**INTRODUCTION TO DATA SCIENCE**

**SALES DATA ANALYSIS PROJECT**

Submitted By

**RISHABH GUPTA**

P.G – Diploma in DATA SCIENCE

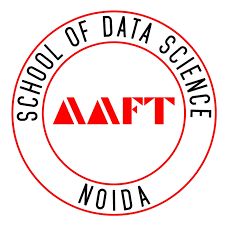
Under the Guidance of

**AISHWARY SHUKLA**

Professor of introduction of data science

AAFT School Of Data Science

**AAFT UNIVERSITY, NOIDA**



**AKNOWLEDGEMENT**

The opportunity of attaining a course based on Data Science using AAFT was worth learning. It was a prestige for me to be part of it. During the period of my course, I received tremendous knowledge related to Microsoft Excel and Data Management.

Pre-eminently, I would like to express my deep gratitude and special thanks to my course teacher Aishwary Shukla for her theoretical knowledge and encouragement on this project and for her valuable guidance and affection for the successful completion of this project.

Secondly, I would like to thank AAFT for giving me an opportunity to learn this course. Lastly, I would like to thank the almighty and my parents for their constant encouragement, moral support, personal attention, and care.

**Rishabh Gupta**

**P.G Diploma in Data Science**

**AAFT , Noida , UP**

**CONTENT**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
| 01.0 | Cover Page | 1 |
| 02.0 | Acknowledgement | 2 |
| 03.0 | Content | 3 |
| 04.0 | Abstract | 4 |
| 05.0 | Introduction | 5 |
| 06.0 | Important Words | 6 |
| 07.0 | Data Analysis | 7 |
| 08.0 | Data Visualization | 7 |
| 09.0 | Conclusions from the analysis: (compare result with scatter plot) | 12 |
| 10.0 | Sales Analysis | 12 |
| 11.0 | Profitability Analysis | 16 |
| 12.0 | Customer Analysis | 17 |
| 13.0 | Product Analysis | 18 |
| 14.0 | Shipping Analysis | 19 |
| 15.0 | Geographical Analysis | 21 |
| 16.0 | Order Analysis | 22 |
| 17.0 | Managerial Analysis | 23 |
| 18.0 | Pivot Table | 23 |
| 19.0 | Pivot Chart | 26 |
| 20.0 | Dashboard | 30 |
| 21.0 | Summary of Project (Conclusion) | 31 |
| 22.0 | Reference | 32 |

**ABSTRACT**

Excel is a software program created by Microsoft that uses

spreadsheets to organize numbers and data with formulas and functions. Excel analysis is ubiquitous around the world and used by businesses of all sizes to perform data analysis. Excel features calculation, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications, and several other features which make Excel a perfect choice to manage and analyze data. My project is an Excel Dashboard. The Excel Dashboard is used to display overviews of large data tracks. Excel Dashboards use dashboard elements like tables, charts, and gauges to show the overviews. The dashboards ease the decision-making process by showing the vital parts of the data in the same window. In this report, I have shared a project where I have done data analysis of Sales data set. This report also presents my learning during my course classes.

**INTRODUCTION**

I have created an Excel dashboard of Sales data set. This dashboard explains and highlights important facts, records, and trends in the Sales history.

The data set used contains information regarding all the previous Summer Sales. It includes information regarding Customer id, Customer name, Order priority, Discount, Unit Price, Shipping cost, Ship Mode, Customer Segment, Product Category, Product Sub- Category, Product Container, Product name, Product base margin, Country, Region, State, City, Postal Code, Order date, Ship date, Profit, Quantity order new, Sales, Order id, Total, Manager. The data set contains details of Approx. **50336** orders.

I have scrubbed and organized the entire data set and performed the analysis of a clean data set. I have deduced and calculated important results from the data set with the help of various Excel features like pivot tables and functions and represented them in the form of a dynamic dashboard using excel visualizing tools and various charts.

Important Word Meaning That You Should Know –

Analysis-  Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

Dashboard- Data dashboards are a summary of different, but related data sets, presented in a way that makes the related information easier to understand .

Mean- Mean is defined as the average of the given numbers. Mean is the sum of all the given data values divided by the total number of data values given in the set.

Median- Median, in statistics, is the middle value of the given list of data when arranged in an order.

Mode- The mode is the most common number that appears in your set of data.

Standard deviation- A standard deviation (or σ) is a measure of how dispersed the data is in relation to the mean

Quartiles- In statistics, Quartiles are the set of values which has three points dividing the data set into four identical parts.

Descriptive Statistics- Descriptive statistics refers to a set of methods used to summarize and describe the main features of a dataset, such as its central tendency, variability, and distribution.

Visualization- Data visualization is the process of using visual elements like charts, graphs, or maps to represent data.

Bar Charts- A bar chart is a way of summarizing a set of categorical data (continuous data can be made categorical by auto-binning).

Scatter plots- A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables.

Box plot- A box and whisker plot or diagram (otherwise known as a boxplot), is a graph summarising a set of data.

Revenue- the total amount of money brought in by a company's operations, measured over a set amount of time.

Trends- A trend is a general change in one variable compared to another over a period of time.

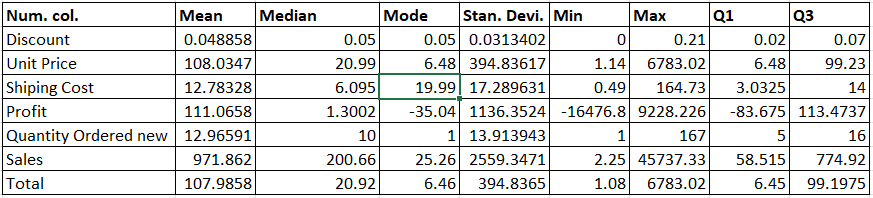
Profitability- Profitability is a measure of an organization's profit relative to its expenses.

Correlation- Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate).

Operational inefficiencies - Operational inefficiencies occur when a business uses more resources than are necessary to maintain or improve its output—be that time, labor, or material goods -and it becomes evident that there's room for optimization.

