**INTRODUCTION TO DATA SCIENCE**

**AN EXCEL PROJECT – SALES DATA ANALYSIS**

SUBMITTED BY

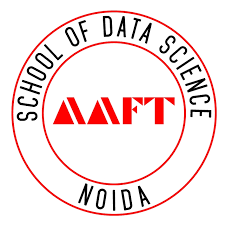
**GUNGUN**

**DATA SCIENCE**

UNDER THE GUIDANCE OF

AISHWARY SHUKLA

AAFT



ACKNOWLEDGEMENT

I express my sincere gratitude to my teacher, Mr. Aishwary Shukla for his invaluable guidance and support throughout the project. His enthusiasm and expertise in the field of Data Science greatly inspired and motivated me to explore and learn new concepts.

I would also like to thank my parents for their unwavering support and encouragement throughout the Course. Their constant encouragement and belief in my abilities allowed me to push myself to the limit and produce a successful project.

I am also grateful to my classmates for their valuable suggestions during the brainstorming and research phases of the project. Their insights and perspectives helped me better understand the topic and improve my project.

Lastly, I would like to acknowledge the various sources that provided valuable information and data for my project. These include excel and different website. I consulted during my research. Their contribution is greatly appreciated.

Gungun

Course-Data Science

AAFT University, Noida

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Topic** | **Page.No** |
| **1** | **Cover page** | **1** |
| **2** | **Acknowledgement** | **2** |
| **3** | **Contents** | **3** |
| **4** | **Abstract** | **4** |
| **5** | **Introduction** | **5** |
| **6** | **Data Analysis** | **6** |
| **7** | **Data visualization** | **7-14** |
| **8** | **Sales analysis** | **15-18** |
| **9** | **Profitability analysis** | **18-20** |
| **10** | **Customer analysis** | **20-21** |
| **11** | **Product analysis** | **21-22** |
| **12** | **Shipping analysis** | **22-27** |
| **13** | **Geographical analysis** | **28** |
| **14** | **Order analysis** | **28-29** |
| **15** | **Managerial insights** | **29-30** |
| **16** | **Pivot table** | **30-33** |
| **17** | **Pivot charts** | **33-37** |
| **18** | **Dashboard** | **38** |
| **19** | **Summary of report**  **And References** | **39-40** |

**ABSTRACT**

Excel is a software program created by Microsoft that uses spreadsheets to organize numbers and data with formula 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 a Sales data set. This report also presents my learning during my course classes.

**INTRODUCTION**

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

The data set used contains information regarding annual year Sales. It includes information regarding all the Data involved in the Sales, the data of Sales like 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 magic, Country, Region, State or province, City, Postal Code, Order date, Profit, Quantity ordered new, Sales, Order id, Total, Manager.

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.

There are some Key points:

**Analysis:** A detailed examination of anything complex in order to understand its nature or to determine its essential features : thorough study. doing a careful analysis of the problem.

**Visualization:** Visualization occurs while reading when readers are able to create mental images that represent what is happening in the text.

**Sale:** A sale is a transaction between two or more parties in which goods or services are exchanged for money or other assets.

