**INTRODUCTION TO DATA SCIENCE**

**AN EXCEL DATA ANALYSIS**

**Submitted by:**

**Gaurav Kamal**

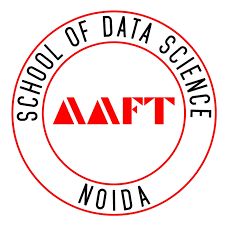
**Diploma in Data Science**

**Submitted to:**

**Mr. Aishwary Shukla**

**Discipline of Introduction to Data Science**

**SCHOOL OF DATA SCIENCE AAFT NOIDA**



**THE ASIAN ACADEMY OF FILM AND TELEVISION**

**DECLARATION**

I Gaurav Kamal, a student of Diploma in Data Science under Diploma Discipline at Asian Academy of Film And Television, Noida, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

**GAURAV KAMAL**

**DIPLOMA IN DATA SCIENCE**

**ASIAN ACADEMY OF FILM & TELEVISION**

**28-05-2024**

**Acknowledgement**

The opportunity of attending a course based on Data Science using Excel at 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 Analysis.

Pre-eminently, I would like to express my deep gratitude and special thanks to my course teacher Mr. Aishwary Shukla for his theoretical as well as practical knowledge and encouragement on this project and for his 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.

**CONTENT**

|  |  |  |
| --- | --- | --- |
| **SR. NO.** | **TOPIC** | **PAGE NO.** |
|  | ABSTRACT | 5 |
| 1 | Sales Data Analysis | 6 - 46 |
|  | * Data Analysis | 6 |
|  | * Data Visualization | 7 - 15 |
|  | * Conclusion for the Analysis | 16 - 31 |
|  | * Pivot Tables | 32 – 35 |
|  | * Pivot Charts | 36 - 45 |
|  | * Dashboard | 46 |

**ABSTRACT**

Microsoft Excel is a robust spreadsheet program that allows users to organize data, perform calculations, and create visual representations through charts and graphs. Pivot tables in Excel facilitate data summarization and analysis, while pivot charts provide visual insights into the summarized data. Users can leverage these tools to extract meaningful information from large datasets efficiently. Excel also enables the creation of interactive dashboards for consolidating key metrics and data visualizations, aiding in performance tracking and trend identification. Through its diverse range of features, Excel serves as a valuable tool for data management, analysis, visualization, and decision-making processes.

**SALES DATA ANALYSIS**

* **Data analysis:**

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

1. Discount
2. Unit Price
3. Shipping Cost
4. Profit
5. Quantity ordered new
6. Sales
7. Total

Ans.

A data analysis with numbers and text

Description automatically generated with medium confidence

* **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.