★ Data analysis:

Perform descriptive statistics (Measures: mean, median, mode, standard deviation, minimum, maximum, and quartiles) for numerical columns:

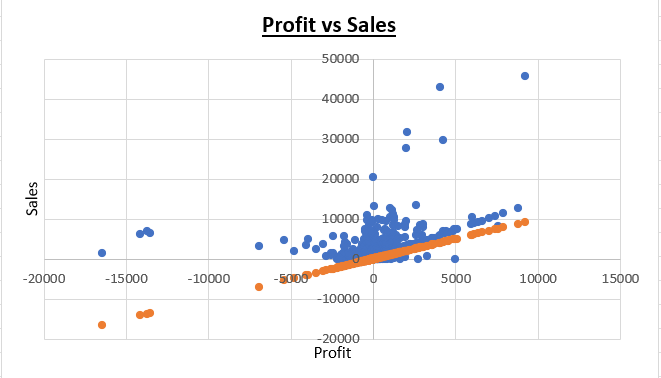


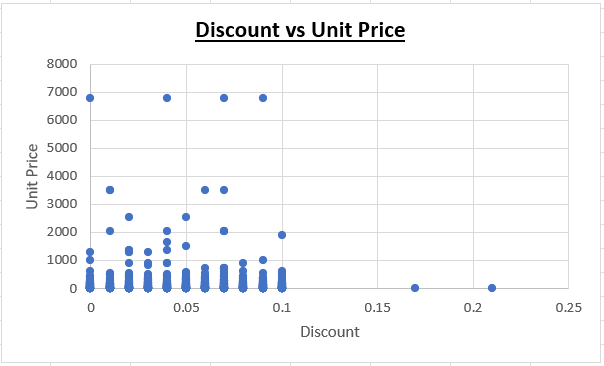
★ Data Visualization:

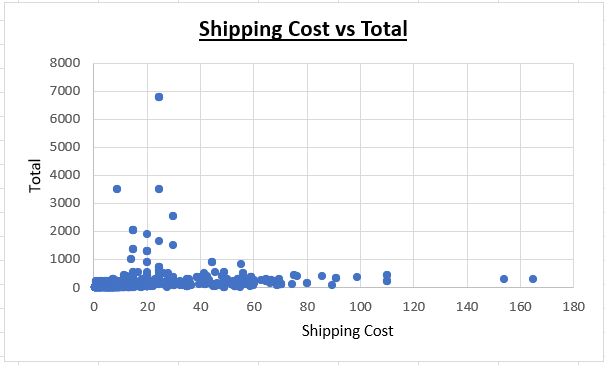
Create various charts and graphs to visualize different aspects of the data. For example.

1. Bar charts for categorical variables like Order Priority, Ship Mode, Customer Segment, etc.

2. Scatter plots to explore relationships between numerical variables like Profit vs. Sales.

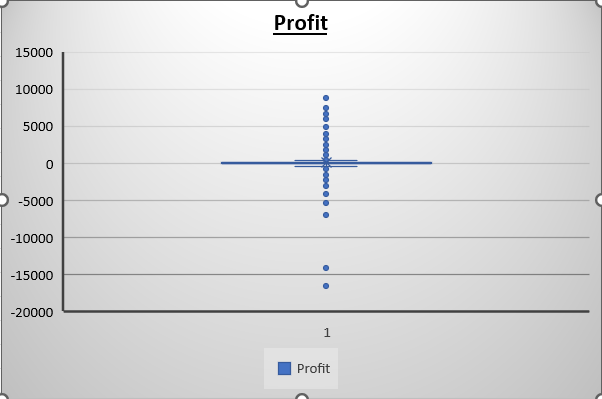




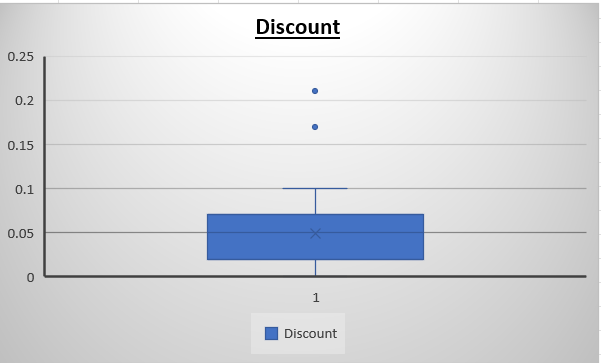


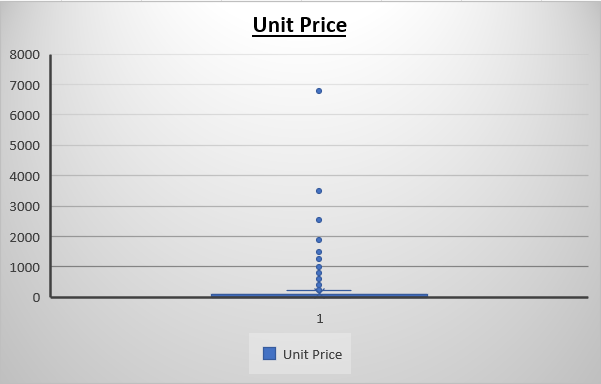
3. Boxplots to identify outliers in numerical data.