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

**Customer:** A person who buys a good, product, or service.

**Product:** A product is the item offered for sale.

**Shipping:** Shipping is the physical movement of goods from one point to another.

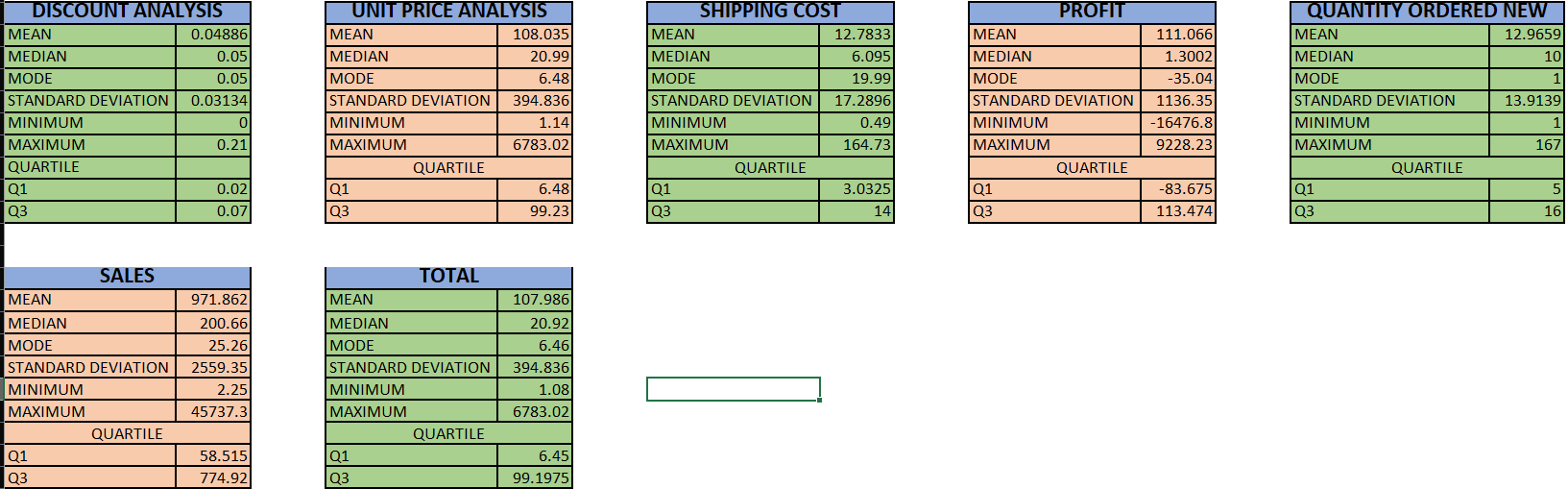
**Geographical:** Relating to the natural features, population, industries, etc., of a region or regions.

**Box Plot:** Box plots are used to show distributions of numeric data values, especially when you want to compare them between multiple groups.

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,

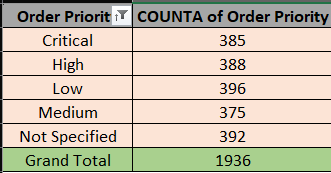
****

DATA VISUALIZATION

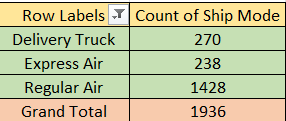
Create various charts and graphs to visualize different aspects of the data. For example:

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

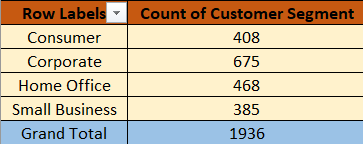
Bar chart of categorical variable of Order priority from the Data.



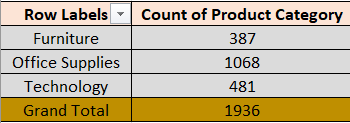
Bar chart of categorical variable of Ship mode from the Data.



Bar chart of categorical variable of Customer segment from the Data.

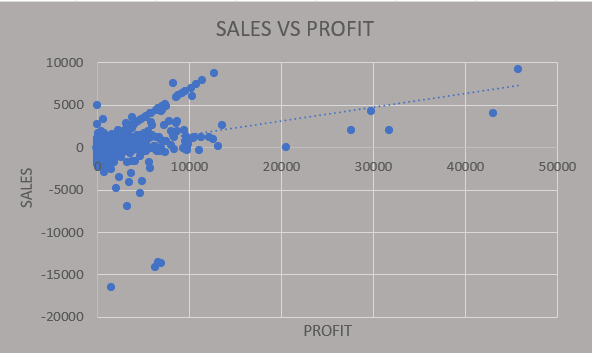


Bar chart of categorical variable of Product category from the Data.





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



There are two scatter plot of different data.

1. Sales VS Profit
2. Unit price VS Total
3. Discount VS Shipping cost

They show the relationship between two variables in the data -set.