A graph of blue rectangular bars

Description automatically generated with medium confidence

Order Priority

A graph with numbers and a bar

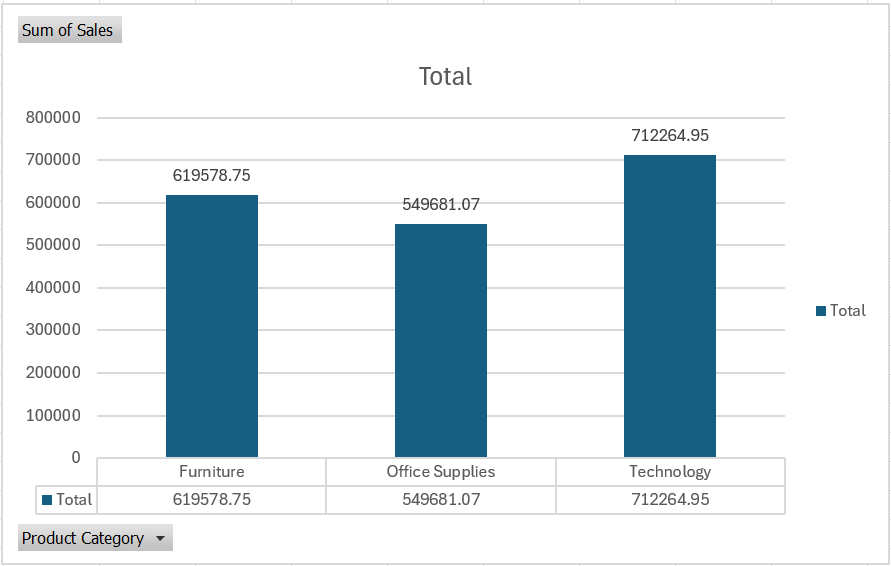
Description automatically generated

Ship Mode

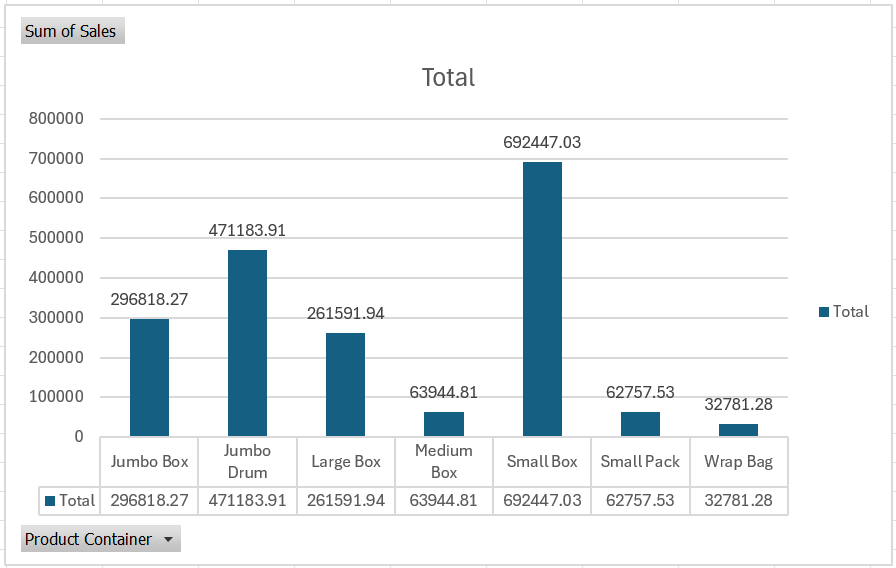
A graph of numbers and a bar

Description automatically generated

Customer Segment



Product Categories

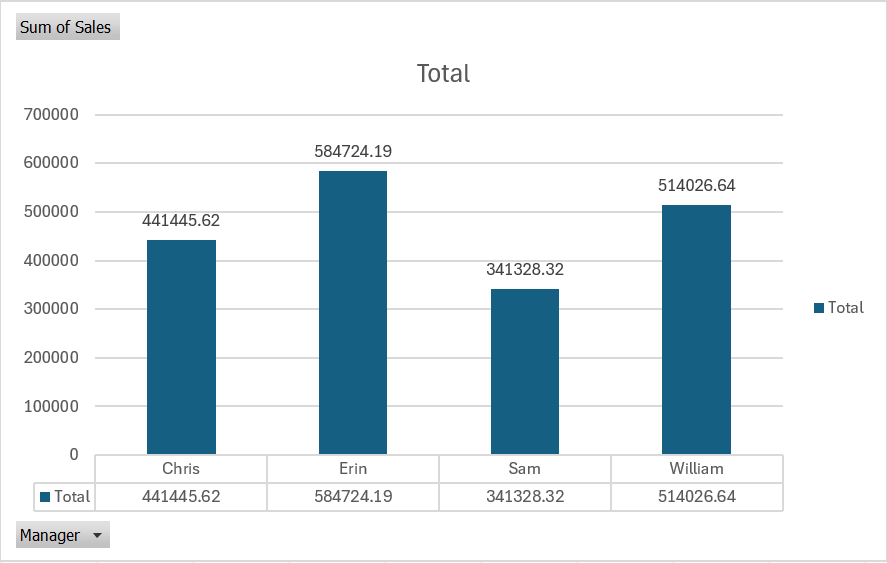


Product Container

A graph of blue rectangular bars

Description automatically generated with medium confidence

Region



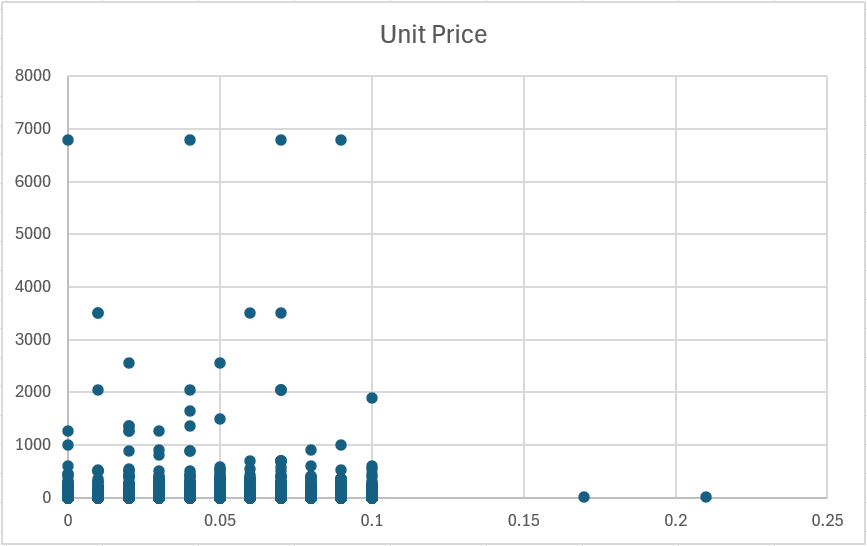
Manager

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

A graph with numbers and dots

Description automatically generated

Profit vs. Sales



Unit Price vs Discount

A graph with blue dots

Description automatically generated

Total vs. Shipping Cost

1. Box plots to identify outliers in numerical data.

A graph with a bar and a number

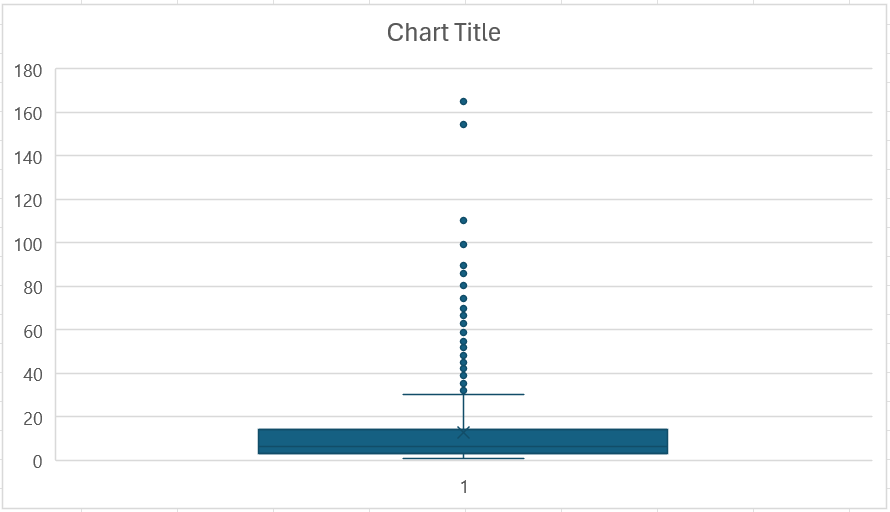
Description automatically generated with medium confidence