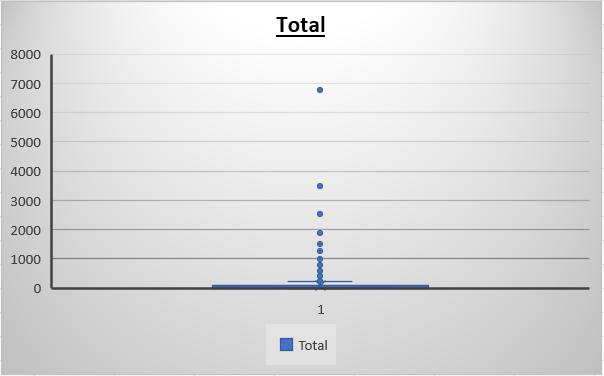
Ans-











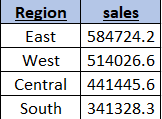


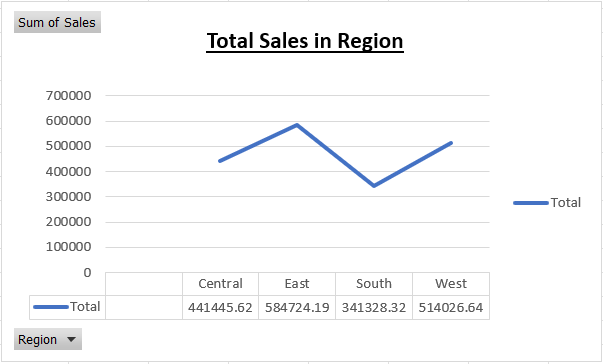
★ **Conclusions from the analysis: (compare result with scatter plot**

**01.Sales Analysis:**

a. What is the total sales revenue for the dataset?

Ans- Total sales revenue = 1881524.77

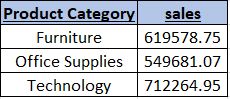


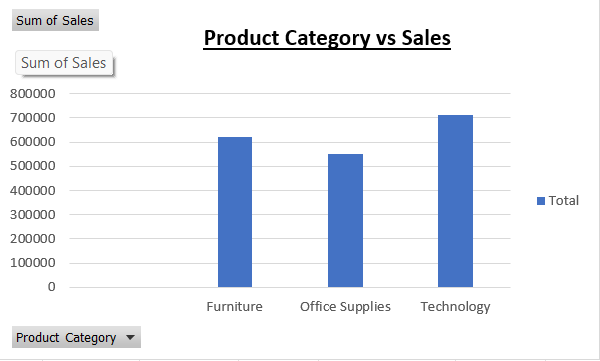


Here you can see the total sales in different regions like central, east, south, west.

b. Which product category generates the highest sales revenue?

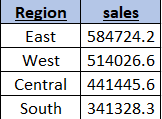
Ans- Technology generates the highest sales revenue.

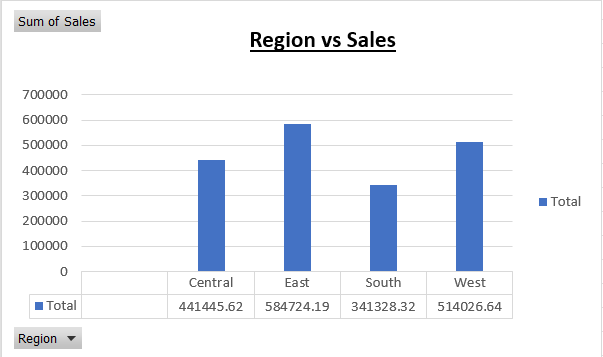




c. How do sales vary by region or country?

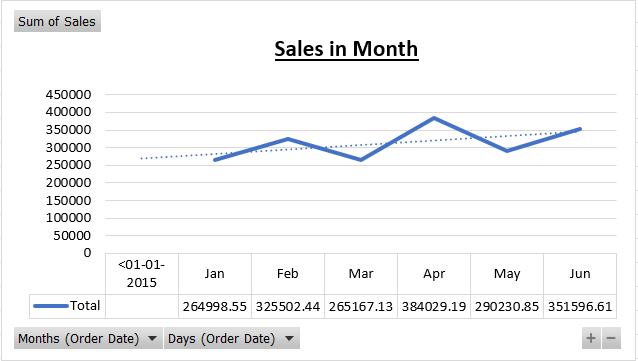
Ans- We have only one country which is United states and region are east, west, central, south. So that’s why sales vary by regions.



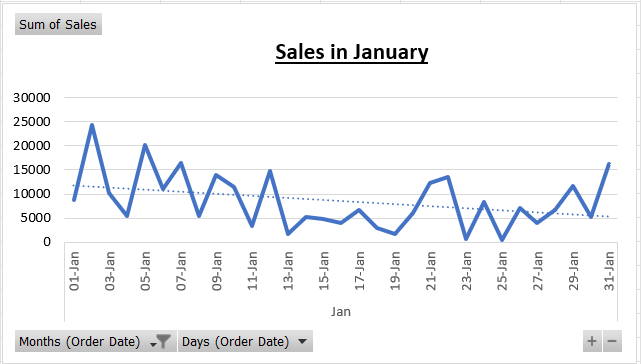


d. Are there any trends or patterns in sales over time?

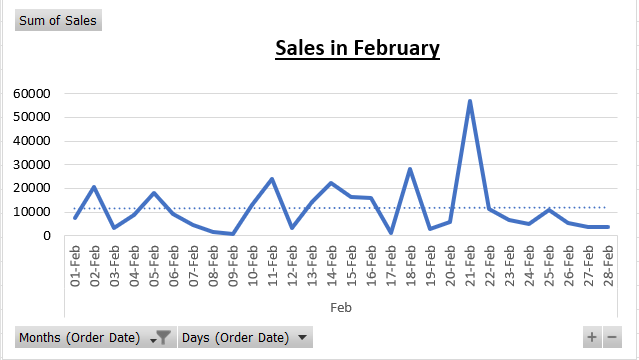
Ans-



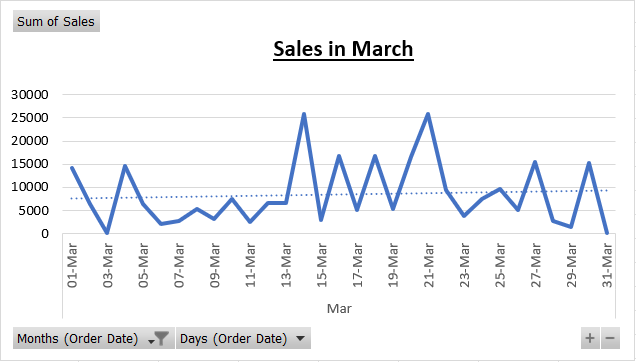
It shows trends of sales in month, which is going up(good).



