INTRODUCTION TO DATA MANAGEMENT: PROJECT REPORT

on

US SUPERSTORE SALES EXCEL DASHBOARD

Submitted by

Dipendra Singh

Reg. No. **12015804**

Programme - P132: B.Tech. (Computer Science & Engineering)

Section: KM079

Course Code:INT217

Under the Guidance of

Ms. Sameeksha Khare: 27946

Discipline of CSE/IT

Lovely school of Computer Science Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Dipendra Singh** bearing Registration no. 12015804 has completed INT:217 project titled, **"US SUPERSTORE SALES DASHBOARD"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

School of Computer Science Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 10-11-2022

DECLARATION

I hereby I, **Dipendra Singh**, student of B.Tech. under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

.

Dipendra Singh

12015804

Date: 10-11-2022

ACKNOWLEDGEMENT

I would like to express my gratitude towards my University Lovely professional University for providing me the golden opportunity to learn this wonderful subject name is INT217 (introduction to data management), which is very vast and deep learning of manage data and how to analyse it. As a result, I came to know so many interesting things about Excel. So, I am really thankful to them.

Moreover, I would like to give a special thanks my friends and family who helped me a lot whenever I got stuck in any problem related to my subject or when I feel depressed. I am really thankful to have a good support of them as they always be with me whenever I need.

I have taken efforts in this subject. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

Deepest thanks to our Instructor Ms. Sameeksha Khare mam for his guidance, monitoring, and constant encouragement, attention and care. They have given more than they give to our project and training sessions and make necessary corrections as when needed and we are very grateful for that.

TABLE OF CONTENT

INTRODUCTION

The Excel Dashboard provides an overview of metrics and other data points in one place.

In simple terms, dashboards are visual representations of data.

Excel dashboards make it easy to perform quick overviews of data reports rather than going through large volumes of data. Overviews help in making quick and urgent decisions since one can skim through a lot of information at once and within a short time.

The dashboards help in tracking Key Performance Indicators (KPIs) with ease, which helps organizations track the progress on their targets. They provide a high-level summary of key aspects of your data to keep everyone at par with the progress, hence giving the organization a timely indicator for necessary action in real-time.

Excel dashboards include various elements such as charts, tables, figures, and gauges that help in presenting the data. They can handle any type of data from different market and purposes, and the information can be used for marketing, financial, or other business projects. The dashboard is most applicable to large volumes of data since it would otherwise be hectic to go through such large volumes of data, especially with limited time.

Objective of Analysis

These are objective of my Analysis of US Superstore Sales dataset: -

1. Perform Analysis on Dataset based on sales
   - Sales and Profit based on region
   - Sales Trend on Time
   - Sales based on State
2. Perform Analysis on Dataset based on Category and subcategory
   - Sales and profits based on category and sub-category
   - Discount based on segment and category
3. Perform Analysis on Dataset based on customer segmentation
   - Top 5 customers
   - Most popular segment
   - Most preferred ship mode by region
   - Top 5 products

Source of Dataset

I have used the Dataset name: -

Sample - Superstore Sales (Excel).xls

Sample data that appears in the December Tableau User Group presentation. Note: Geographic locations have been altered to include Canadian locations (provinces / regions).

It was updated Superstore Excel file to the version shipping with 10.4 - 11/28/2017

It contains the data of 4 years (2014 to 2017)