Discount

A graph with blue dots

Description automatically generated

Unit Price



Shipping Cost

A graph with a bar and a line

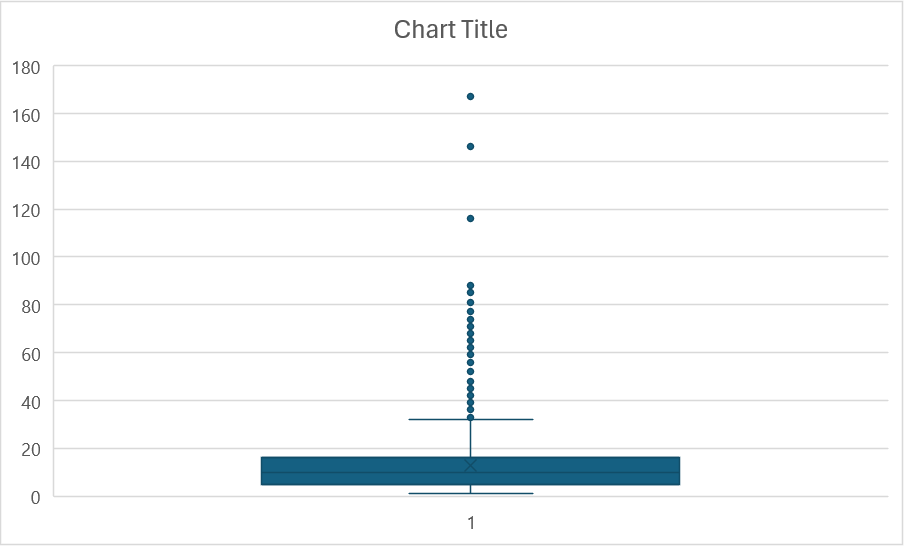
Description automatically generated with medium confidence

Base Margine

A graph with a cross

Description automatically generated

Profit



Quantity Ordered New

A graph with blue dots and white lines

Description automatically generated

Sales

A graph with blue dots and white lines

Description automatically generated

Total

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

1. **Sales Analysis:**
2. What is the total sales revenue for the dataset?

Ans.  The total sales for the dataset are = 1881524.77. Below are mentioned the sales by region. The graph shows that the southern region generated less sales revenue as compared to others and East region generated highest sales revenue as compared to others.

A screenshot of a calculator

Description automatically generated

**A graph with numbers and lines

Description automatically generated**

1. Which product category generates the highest sales revenue?

Ans. As we can see in the bar chart that Technology product category generates the highest sales revenue.

A screenshot of a calculator

Description automatically generated

A screenshot of a graph

Description automatically generated

1. How do sales vary by region or country?

Ans. Here sales vary by region not country because this data is only from one country and hence in that country the sales will vary by region as you can see in the line graph below.