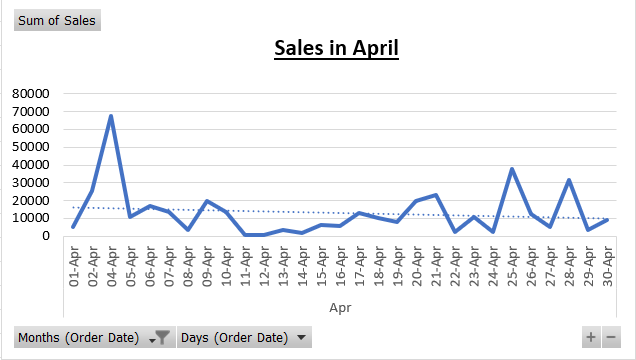
It shows trend of sales in January data, which is going down(bad).



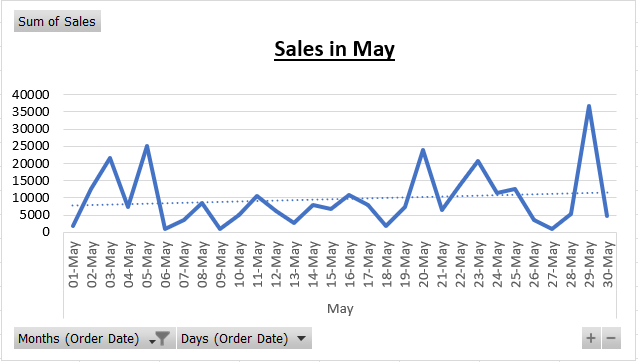
It shows trend of sales in february data, which is stable .



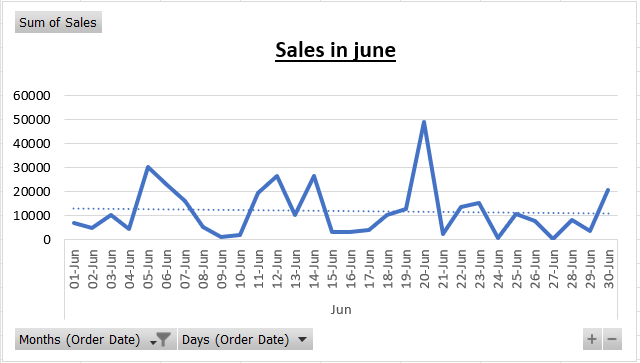
It shows trend of sales in march data, which is going up.



It shows trend of sales in April data, which is going down.



It shows trend of sales in May data, which is going up.

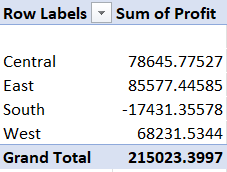


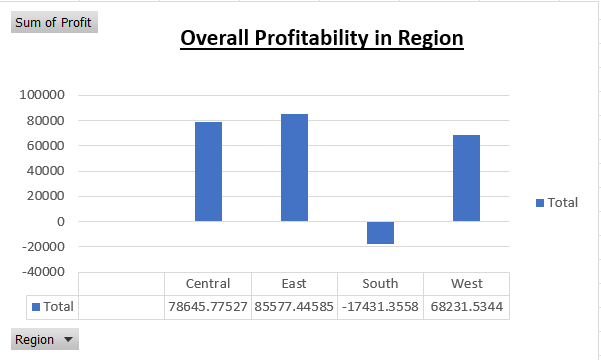
It shows trend of sales in June data, which is going down.

**02.Profitability Analysis:**

a. What is the overall profitability of the business?

Ans-





It shows overall profitability in different regions and the highest profit shows in east region and lowest in south region.

b. Which products have the highest profit margins?

Ans- Binders and Binder Accessories

c. Are there any customer segments or regions that are more profitable than others?

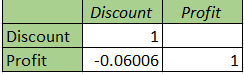
Ans- Corporate in East

A graph of a bar chart

Description automatically generated with medium confidence

d. Is there a correlation between discounts offered and profitability?

Ans-

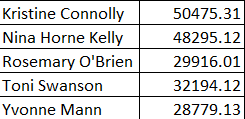


Here you can see the correlation between discount offered and profitability is -0.06006 .

**03.Customer Analysis:**

a. Who are the top customers by total sales?

Ans-



Here are top 5 customers by total sales

b. What is the average order size for different customer segments?

Ans- Average order size = 12.96591

A table with text and numbers

Description automatically generated with medium confidence

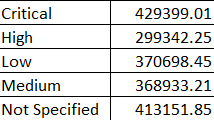
A screenshot of a graph

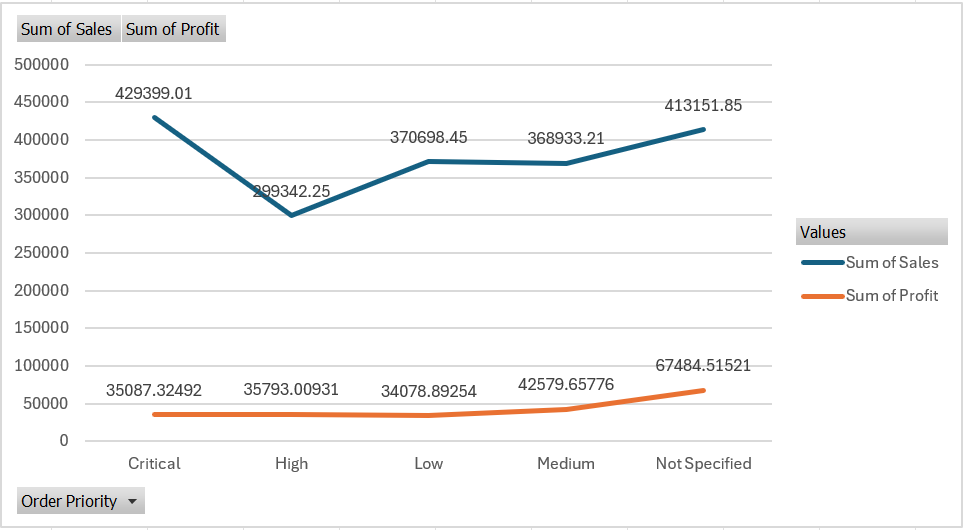
Description automatically generated

As you can see the small business has highest and corporate has lowest average quantity order

c. How does order priority impact sales or profitability?