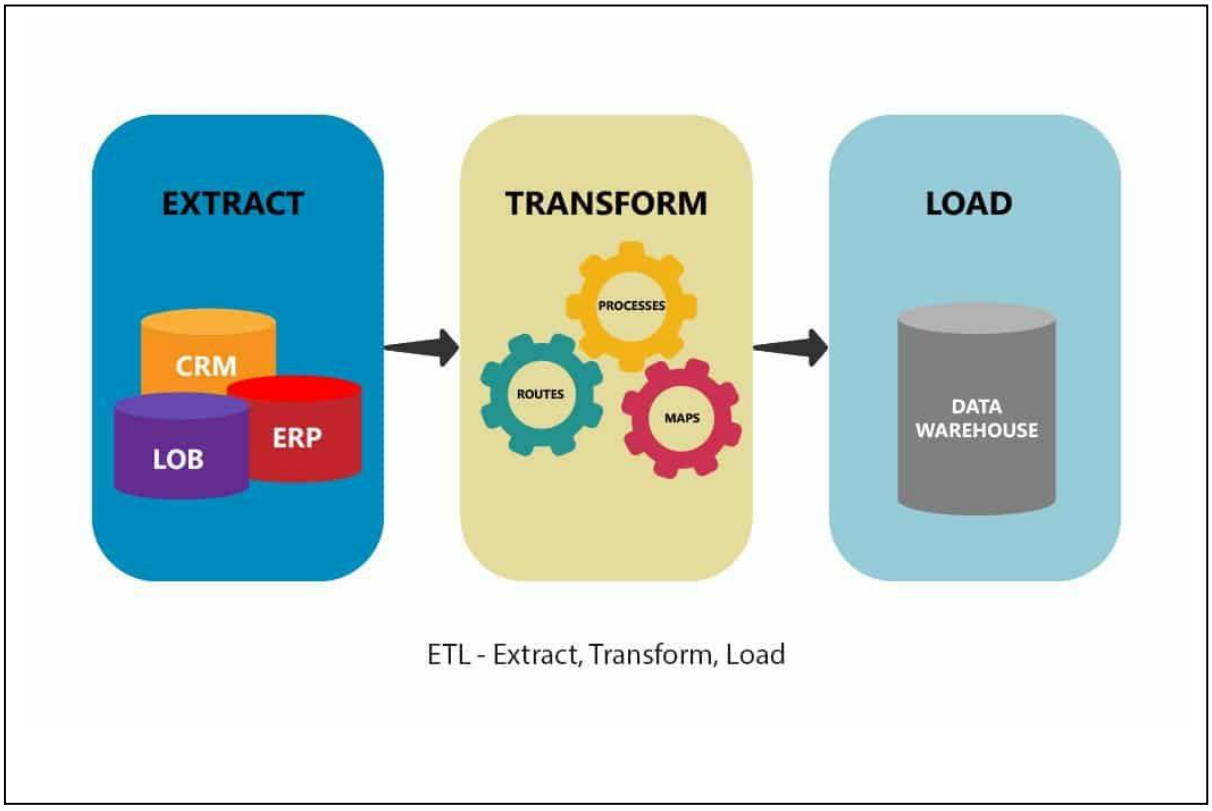
Link of Accessing the dataset is: -

https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls

ETL Process

ETL stands for Extract, Transform and Load.

- ETL is used to integrate the data with the help of three steps Extract, Transform, and Load, and it is used to blend the data from multiple sources. It is often used to build a data warehouse.
- In the ETL process, data is extracted from the source system and convert into a format that can be examined and stored into a data warehouse or any other system.



ETL - Extract, Transform, Load

Extract

- **Extract** is the process of fetching (reading) the information from the database. At this stage, data is collected from multiple or different types of sources.
- A staging area is required during ETL load. There are various reasons why staging area is required.
- The source systems are only available for specific period to extract data. This period is less than the total data-load time. Therefore, staging area allows you to extract the data from the source system and keeps it in the staging area before the time slot ends.
- Staging area is required when you want to get the data from multiple data sources together or if you want to join two or more systems together. For example, you will not be able to perform a SQL query joining two tables from two physically different databases.
- Data extracted from source systems can be used in multiple data warehouse system, Operation Data stores, etc.

Transform

- In data transformation, you apply a set of functions on extracted data to load it into the target system. Data, which does not require any transformation is known as direct move or pass-through data.
- *Transform* is the process of converting the extracted data from its previous form into the required form.
- You can apply different transformations on extracted data from the source system.
- For example, you can perform customized calculations. If you want sum-of-sales revenue and this is not in database, you can apply the SUM formula during transformation and load the data.
- For example, if you have the first name and the last name in a table in different columns, you can use concatenate before loading.

Load

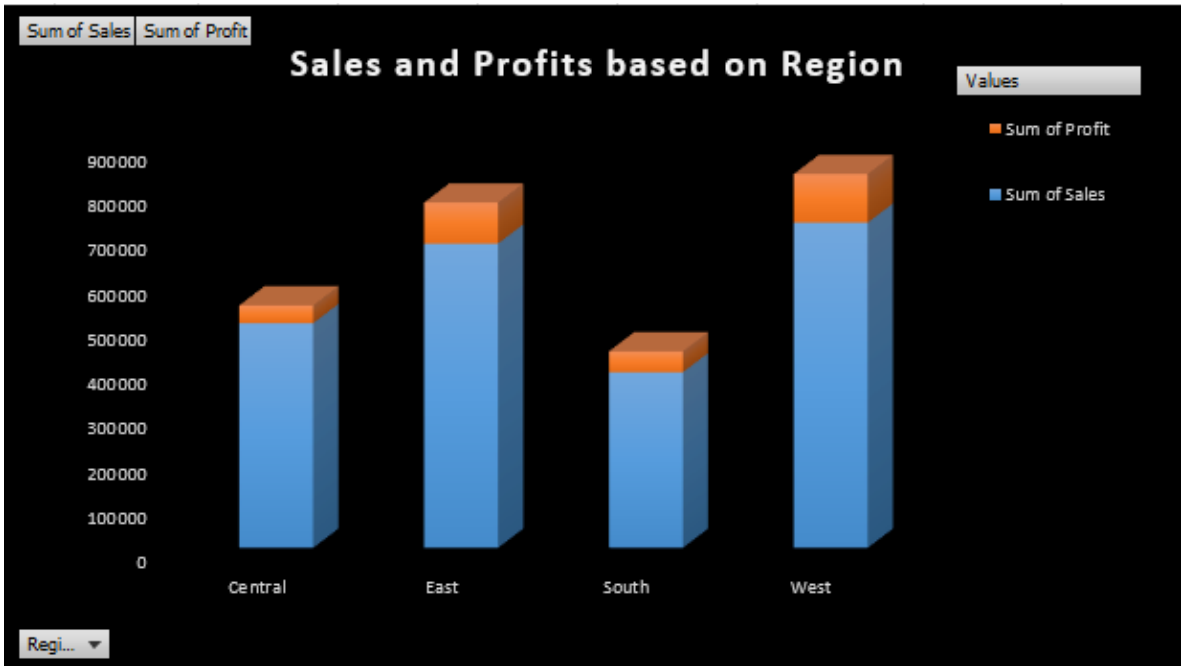
- In this last step, the transformed data is moved from the staging area into a target data warehouse. During Load phase, data is loaded into the end-target system, and it can be a flat file or a Data Warehouse system.
- Typically, this involves an initial loading of all data, followed by periodic loading of incremental data changes and, less often, full refreshes to erase and replace data in the warehouse. For most organizations that use ETL, the process is automated, well-defined, continuous and batch driven. Typically, ETL takes place during off-hours when traffic on the source systems and the data warehouse is at its lowest.

