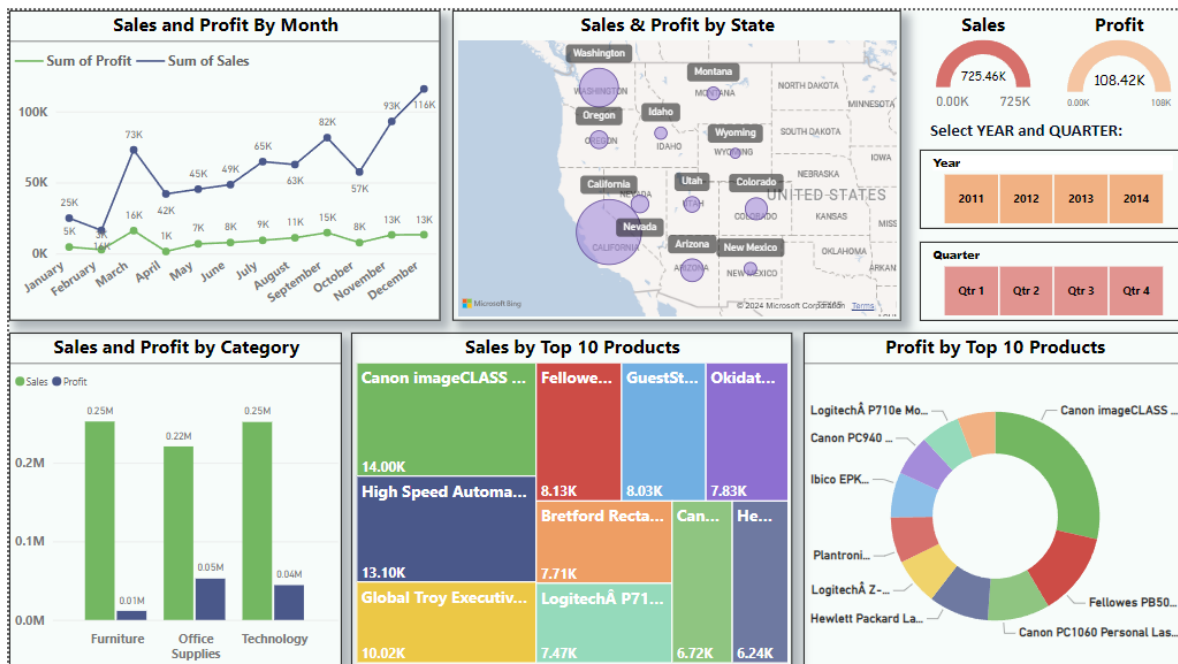


Figure 7: Sales and Profit Analysis Dashboard

- Firstly, the trend/ line chart shows the series of profit and sales by months (cumulative values of four years). It can be changed to year and quarters by drill down and up options. There is a consistent rise in sales. And during March, September, November and December there is increased sales.
- The profit line is somewhat consistent in nature with peak in march.
- Select years in filter:** In all years December sales stands the highest of that year. This may be due to Christmas and New year discounts/offers.
- Next is a simple map with size bubbles of sales in each state. California has the highest sales and profit. The lowest sales is from Wyoming. On further analysis we can see that from Wyoming there is only one customer who has ordered chairs (furniture-home office) only once on 13<sup>th</sup> November 2013.
- In Wyoming products are bought only from home office segment and in Montana there are customers who get products only from home office and consumer segment. In order to sell all segments in all states, we need to increase the marketing for other categories but also focus on providing most purchased products with maximum customer satisfaction.
- The states which generate loss: Colorado, Arizona, Oregon.
- A clustered bar chart with sales and profit by product category shows that sales is even for all the categories (>0.2M), but profit is more from the category Office supplies and Technology.
- Sales and profit by segment chart show that sales and profit are high in consumer segment.
- With the help of heat map, we can visualize the top 10 products contributing to sales.
- With the help of donut map we can visualize the top 10 products contributing to profit.
- The cards in top right corner shows total sales and profit.
- With the help of filters, we can use year and quarters to see it in all visual charts.