A screenshot of a table

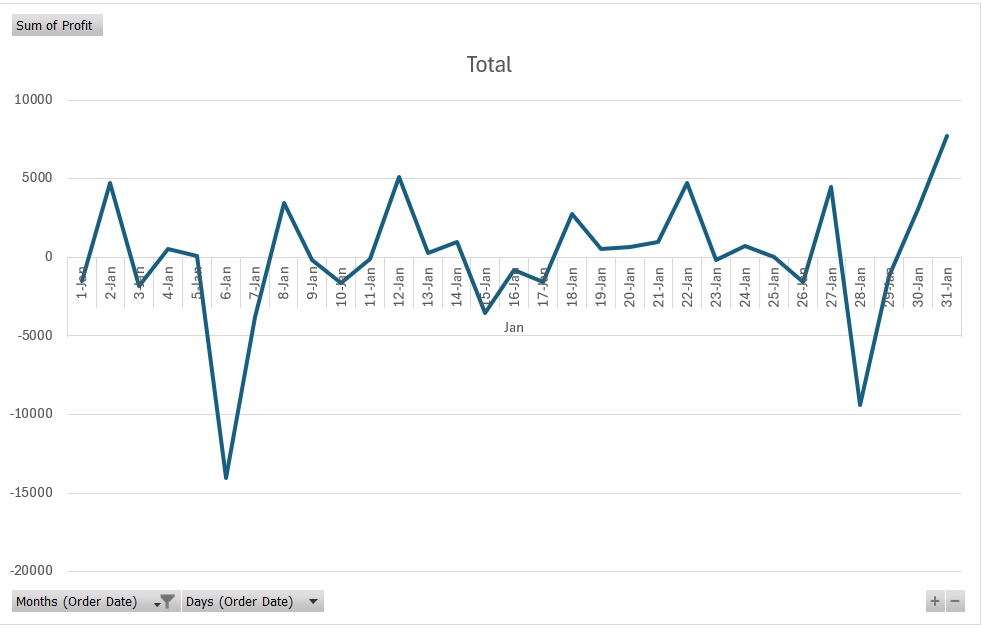
Description automatically generated

**A graph with numbers and lines

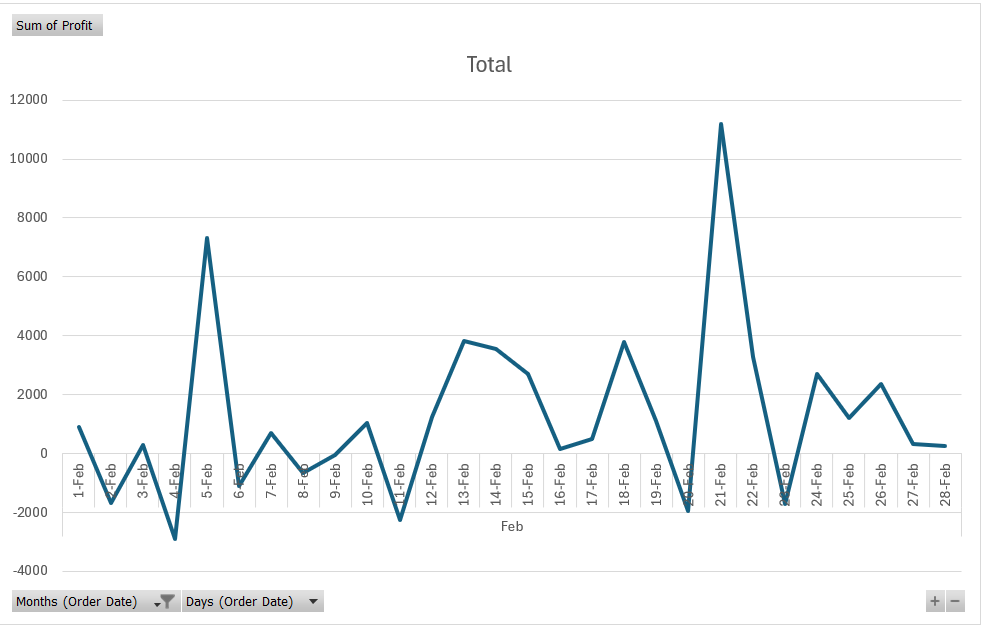
Description automatically generated**

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

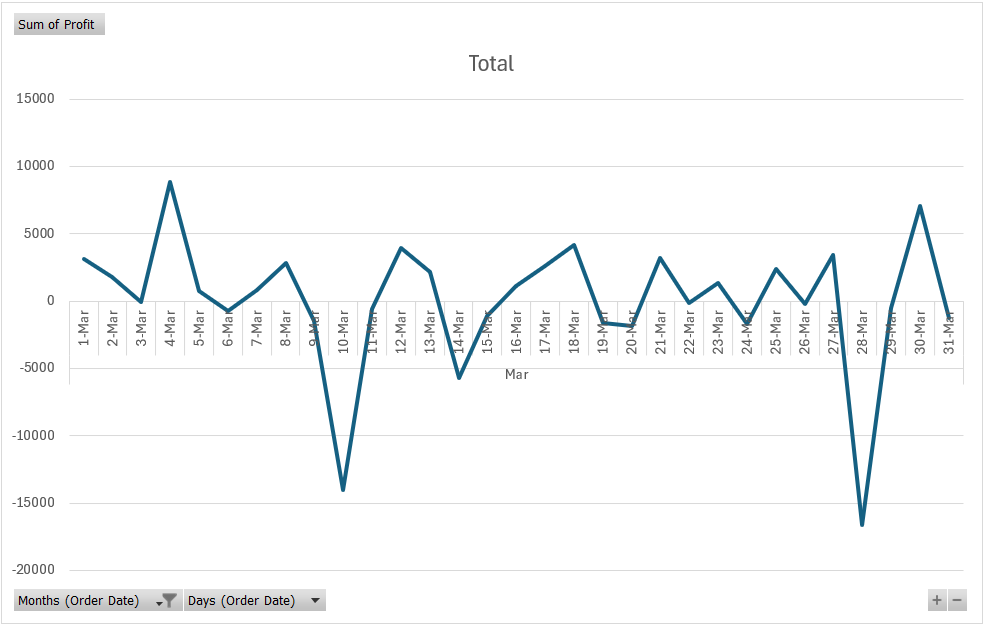
Ans. There are patterns in sales over month because we are not given any particular time. As you can see below there are six graphs according to months from January to June. These graphs shows sales of that particular month.



Sales in January



Sales in February



Sales in March

A graph with a line

Description automatically generated

Sales in April

A graph with blue lines and numbers

Description automatically generated

Sales in May

A graph with blue lines and numbers

Description automatically generated

Sales in June

1. **Profitability Analysis:**
   1. What is the overall profitability of the business?

Ans. The overall profitability of the business is very good, as you can see the maximum profit from this business is from east region and also there is loss from south region but if we see the overall profit from all the region then this business is generating very good profit.

A screenshot of a calculator

Description automatically generated

A graph of blue rectangular bars

Description automatically generated with medium confidence

* 1. Which products have the highest profit margins?

Ans. Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind.

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

Ans. As you can see in the bar graph that corporate segment in east region are more profitability than others where as in central region corporate segment is in loss with -25.78 average.

A graph of a bar chart

Description automatically generated with medium confidence

* 1. Is there a correlation between discounts offered and profitability?  
     Ans. The correlation between discounts offered and profitability is given below.

A black and white discount tag

Description automatically generated

1. **Customer Analysis:**
   1. Who are the top customers by total sales?

Ans. These are the top three customers by total sales.

A table with text on it

Description automatically generated

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

Ans. The average order size for different customer segments are given below. Small business orders the most as compared to other segments like home office, corporate, consumer, etc. The corporate segment where as orders less from all the customer segments.