Types of Loading:

- **Initial Load** — populating all the Data Warehouse tables
- **Incremental Load** — applying ongoing changes as when needed periodically.
- **Full Refresh** —erasing the contents of one or more tables and reloading with fresh data.

Perform Analysis on Dataset based on sales

1)Sales and Profit based on region



In this chart, I have shown the Sales and the Profits based on region wise. ‘

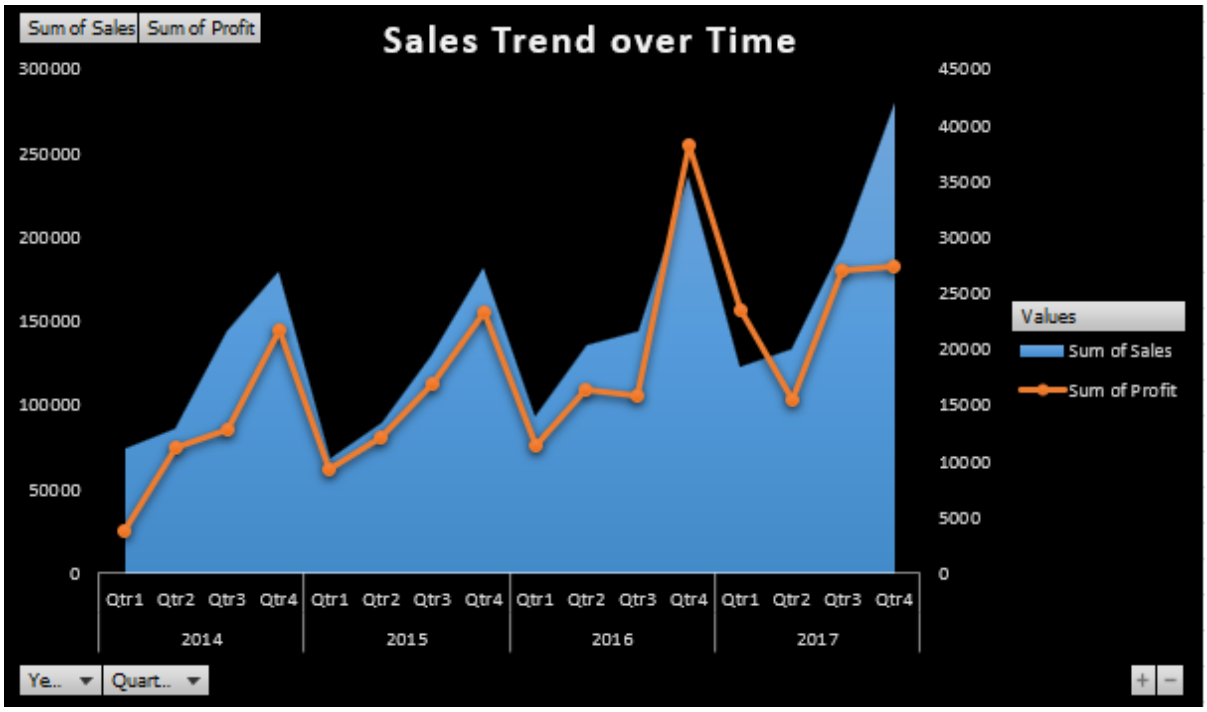
This char shows us how much benefits we are getting from different region based on sales and profits and it tell us the sum of sales and profits of that particular region with **3D stacked column chart**.

Basically, the formula used for preparing this chart is: -

- Region -> Category Axis
- Sales and profits -> Values Section

Perform Analysis on Dataset based on sales

2) Sales Trend on Time



In this combo-chart, I have shown the Sales trend over a period of time. This chart show us in which quarter of year we are getting the maximum sales and in which quarter year we are getting the maximum profits and years further divided into four Quarter.

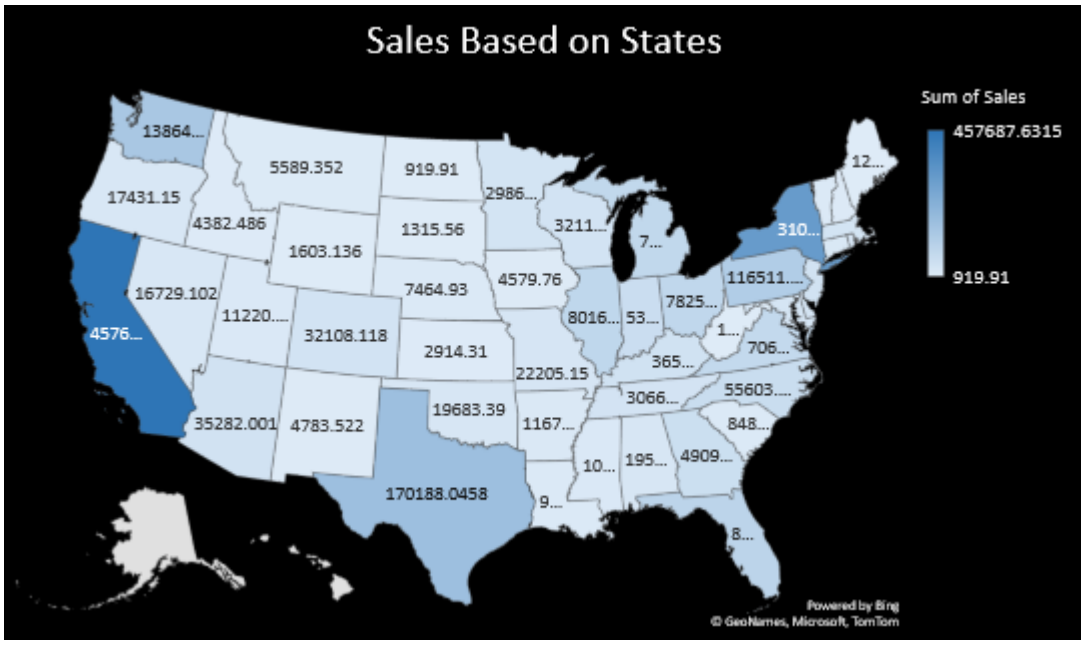
I have shown a line chart showing the sales trend over the quarter of years and it tell us the sum of sales and profits of that particular quarter with **combo chart**.