## 7. CUSTOMER ANALYSIS DASHBOARD:

A customer analysis dashboard is used to analyse customer data, such as demographics and purchase history, to gain insights into their behaviour and preferences.

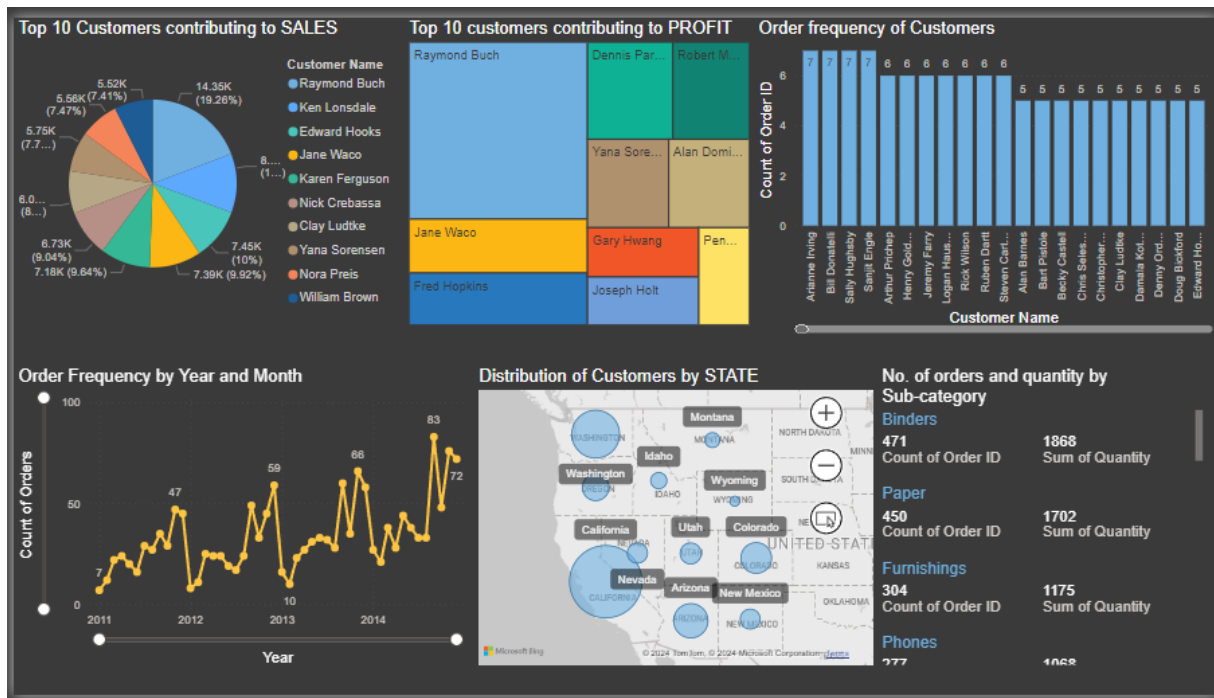


Figure 8: Customer Analysis Dashboard

- The pie and tree map shows the top 10 customers contributing to sales and profit.
- The frequency of customer orders shows how many times the customer has placed orders in the last four years. The highest frequency is 7 by Arianne Irving, Bill Donatelli, Sally Hughsby and Sanjit Engle. We can say that the customers who ordered 4 and more times are our loyal customers.
- Order frequency from line chart can show what is the order traffic each month or year. We can observe that during the last quarter i.e. mostly during November and December there is a high traffic in orders. This is a seasonal trend seen in all years.
- The demography of customers can show where our most customers place order. California, Washington and Arizona are the top three states where our customers order products.
- A card shows the no. of orders and quantity by product sub-category to see the popular product categories.
- All the charts are responsive and interactive.
- We can select a customer from top 10 list or most ordered customers to see the states he/she has placed order, what's their order frequency and history, and the product category he/she ordered.

## 8. INVENTORY AND STORES ANALYSIS:

To monitor the inventory levels for ensuring supply of popular products, and to optimize the store performance by analysing customer traffic a dashboard is created to visually analyse it.