Ans- If order priority is high then it increase the sales and also profitability.



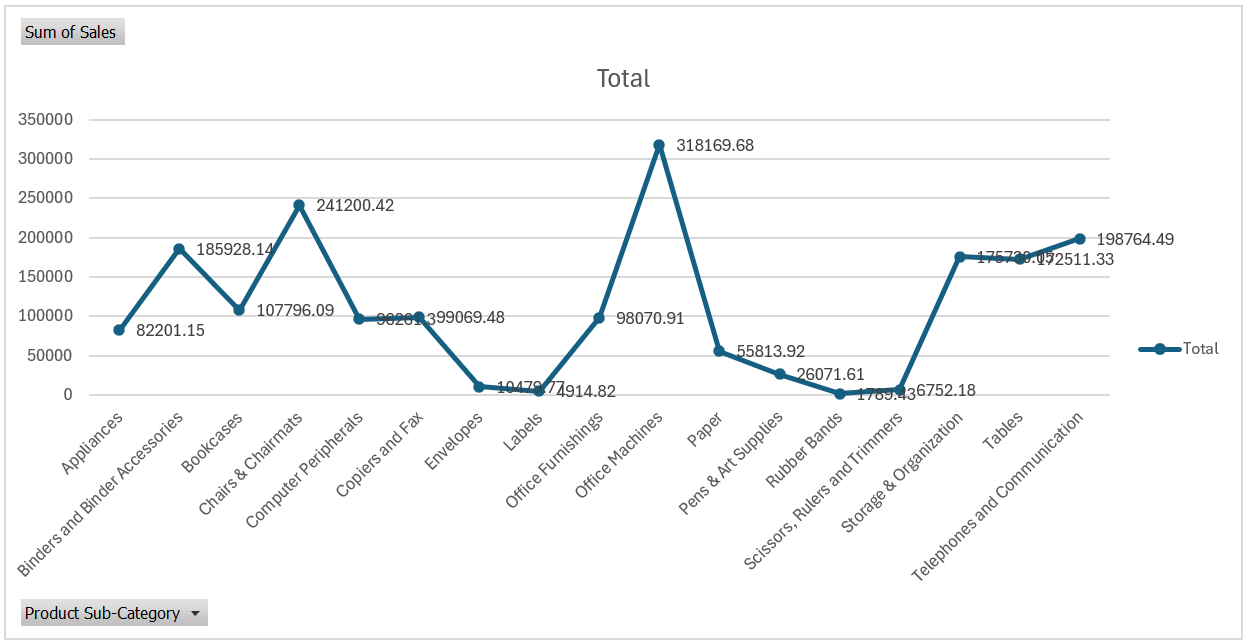


It shows sum of sales and profit with order priority

**04.Product Analysis:**

a. Which product sub-categories are the best-selling?

Ans- Office Machines



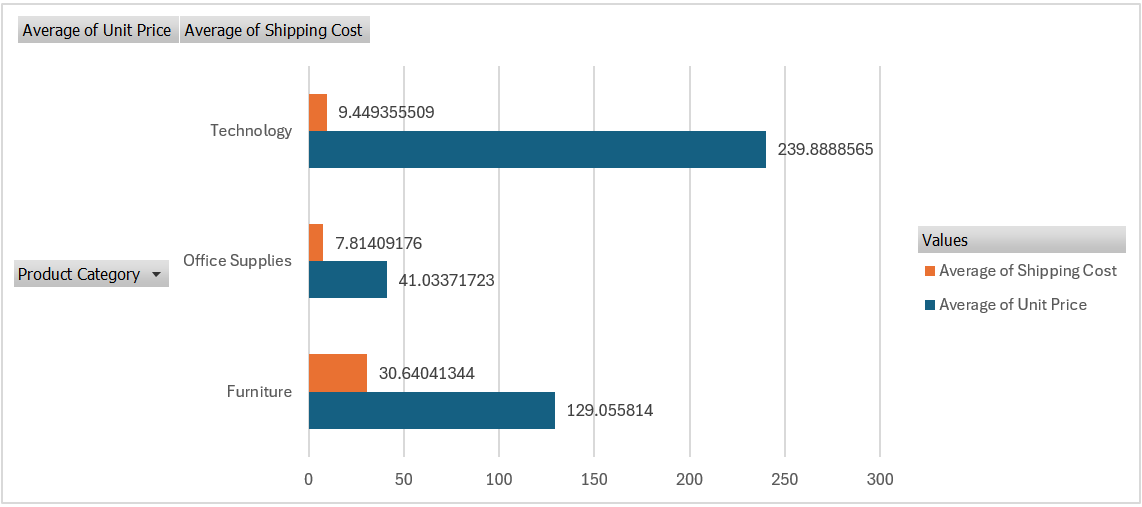
As you see in line chart office machines has best selling product.

b. What is the average unit price and shipping cost for each product category?

Ans-

A blue and white rectangular sign with black text

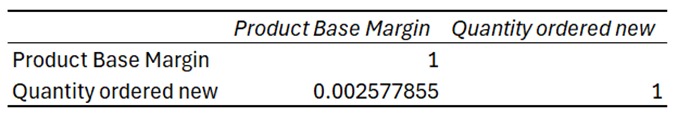
Description automatically generated



As you see the bar chart shows Average of unit price and shipping cost with product category

c. Is there a relationship between product base margin and sales volume?

Ans-



As you see the chart shows correlation between product base margin and quantity ordered new

**05.** **Shipping Analysis:**

a. Which shipping mode is most commonly used?

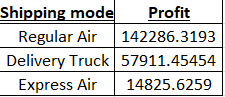
Ans- Regular Air

A graph with blue bars

Description automatically generated

b. Does the choice of shipping mode affect profitability or customer satisfaction?