Basically, the formula used for preparing this chart is: -

- Years & Quarter-> Category Axis
- Sales and profits -> Values Section

Perform Analysis on Dataset based on sales

3)Sales based on State



In this Map, I have shown the total Sales of the products based on state using Map chart.

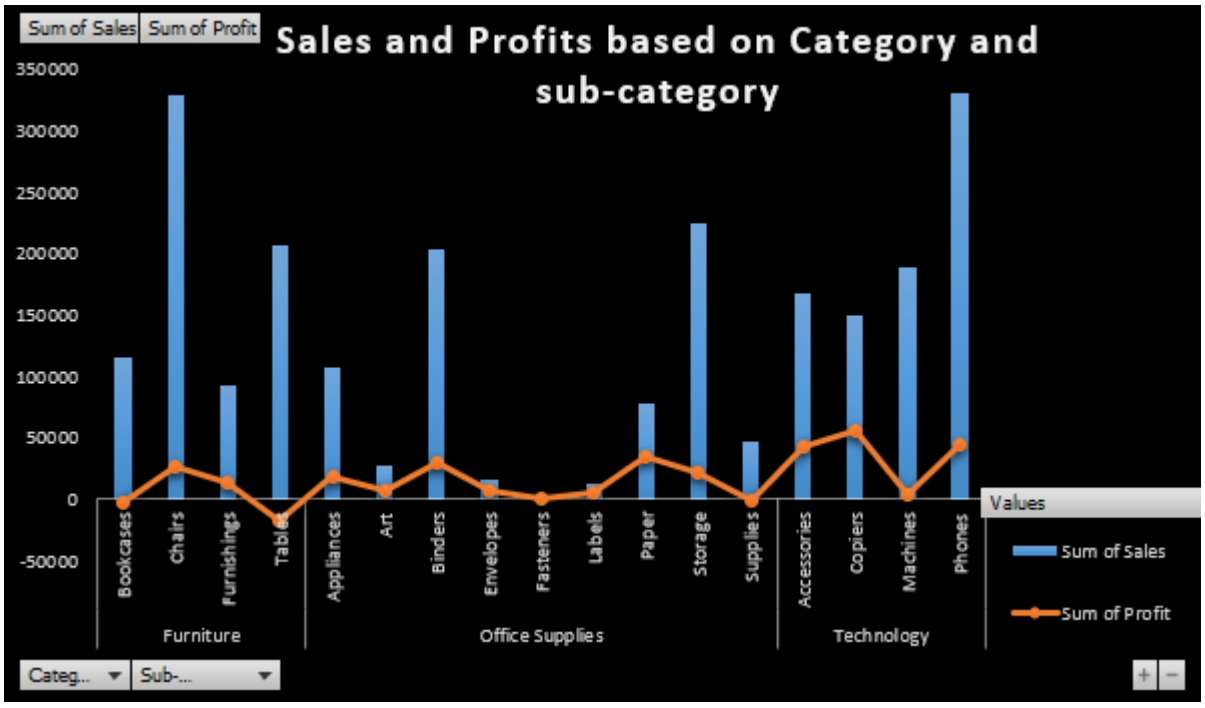
This Map shows us how much benefits we are getting from different states based on total sales of that state and it tell us the sum of sales of that particular state inside **Map** particular state region.

Basically, the procedure for inserting the Map in excel: -

- ✓ **Go to Insert Tab**
- ✓ **Search for chart section**
- ✓ **Click on Map**

Perform Analysis on Dataset based on Category and subcategory

1)Sales and Profit based on category and sub-category



In this **combo-chart**, I have shown the Sales and the Profits based on category and sub-category.

This chart shows us how much benefits we are getting from different category and sub-category, and I also have shown the line chart to show the sales of particular sub-category.