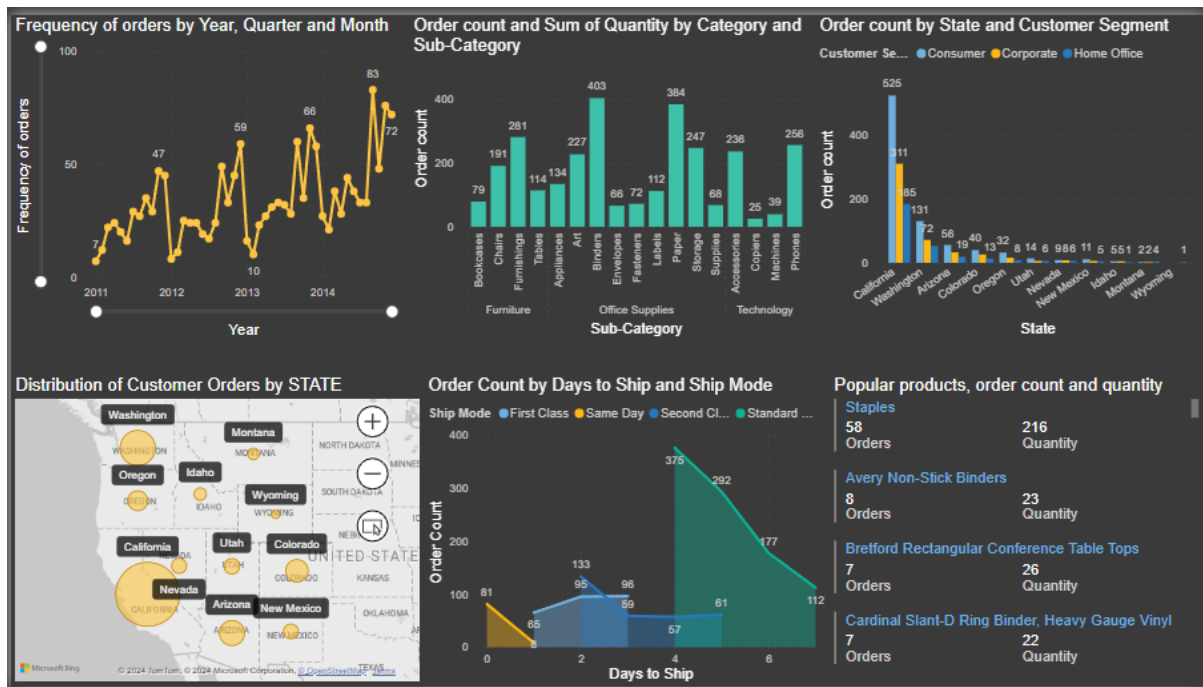


Figure 9: Inventory and Store Analysis Dashboard

- Order frequency from line chart can show what is the order traffic each month or year. We can see that the order traffic is high during fourth quarter so we must be prepared for adequate stock of products. This is a seasonal trend seen in all years.
- Next is a column chart displaying the number of orders placed in each sub-category. We can see that the binders and paper in office supplies are the popular product sub-category.
- A clustered column chart displays the no. of orders placed in each customer segment by state. We can see that consumer segment is dominant in all states. And most of our orders are from California.
- From a simple map we can identify the order traffic from each state to manage/plan the inventory and shipping in that state.
- An area chart shows the order traffic in each shipping mode and the days for those orders to get shipped. We can see that most customers place orders with standard shipping of 4 days.
- A multi-row card shows the popular products with their order count and quantity.
- All the charts are responsive. Once we click drill down and select a data point in line graph, we can see the days where the traffic was high, popular products and their order and quantity, from where the orders are placed, customer segment and product category and the shipping mode.
- With this we can be well planned to provide supply of products in all states.

## 9. ROOT CAUSE ANALYSIS:

### 9.1. Problem statement:

The problem statement for root cause analysis is: Loss of \$1867.75 on April 2014.