A table with text and numbers

Description automatically generated with medium confidence

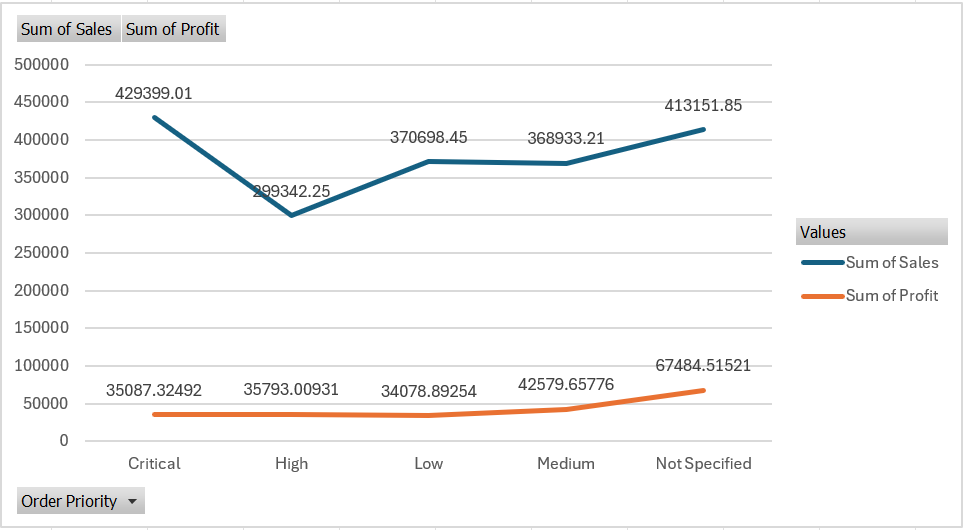
A screenshot of a graph

Description automatically generated

* 1. How does order priority impact sales or profitability?  
     Ans. As we can see from the graph and data that the orders with very critical priority generated less profit but those orders which are not specified generated the highest profit among them. As the priority increases the profit decreases, hence order priority is inversely proportional to profit which is shown below.

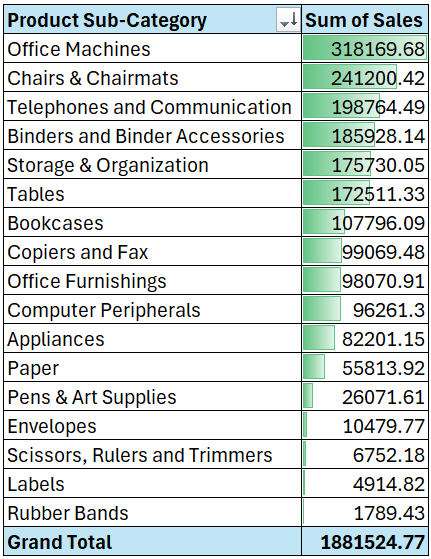
A screenshot of a spreadsheet

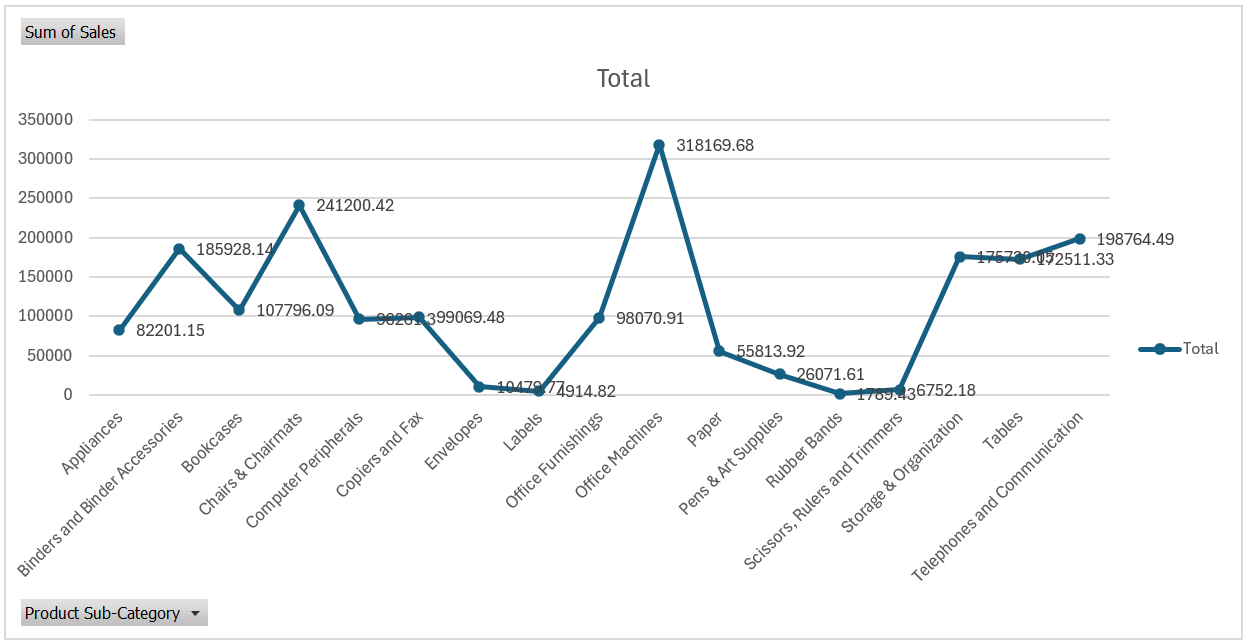
Description automatically generated



1. **Product Analysis:**
   1. Which product sub-categories are the best-selling?

Ans. From the line chart and the data we can see that office machines are the best selling in this 6 month data.