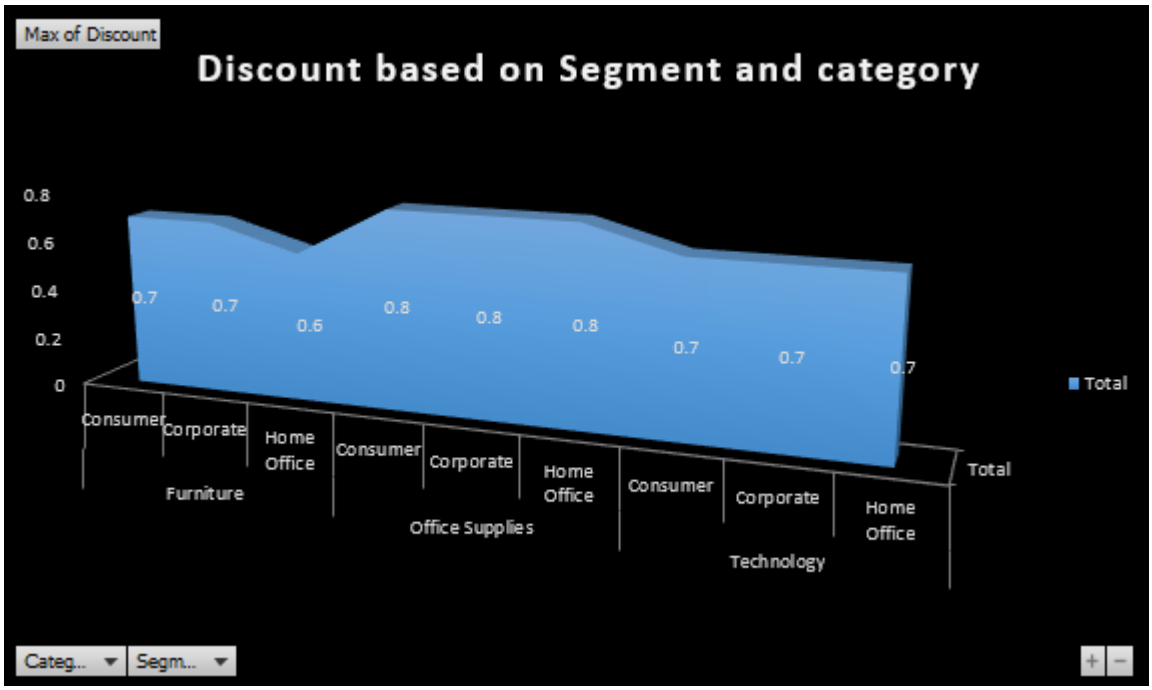
I have shown total sales with the help of clustered column chart and total sales with the help of line marker.

Basically, the formula used for preparing this chart is: -

- Category & Sub-category-> Category Axis
- Sales and profits -> Values Section

Perform Analysis on Dataset based on Category and subcategory

2)Discount based on Segment and Category



In this chart, I have shown the Discount based on Segment and Category.

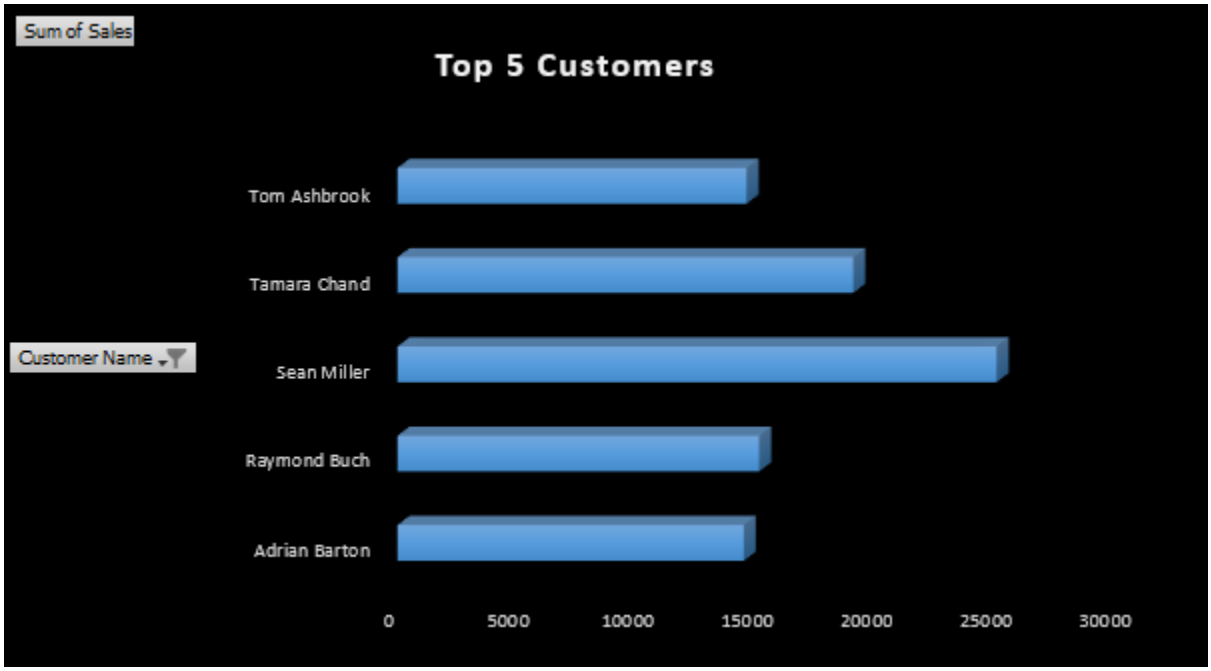
This chart shows us the how much maximum discount the superstore is offering to the different category and segments, and it tell us the Maximum of discount each segment and each category are getting with **Area chart**.

Basically, the formula used for preparing this chart is: -

- Category & Segment -> Category Axis
- Maximum of discount-> Values Section

Perform Analysis on Dataset based on customer segment

1) Top 5 customers



In this chart, I have shown the Top 5 (five) customers based on sales.