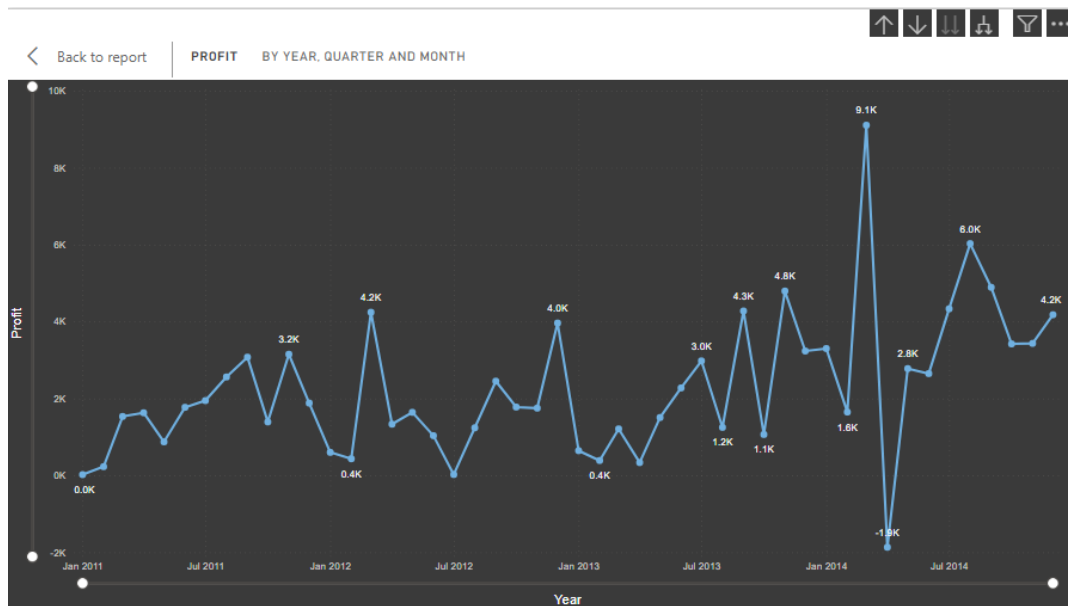


Figure 10: Profit Trendline by year, Quarter and Month

- By seeing the profit line chart, we can observe that there is a huge loss of \$1867.75 on April 2014.
- Normally during September, November and December the sales and profit value is high. But In 2012 and 2014 profit in March is the highest.
- Since the profit values of march 2014 and April 2014 are outliers, this might not happen every year. But the profit of last two years 2013 and 2014 shows April as their lowest value.
- Hence, it is taken as a problem statement to analyse the factors causing the loss.

### 9.2. Correlation Analysis:

- By using scatter plots we can see the correlation of profit with various factors like order quantity, sales and discount. Quick measures can be used to calculate the correlation coefficient of the trend line.
- When categorised by order ID, profit has a positive correlation with sales. Increase in Sales or higher the Sales value higher is the profit. The correlation coefficient ( $R^2$ ) is 0.67 which indicates moderate relation between the two variables.
- When categorised by order ID, Profit and discount have a negative correlation. When discount is high the profit is low or there can be loss. The correlation coefficient ( $R^2$ ) is 0.63.
- When categorised by order ID, profit has no/very mild correlation ( $R^2=0.17$ ) with Quantity. The highest order quantity in an order is 46 which was made by Keith Herrera on 9<sup>th</sup> December 2014.
- However, when we categorise the profit and quantity by order date there is a positive correlation with coefficient  $R^2= 0.56$  which indicates moderate correlation. The maximum Total quantity is 634 on December 2014. This chart has total order quantity of every month in all four years.
- So, we can conclude that if the order quantity is higher the profit will be high.

- The outliers march 2014 and April 2014 can be seen in all the scatter plots.