Ans- yes the choice of shipping mode affect profitability



A graph with numbers and a bar

Description automatically generated

c. Are there any trends in shipping costs over date or across regions?

Ans- Yes shipping cost vary across region.

A graph showing the number of months and months

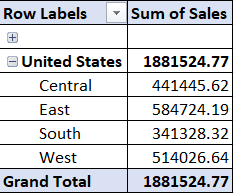
Description automatically generated with medium confidence

As you see the line chart it shows that in July trend goes down in every region

**06.Geographical Analysis:**

a. Which regions or countries contribute the most to sales or profitability?

Ans-



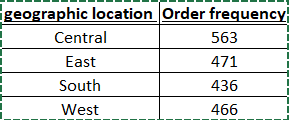
A graph of different colored bars

Description automatically generated with medium confidence

As you see this bar chart shows most sales and profit with regions

b. Is there a correlation between geographic location and order frequency?

Ans-



As you see the order frequency with different geographic location

**07.Order Analysis:**

a. What is the average order value of products ?

Ans- Average Order value = 8.332211

A blue and white rectangular object with black text

Description automatically generated

A graph of a bar chart

Description automatically generated with medium confidence

As we show here average of order value of product category

b. How does order volume vary by month or season?

Ans- order volume vary by month

.A screenshot of a table

Description automatically generated

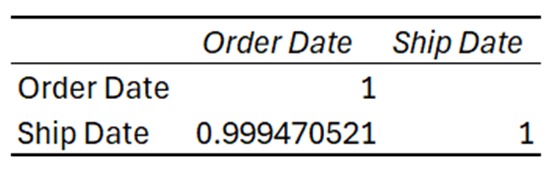
A graph with a line and numbers

Description automatically generated

As you see the count of orders changes in months

c. Are there any correlations between order date and shipping date?

Ans-



As you see in this chart gives us the correlation between order date and shipping date

**08.Managerial Insights:**

a. Based on the analysis, what recommendations can be made to improve profitability?

Ans- To improve profitability you have to increase the unit price and quantity of order.

b. Are there any operational inefficiencies that can be addressed?

Ans- In Shipping mode you can make your deployment in time so that your operational inefficiencies will decrease .

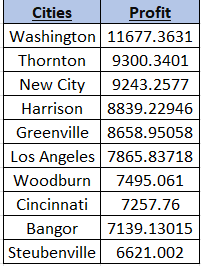
c. How can the business better tailor its offerings to meet customer needs?

Ans- Conduct market research, analyze customer feedback, and gather data to identify their pain points, preferences, and desires. By gaining insights into what your audience truly wants, you can tailor your offerings to meet their specific needs.

**★ Pivot Table: 1.**

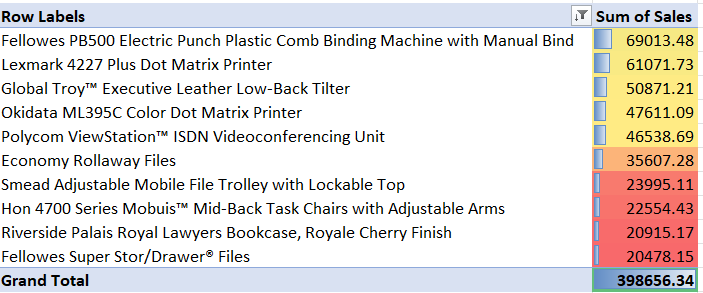
1. Which are the top 10 cities in terms of profit?

Ans-



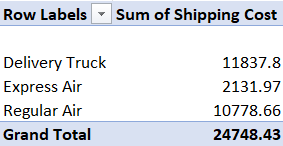
2. What are the most revenue-generating products in terms of profit?

Ans-



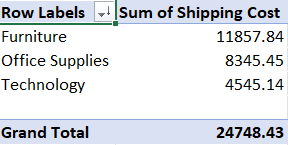
3. How does shipping expense vary for different transportation methods?

Ans-



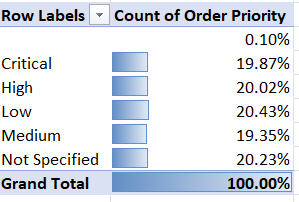
4. How do shipping charges vary for different products?

Ans-



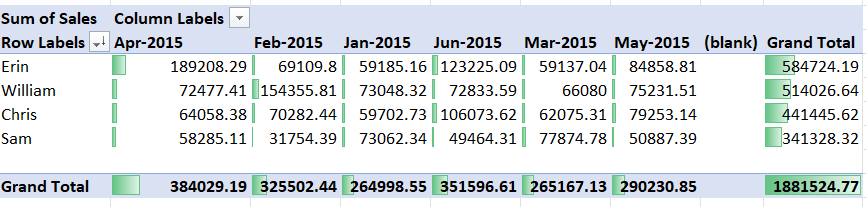
5. What is the percentage of orders based on priority?

Ans-



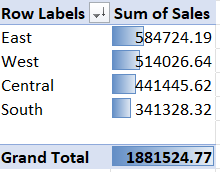
6. How does the performance of managers vary across months (Jan-Jun)?

Ans-



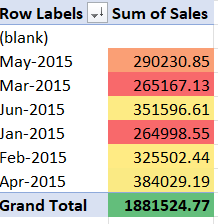
7. How do sales vary across different regions?

Ans-



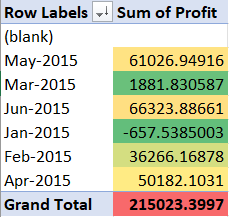
8. What is the trend of sales from January to June?

Ans-



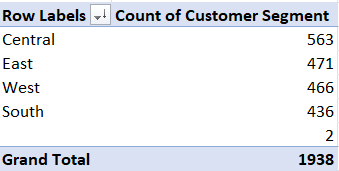
9. How does monthly profit vary throughout the first half of the year?