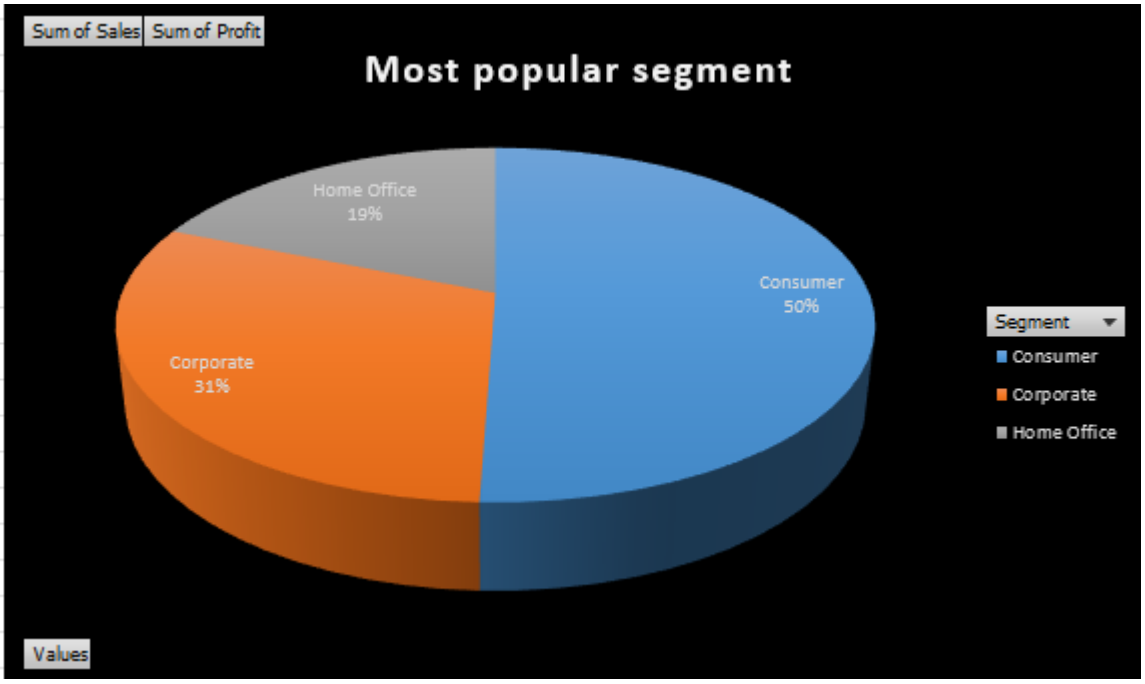
This chart shows us out of all customers which 5 customers that are doing Most sales based on sales and it tell us the sum of sales of those 5 customers with **bar chart**.

Basically, the formula used for preparing this chart is: -

- Customer Name -> Category Axis
- Sum of Sales-> Values Section

Perform Analysis on Dataset based on customer segment

2) Most popular segment



In this chart, I have shown the most popular segment based on sum of Sales and the Profits.

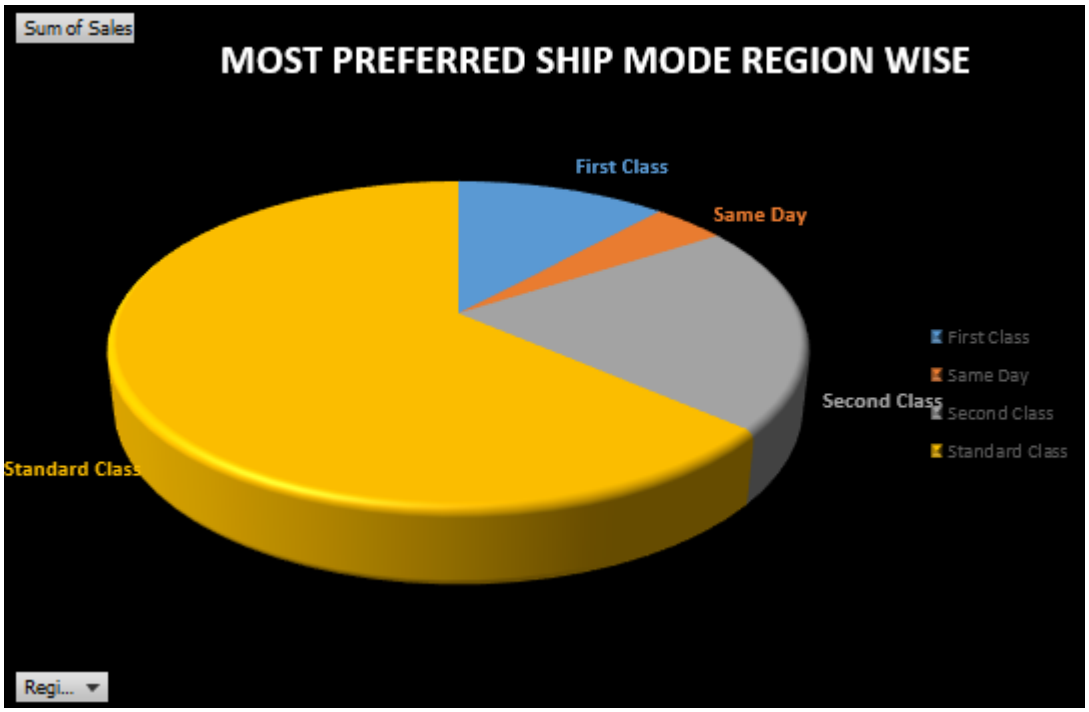
This pie chart shows us the how much benefits out of proportional total benefits we are getting from different segment based on sales and profits and it tell us the sum of sales and profits of that particular segment with **pie chart**.

Basically, the formula used for preparing this chart is: -

- Segment -> Category Axis
- Sales and profits -> Values Section

Perform Analysis on Dataset based on customer segment

3) Most preferred Ship mode by region



In this chart, I have shown the Most preferred Ship mode based on region wise.