**3. Box plots to identify outliers in numerical data.**

**KEY POINTS:**

**Q1=FIRST QUARTILE**

**Q2= SECOND QUARTILE AND MEDIAN**

**Q3=THIRD QUARTILE**

**MIN=MINIMUM VALUE**

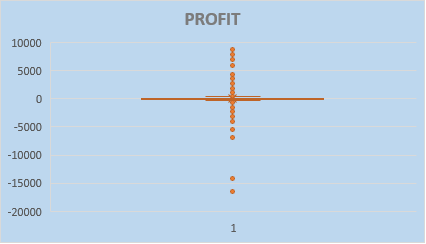
**MAX=MAXIMUM VALUE**

**IQR=INTERQUARTILE RANGE**

**RANGE=MAXIMUM VALUE – MINIMUM VALUE**

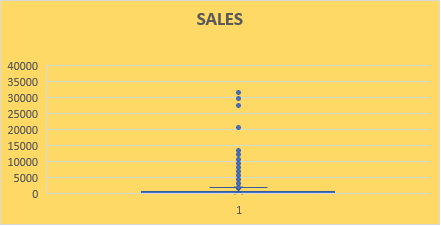
**IQR=Q3-Q1**

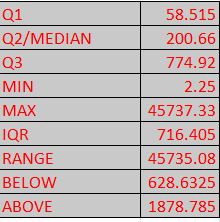
**Box plot of profit to identify the outliers.**



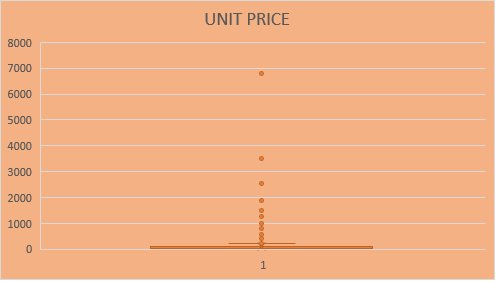


**Box plot of Sales to identify the outliers.**



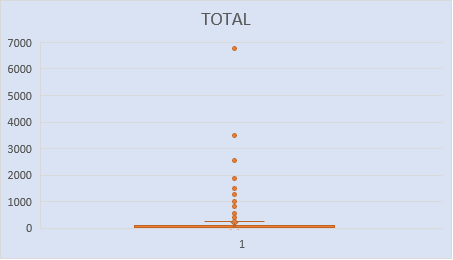


**Box plot of Unit price to identify the outliers.**





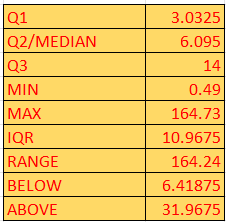
**Box plot of Total to identify the outliers.**





**Box plot of Shipping cost to identify the outliers.**





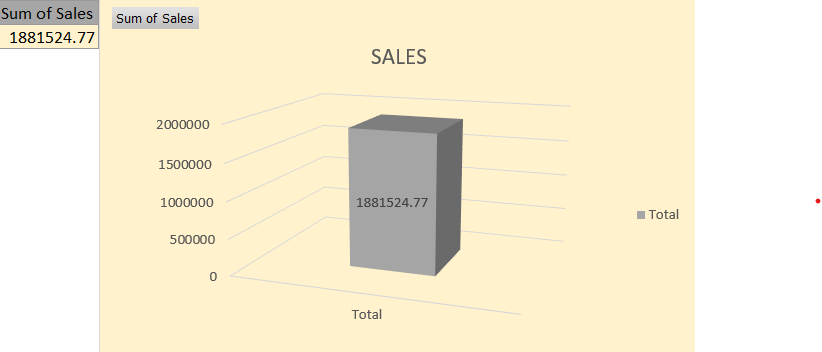
**Box plot of Discount to identify the outliers.**