Ans-



10. How can customers regions be analyzed based on segments?

Ans-

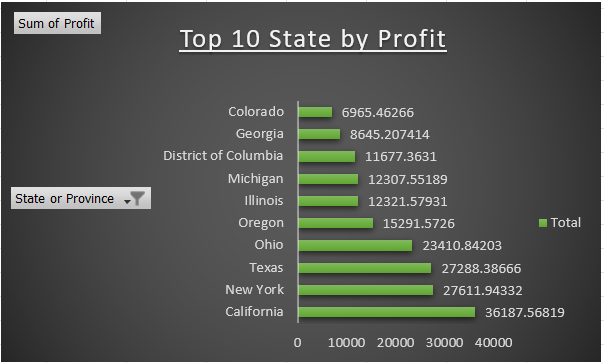


This explans the

**★ Pivot Charts:**

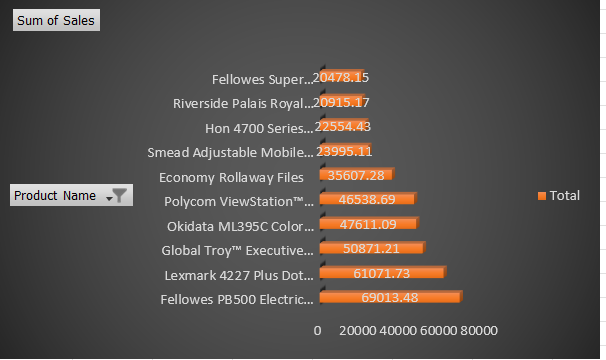
1. How can we visually represent the top 10 States in terms of profit on the dashboard?

Ans-



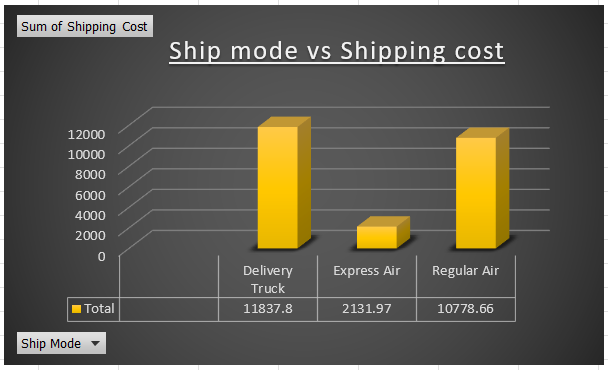
2. What visualizations can be used to display the most revenue-generating products?

Ans-



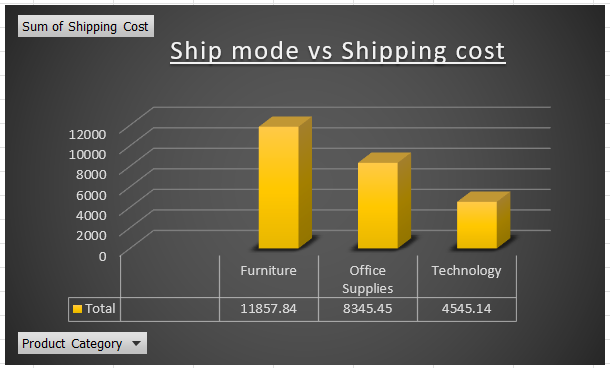
3. How can we visualize the shipping expense for different transport methods?

Ans-



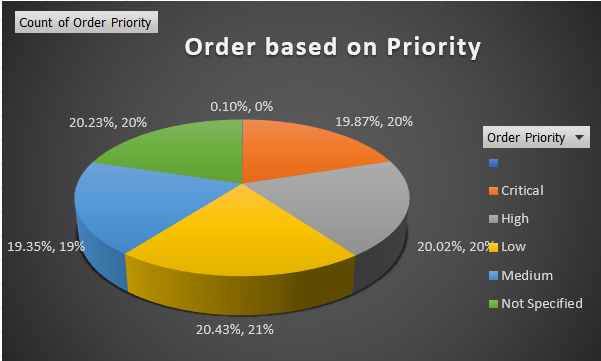
4. What type of chart would effectively represent shipping charges for different products?

Ans-



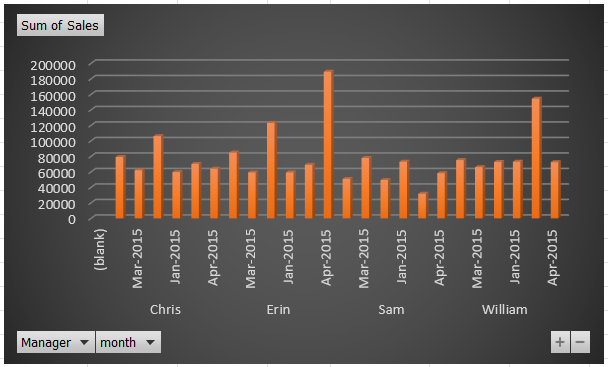
5. How can we visually display the percentage of orders based on priority?

Ans-



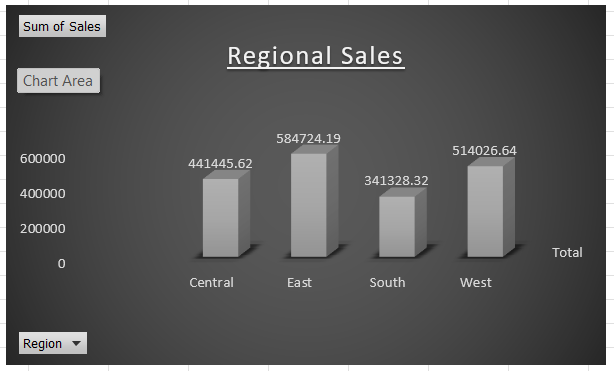
6. What type of chart or graph can effectively showcase the performance of managers for each month?

Ans-



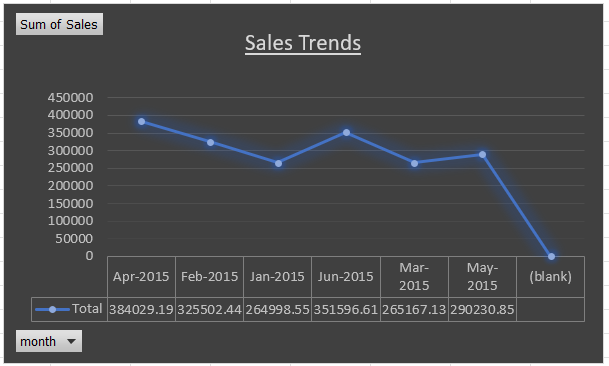
7. How can regional sales be displayed on the dashboard for easy interpretation?