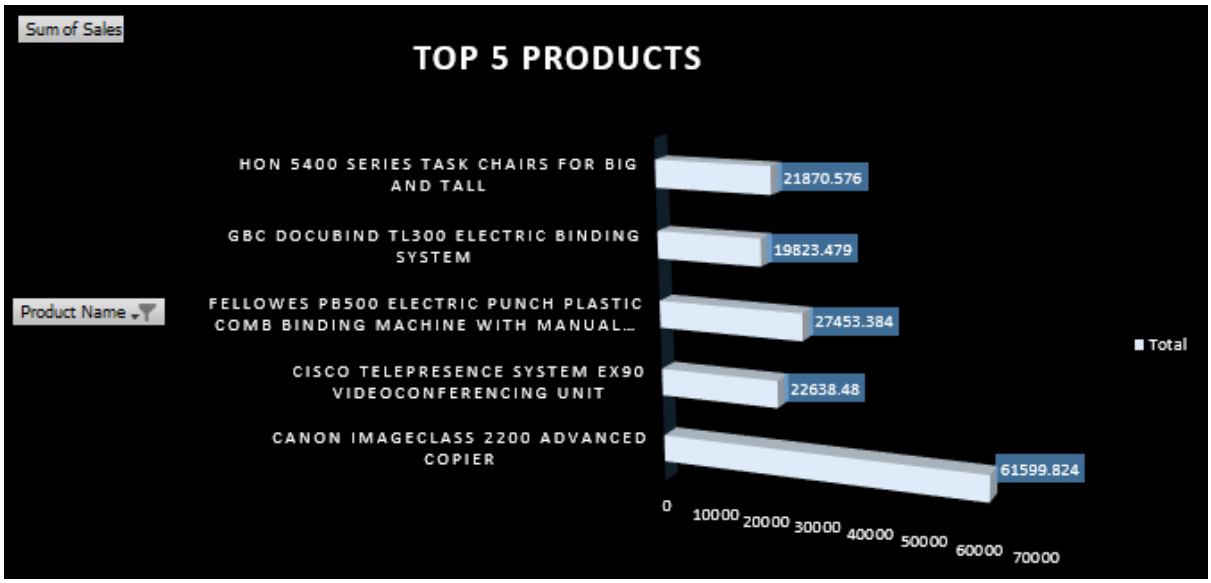
This chart shows us which mode is most preferrable by the customers from different region based on sales and profits and it tell us the total of ship mode preferred of that particular region with **pie chart**.

Basically, the formula used for preparing this chart is: -

- Ship mode-> Category Axis
- Sum of Sales -> Values Section

Perform Analysis on Dataset based on customer segment

4) Top 5 Products



In this chart, I have shown the Top 5 (five) products based on sales.

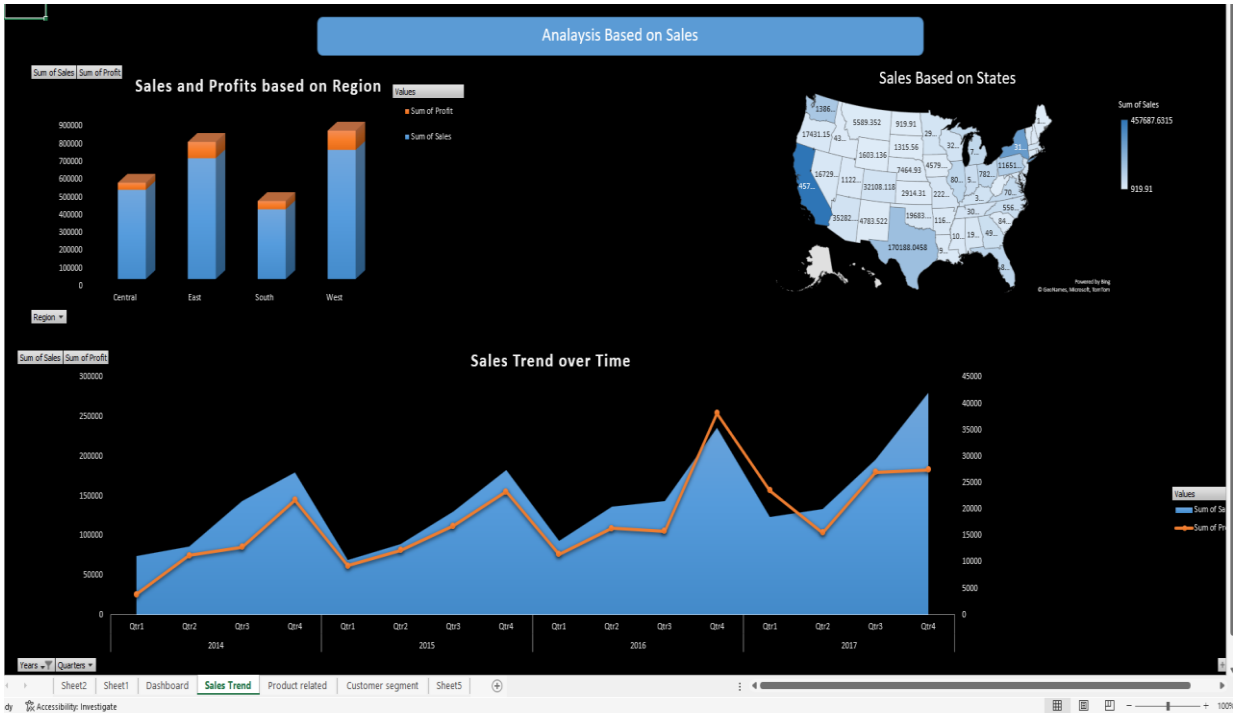
This chart shows us out of all products which 5 products that are liked by the customers most based on sales and it tell us the sum of sales of that 5 products as well with **clustered bar chart.**