**Sales Analysis:**

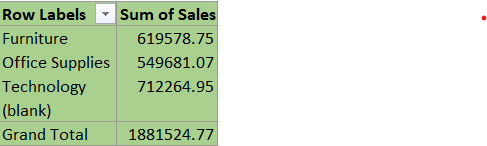
* **Conclusions from the analysis: (compare result with scatter plot)**
  1. **What is the total sales revenue for the dataset?**

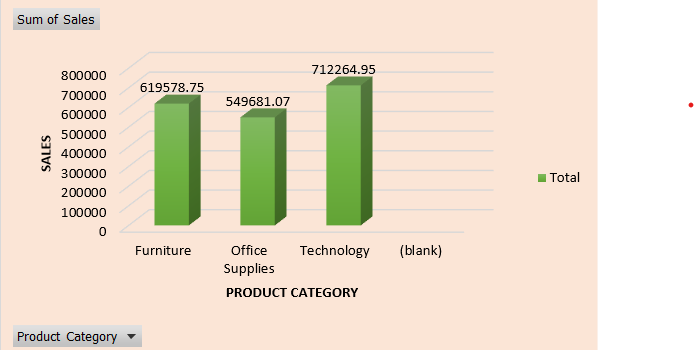
****

CONCLUSION:

The total sales revenue of the dataset is 1881542.77.

**B. Which product category generates the highest sales revenue?**

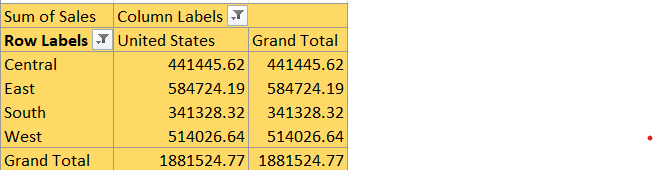


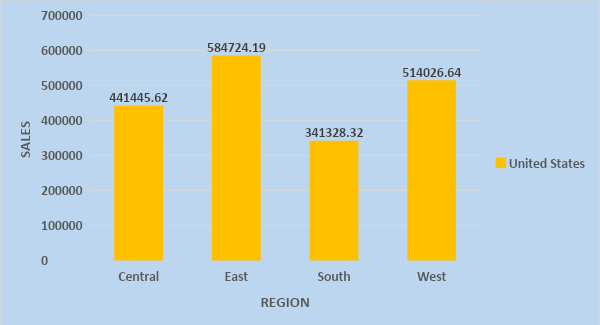


**CONCLUSION:**

As we can see in the above graph. It shows that the technology sale is more than other product category sales.

**C. How do sales vary by region or country?**

****



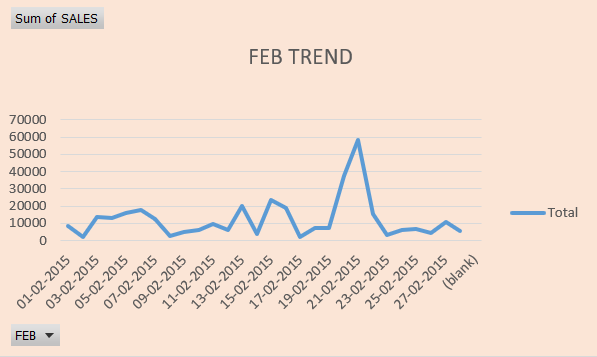
D. Are there any trends or patterns in sales over time?

A general direction of change : a way of behaving, proceeding, etc., that is developing and becoming more common. the downward/upward trend of the stock market. Digital technology is the latest/current trend in television.

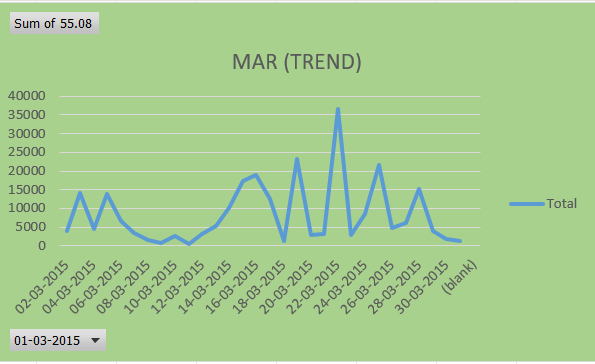
**CONCLUSION:**A trend describes the general direction of change of some variable. The diagram shows the change in the sales of Jan. The trend shows that it is not constant. A decreasing trend means the sales are decreasing and vice versa.