Ans-



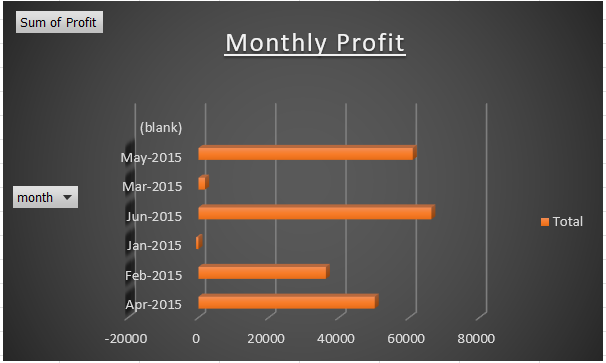
8. What type of chart can effectively represent the sales trend from January to June?

Ans-



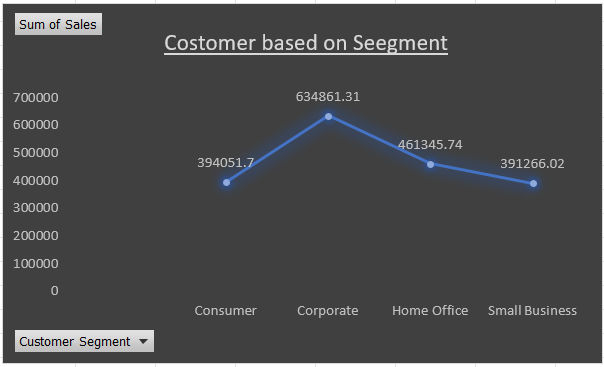
9. How can we visualize monthly profit variations on the dashboard?

Ans-



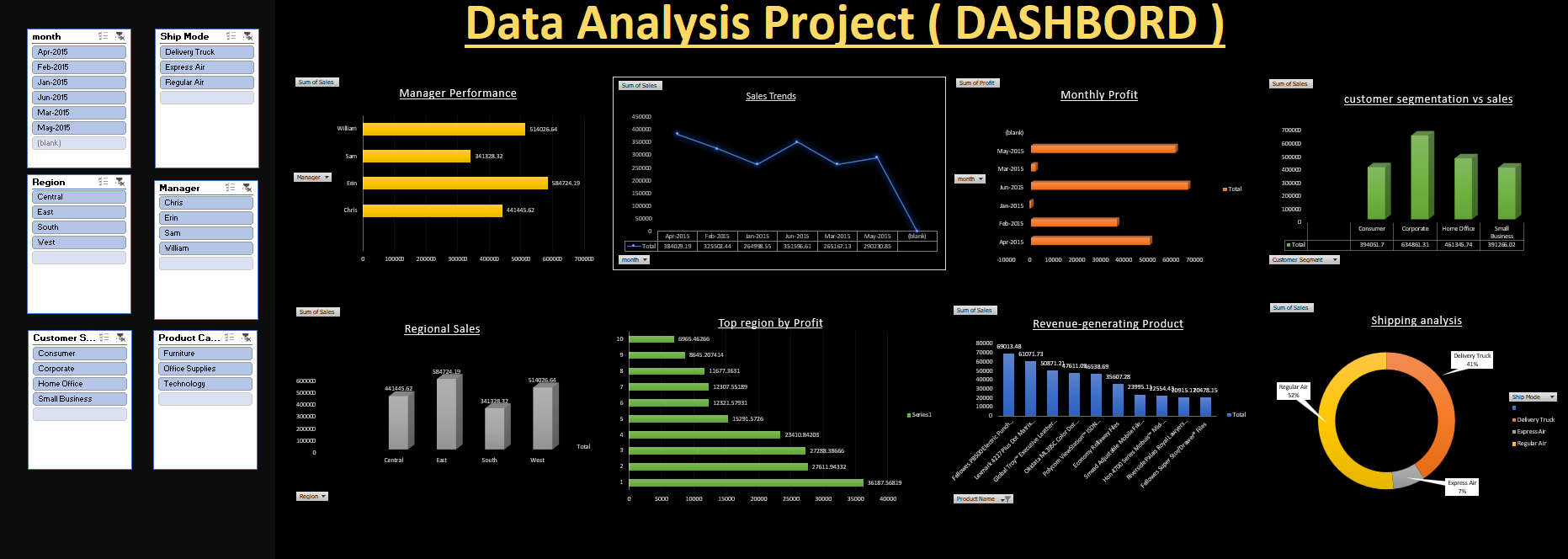
10. What visualizations can be used to analyze customers based on segments?

Ans-



**★ DashBoard:**

A dashboard that provides an overview of all key metrics, such as top regions by profit, revenue-generating products, shipping analysis, manager performance, regional sales, sales trend, monthly profit, and customer segmentation, in one comprehensive sheet?

****

**PROJECT SUMMARY**

**( CONCLUSION )**

In this report I have discussed in detail my project, its working,

making, features, and applications. I have explained each questions using from a raw data using Excel.

This report highlights all the processes involved in the making in

serial order from ETL processes extracting, transforming, and loading

data to using several excel features like pivot table, filtering, sorting,

formulas to perform data analysis and deduce important results then

representing them graphically using charts.

I have also attached the preview of the dashboard, and all the

objectives in this report.

REFERENCES

* Using Sales data sheet : <https://docs.google.com/spreadsheets/d/1B7w444UZ1eWQtnE3Z5Cmozipr3RkOAhe/edit?usp=drive_link&ouid=109618786931407309161&rtpof=true&sd=true>
* https://drive.google.com/drive/folders/1k9KYGnlfZpEAUau1y2VIznwD8eNDqcNt?usp=sharing