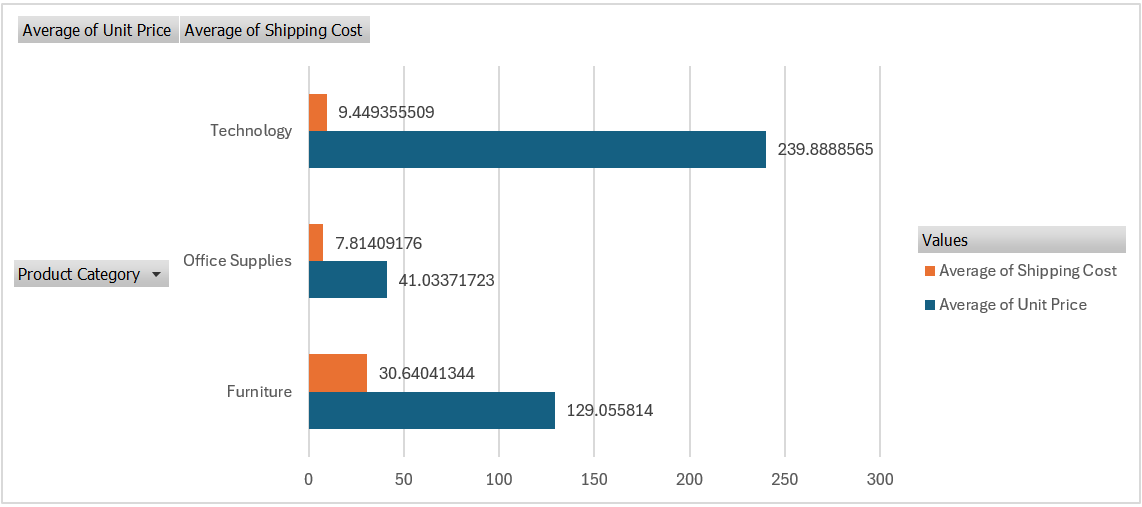


* 1. What is the average unit price and shipping cost for each product category?

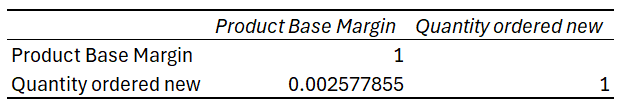
Ans. The average unit price and shipping cost for each product category is given below. The line chart shows that the average unit price for technologies is higher among them where as the average shipping cost for furniture is higher.

A blue and white rectangular sign with black text

Description automatically generated



* 1. Is there a relationship between product base margin and sales volume?  
     Ans. It was not possible to find relationship between numerical values so I find correlation between product base margin and sales volume.



1. **Shipping Analysis:**
   1. Which shipping mode is most commonly used?

Ans. The shipping mode which is most commonly used is Regular Air because its cheaper and faster thus Delivery truck will be costing much less but it will consume more time to get delivered than Regular Air.

A screenshot of a computer

Description automatically generated

A graph with blue bars

Description automatically generated

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

Ans. Yes, choice of shipping mode affects profit and customer satisfaction. The better the shipping mode is the better the satisfaction of customer will be.

A blue and white rectangular object with black text

Description automatically generated

A graph with numbers and a bar

Description automatically generated

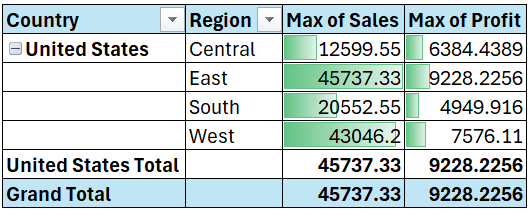
* 1. Are there any trends in shipping costs over time or across regions?  
     Ans. Yes shipping cost vary across region. As you can see in the line chart below it shows shipping cost in every region with respect to months and we can see that in central region the shipping costs were highest and in the south region the shipping costs were lowest.

A graph showing the number of months and months

Description automatically generated with medium confidence

1. **Geographical Analysis:**
   1. Which regions or countries contribute the most to sales or profitability?

Ans. The east region contributes the most to sales and profitability. As we can see that this data is of 1 country i.e. united states so we have to look in the regions as given below.



A graph of different colored bars

Description automatically generated with medium confidence

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

Ans. There is no geographical location given in the data so I made it with order frequency and sales.

A white rectangular sign with black text

Description automatically generated

1. **Order Analysis:**
   1. What is the average order value?

Ans. The average order value according to product category are given below where you can see that the highest average order is of Office Supplies.

A blue and white rectangular object with black text

Description automatically generated

A graph of a bar chart

Description automatically generated with medium confidence

* 1. How does order volume vary by month or season?

Ans. The order volume vary by month as you can see in the chart below. Every month you can see the order frequency i.e. order count. The maximum order count was in January(2015-01) and the minimum order count was in June(2015-06). Yes, we can say that order volume vary by month.

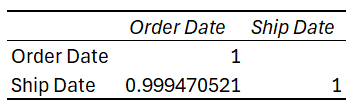
A screenshot of a table

Description automatically generated

A graph with a line and numbers

Description automatically generated

* 1. Are there any correlations between order date and shipping date?  
     Ans. The correlations between order date and shipping date is given below.



1. **Managerial Insights:**
   1. Based on the analysis, what recommendations can be made to improve profitability?

Ans. The shipping duration can be improved, focusing on those region where sales are going down so the profit can be increased.

* 1. Are there any operational inefficiencies that can be addressed?