Basically, the formula used for preparing this chart is: -

- Product Name -> Category Axis
- Sum of Sales-> Values Section

List of Analysis with result

Summary on Perform Analysis on Dataset based on sales



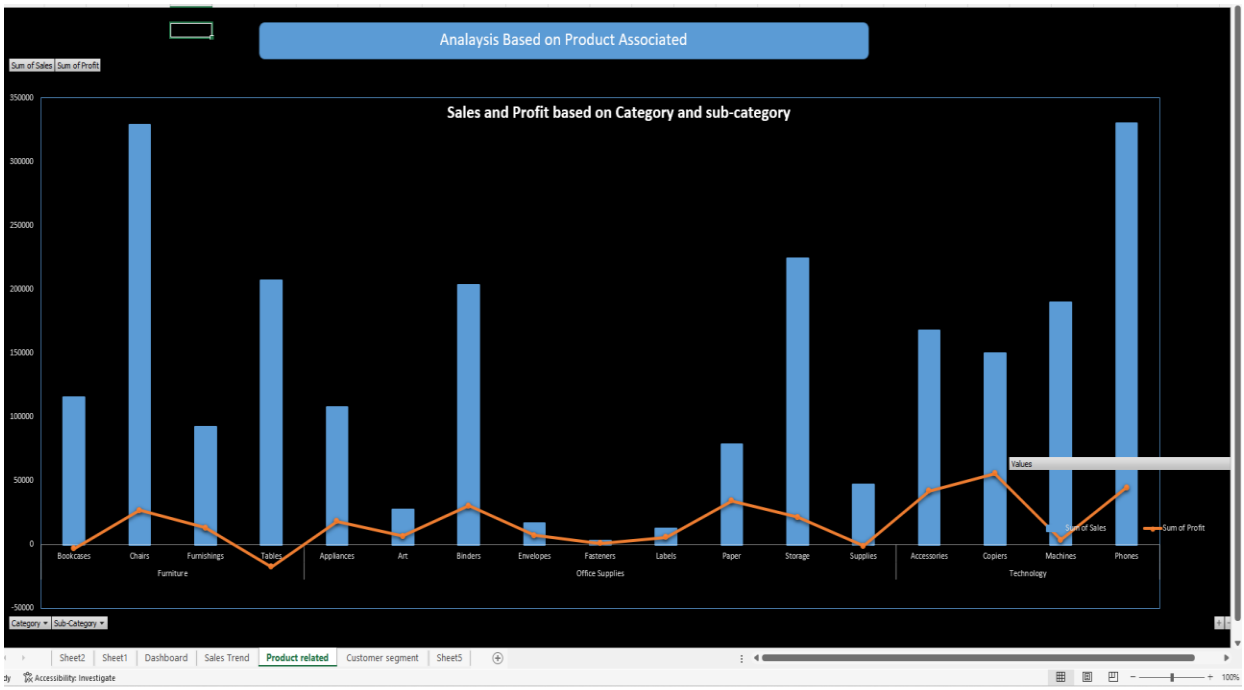
In this Worksheet, I have shown Analysis with respect to Sales: -

- **Sales and Profit based on region**
- **Sales based on State**
- **Sales Trend on Time**

In this worksheet, I have shown summary on analysis based on sales. These three charts show the prediction of sales and profits based on region, states and it predict the sales and profits over quarter of years.

By looking this working , we can easily identify in which region and states and in which period of time, the sales of Superstore increases and decreased and we can take further decision based on that analysis for improving the sales.

Summary on Perform Analysis on Dataset based on products (category and subcategory)



In this Worksheet, I have shown Analysis with respect to products: -

- Sales and Profit based on category and sub-category

In this worksheet, I have shown summary on analysis based on products. Basically, this is a combo chart compromises of Clustered column chart, and Line marker chart which conclude give a combo chart. This chart shows the Sales and profits based on Category and sub-category.

By looking this worksheet, we can easily identify the sales by clustered column chart and profit by line marker chart, and also, we can take further decision based on that analysis for improving the sales of superstore.

Summary on Perform Analysis on Dataset based on customer segmentation

In this Worksheet, I have shown Analysis with respect to customer segmentation: -

- Top 5 customers
- Most popular segment
- Most preferred ship mode by region
- Top 5 products

In this worksheet, I have shown summary on analysis based on customer segmentation. These 5 charts are showing the top 5 customer and top 5 products and Most preferred ship mode region wise and Most popular segment and Discount given to customers based on segment and category.

By looking this worksheet, we can easily identify the customers side, it will give the full details of customer, and also, we can take further decision based on that analysis for improving the interaction with customers.

List of Analysis with result

Here, this is my Resultant Dashboard, in this I have shown the all the charts and map with some filter.

These are the values which dynamically changes after applying the filter: -

- **Average of Sales**
- **Average of Quantity**
- **Average of Discount**
- **Average of Profits**

List of Analysis with result

In this Dashboard, I have applied the filter on various fields and changes occurs in each chart and map: -

Here, I have mentioned some fields like

- **Years**
- **Segment**
- **Region**
- **Category**
- **Ship Mode**
- **Months**
- **States**

References and Bibliography

- www.google.com
- https://www.tutorialspoint.com/excel/index.htm
- https://en.wikipedia.org/wiki/Microsoft_Excel
- https://support.microsoft.com/en-us/office/excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188
- PowerPoint presentation of ETL