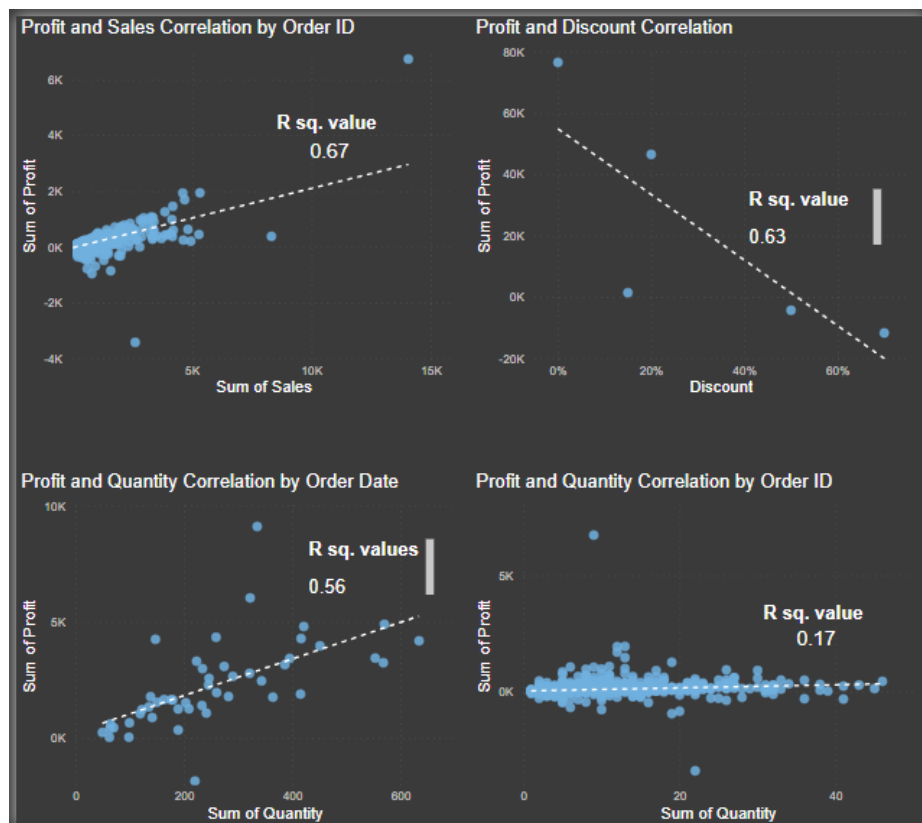


Figure 11: Profit Correlation dashboard

### 9.3. Dashboard Creation:

A dashboard can be created to drill down and analyse the loss with various charts.