Ans. Shipping could be better because for some products the shipping date is more than the order date i.e. the duration of shipping can be decreased that will help in the growth of the company.

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

Ans. The business better tailor its offerings to meet customer needs by adding more supplies other than these product categories and decrease the duration of shipping items.

**Pivot Table**

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

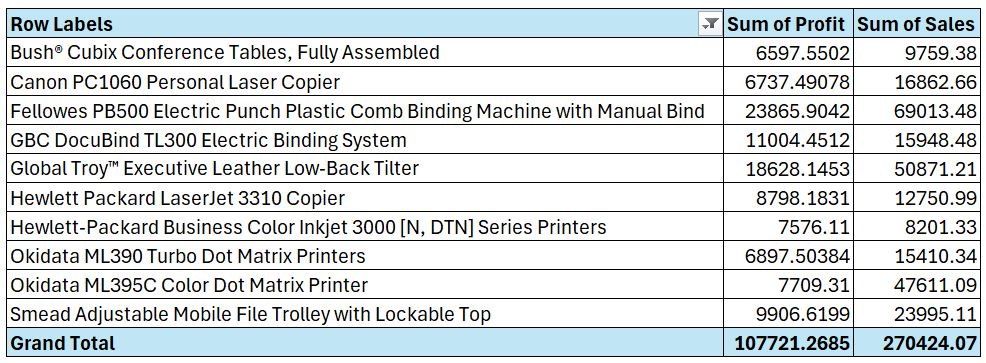
Ans. These are the top 10 cities in terms of profit.

A table with numbers and a few black text

Description automatically generated with medium confidence

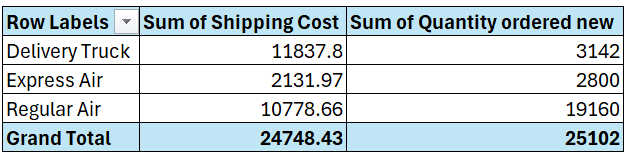
1. What are the most revenue-generating products in terms of profit?

Ans. These are the top 10 most revenue generating products in terms of profit.



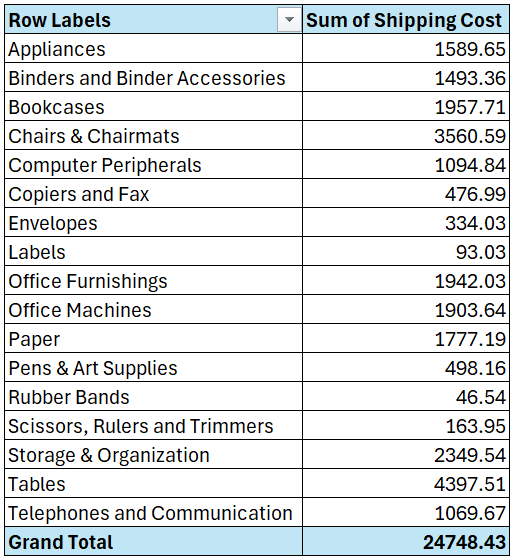
1. How does shipping expense vary for different transportation methods?

Ans. As we can see that most product have been ordered from regular air and it also cost less than any other ship mode. The delivery truck takes lot of time to deliver that is why most order are from regular air its faster than any other.



1. How do shipping charges vary for different products?

Ans. The shipping charges vary for different products are given below. As we can see that the ‘Rubber Bands’ has lowest shipping cost and ‘Tables’ has the highest shipping cost.



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

Ans. Below is the percentage of order based on priority.

A table with numbers and text

Description automatically generated

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

Ans. The performance of managers from Jan-Jun are given below.

A screenshot of a data table

Description automatically generated

A graph of different colored bars

Description automatically generated

1. How do sales vary across different regions?

Ans. Sales vary across different regions are given below. South region has lowest sales and East region has highest sales.

A table with numbers and text

Description automatically generated

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

Ans. The Sales from January to June are given below.

A table with numbers and a row of labels

Description automatically generated

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

Ans. Monthly profit throughout the first half of the year are given below.

A screenshot of a calculator

Description automatically generated

1. How can customers be analyzed based on segments?

Ans. Customers can be analyzed on the basis of customer segment and total sales are given below.

A screenshot of a computer screen

Description automatically generated

**Pivot Charts**

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

Ans. These are the top 10 products that generated the most revenue in terms of profit.

A graph of a product

Description automatically generated with medium confidence

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

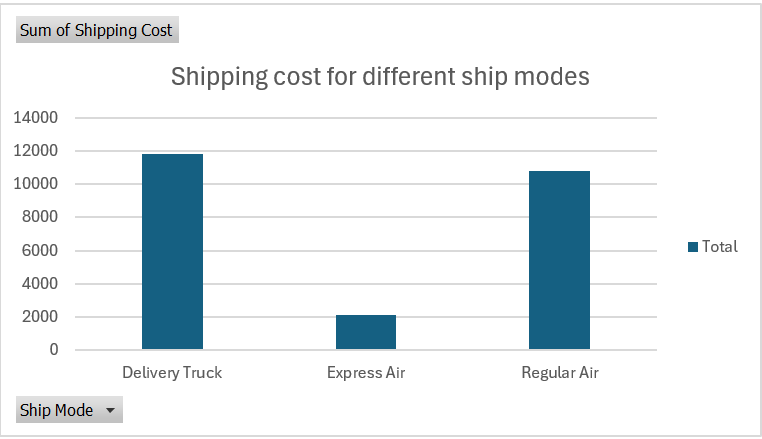
Ans. These are the top 10 most revenue generating products in terms of sales.