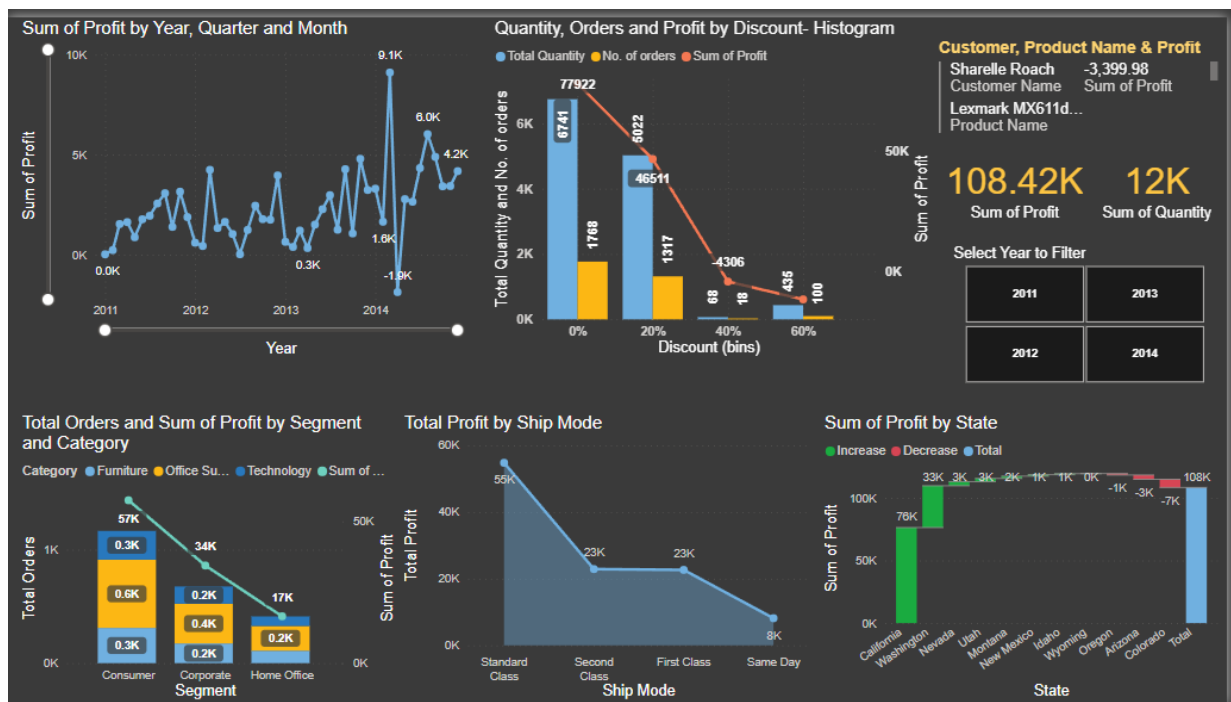


Figure 12: Profit Root Cause Analysis Dashboard

The dashboard consists of:

- Line chart – to represent profit by Order date. Can be drilled down from year to days.
- Histogram and Line chart- discount bins with bin width of 20% is created to see order quantity and number of orders in each discount bin. A line represents the total profit in each discount bin.
  - The minimum discount is 0% and maximum is 70%. We have 0%, 15%, 20%, 50% and 70% of discount in our dataset.
  - A total of four bins will be created with width of 20% (0%-20%, 20%-40%, 40%-60%, 60%-70%). Note that each bin includes the value of the lower limit but not the upper limit. Example the 20% bin consists of quantity, orders, profit for 20% to 39%.
- Stacked column and line Chart: To visualize the order count by customer segment ,product category and the total profit from each consumer segment.
- Area Chart: To visualize the total profit for each shipping mode.
- Waterfall Chart: To visualize the profit from each state adding up to the total profit.

#### 9.4. Factors causing loss:

From the dashboard we can gather these important insights:

- As the order quantity is highest in 0%, 15% and 20% the profit is more from these discount bins.
- With the increase in discount there is a significant drop of profit. Both the 40% (having 50% discount data) and 60% bins (having 70% discount data) show loss of \$4306 and \$11709 respectively. This indicates that most of the products sold with 50% and 70% discount are causing loss.
- The huge loss of \$11k when the discount is 70% is due to products in binders, bookcase and machines sub-category.
- The top 5 products causing loss are:

S.no	Product Name	Category	Sub-Category	Quantity	Profit Value (\$)
1	Lexmark MX611dhe Monochrome Laser Printer	Technology	Machines	5	-3399.98
2	Zebra GK420t Direct Thermal/ Thermal Transfer Printer	Technology	Machines	6	-938.28
3	Atlantic Metals Mobile 4-Shelf Bookcases, Custom Colors	Furniture	Bookcase	7	-786.74
4	GBC DocuBin 300 Electric Binding Machine	Office Supplies	Binders	6	-694.29
5	Swingline SM12-08 MicroCut Jam Free Shredder	Technology	Machines	5	-479.99

Table 2: Top 5 Loss generating Product Details.

- The consumer segment has the highest order quantity hence the profit is high in this segment. Consumer > Corporate > Home office
- Standard class shipping mode provides the highest profit to the company. The order quantity in this shipping mode is the highest. This may be because the standard class shipping is the cheapest option for the customers. Standard Class > Second Class > First Class > Same Day

- In the waterfall chart green indicates profit and red indicates loss. Oregon, Arizona and Colorado are the states which contributes to loss.

We need to use the drill down option in line chart and select the value: April 2014 to further know the factors that are leading to this loss.

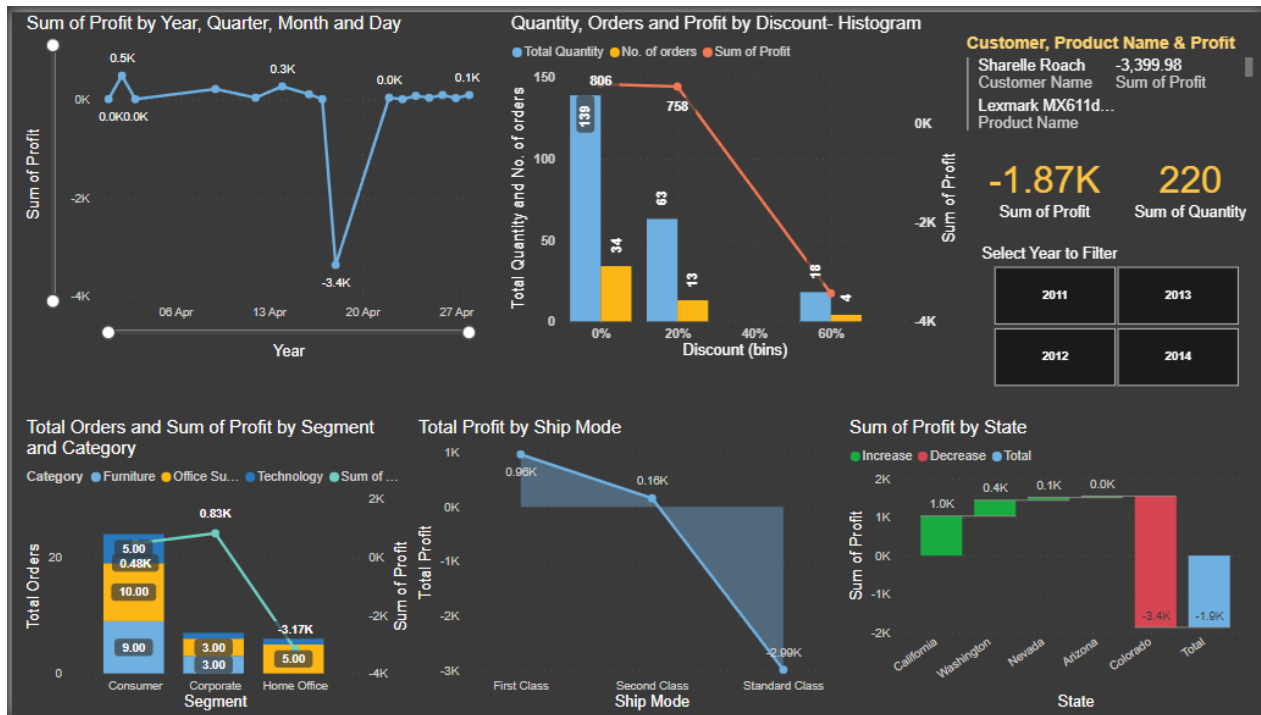


Figure 13: Profit Root Cause Analysis Dashboard for April 2014

- We can see through-out this month there has been very less profit of maximum 0.5k. But on 18<sup>th</sup> April the loss is \$3364.87. This is the main value leading to the total loss of \$1.87k in April 2014
- About **\$3431** of loss is coming from **18 nos.** of products from home office segment with **70%** discount.
- If we select the data point in 60% discount bin we can find the following:
  - That all these orders of different **home office** products were made by **Sharelle Roach** from home office and technology product-category.
  - The top most product which is contributing to a loss of **\$3399.98** is **Lexmark MX611dhe Monochrome Laser Printer**. The order quantity of this product is **5nos.**
  - All these orders were made from the state **Colorado** with **standard** ship mode.

## 9.5. Solution:

- To reduce loss, we must make a decision to not sell most valued (high priced) products of binders, bookcase and machines with 70% or 50% discounts.
- Instead, we can discount these products by 20% or 30% with 5% or 10% extra off on order quantity more than 3.



## **10. CONCLUSION:**

The learning outcomes achieved by completing this project:

1. Data Modelling:
  - Cleaning the data, analysing the data types and removing symbols and null values.
  - Importing new excel file for sales: customer pareto analysis and auto-created relation with both tables by customer name.
  - Creating new column (Days to Ship) to know the time taken to ship orders.
  - Creating quick measures to calculate correlation coefficient.
  - Grouping Discount to create Bins for using it in Histogram.
2. Data Visualization and Report Development:
  - Creating dashboards with slicers and cards to interact with charts like: Line Chart, Column Chart, Bar Chart, Simple Map, Area Chart, Waterfall Chart, Pie Chart, Donut Chart, Tree Map, Histogram, Stacked Column Chart, Decomposition Tree.
3. Business Solution:
  - Doing Root cause analysis to find solution for business problems.
  - Analysing sales, profit and customer to find areas of improvement.