A graph of sales

Description automatically generated

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

Ans. Below are the shipping cost for different ship modes mentioned in the dataset.



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

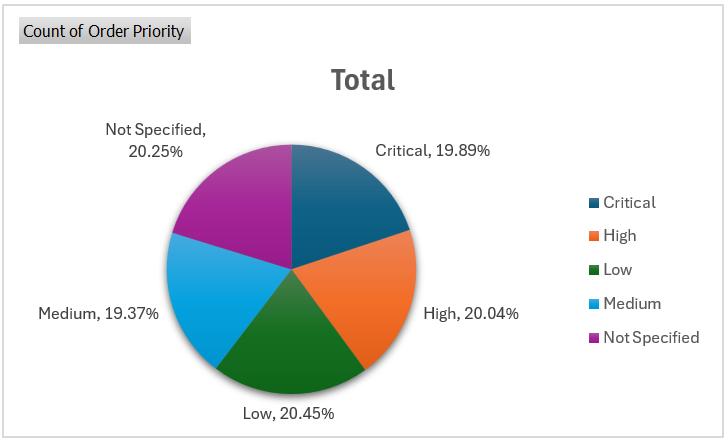
Ans. These are the shipping cost for different product sub-category.

A graph with blue and black lines

Description automatically generated with medium confidence

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

Ans.



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

Ans.

A graph with blue lines

Description automatically generated

Sales in January

A graph with blue lines and numbers

Description automatically generated

Sales in February

A graph with blue lines

Description automatically generated

Sales in March

A graph showing sales

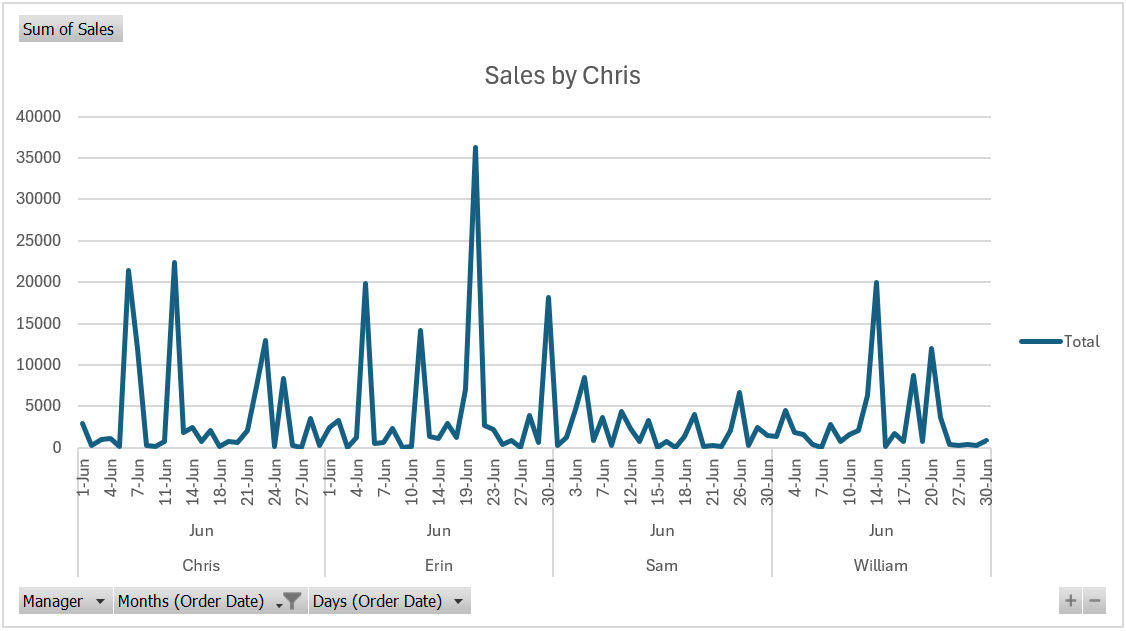
Description automatically generated

Sales in April

A graph with blue lines and numbers

Description automatically generated

Sales in May



Sales in June

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

Ans. Below is the display of the dashboard to see regional sales.

A screenshot of a phone

Description automatically generated A graph with a line and numbers

Description automatically generated

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

Ans. Below is the line chart represents the sales trend from January to June.

A graph with a line

Description automatically generated

Sales in January

A graph with blue lines and numbers

Description automatically generated

Sales in February

A graph with blue lines and numbers

Description automatically generated

Sales in March

A graph with a line

Description automatically generated

Sales in April

A graph with blue lines and numbers

Description automatically generated

Sales in May

A graph with blue lines and numbers

Description automatically generated

Sales in June

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

Ans. Below is the dashboard visualization of monthly profit variations.

A screenshot of a calendar

Description automatically generated A graph of a graph

Description automatically generated with medium confidence

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

Ans.

A graph with a line

Description automatically generated

**Dashboard**

How can we design 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?



**SUMMARY**

In this report I have discussed about analysis on excel on a dataset, its working, graphs, dashboard, etc. I tried to explain almost every steps of how to analyze a dataset using Microsoft Excel from a given dataset.

This report highlights all the process involved in analyzing data using Excel features like pivot tables, pivot charts, dashboards, filtering, sorting, formulas etc. I have attached previews of charts, pivot charts, dashboards and all objective in this report.

**REFERENCES**

Datasheet –

<https://docs.google.com/spreadsheets/d/1B7w444UZ1eWQtnE3Z5Cmozipr3RkOAhe/edit?usp=drive_link&ouid=109618786931407309161&rtpof=true&sd=true>