In this graph the highest Sale on 10TH TO 12TH jan.

FEB DATA LINE CHART

 CONCLUSION:

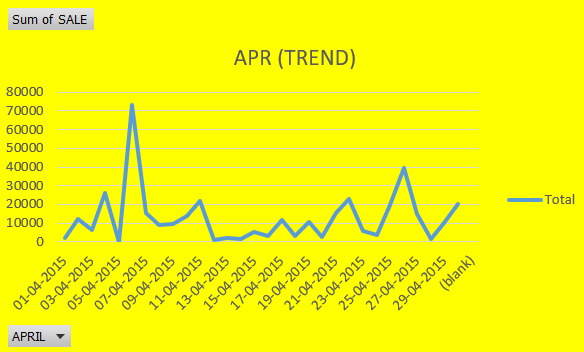
In above graph highest sale on 21th to 23th feb.

MAR DATA LINE CHART 

CONCLUSION:

As per the march data highest sale on 21th to 22th mar.

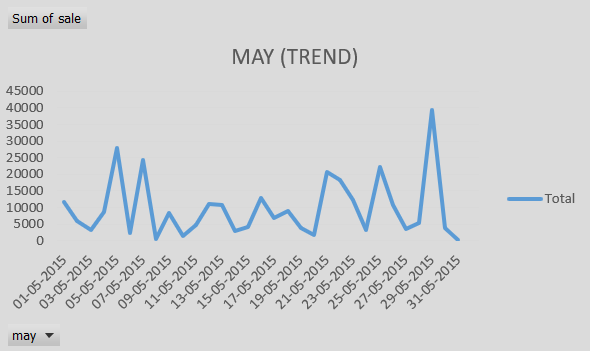
APR DATA LINE CHART



CONCLUSION:

In Apr sales data highest sales on 5th to 8th Apr.

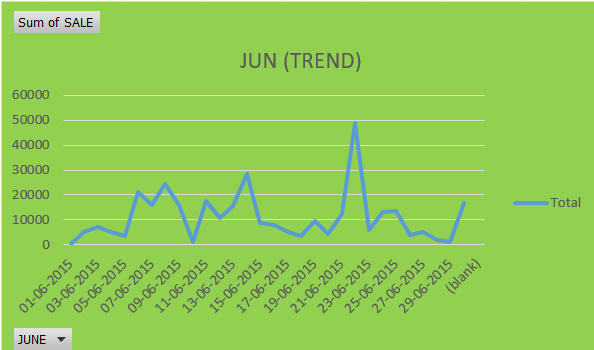
MAY DATA LINE CHART



CONCLUSION:

As per the diagram, The highest sale was on 28th to 30th May.

JUNE DATA LINE CHART



CONCLUSION:

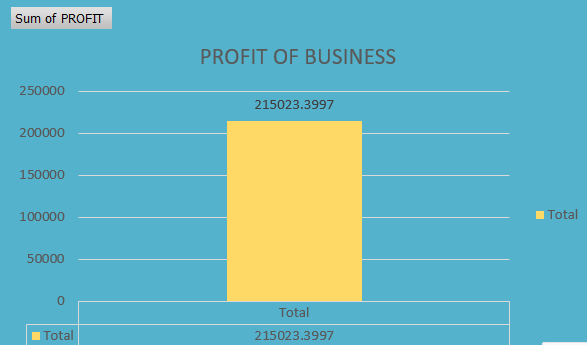
In this graph the highest sale was on 20th to 22th June.

**Profitability Analysis:**

A:- What is the overall profitability of the business?

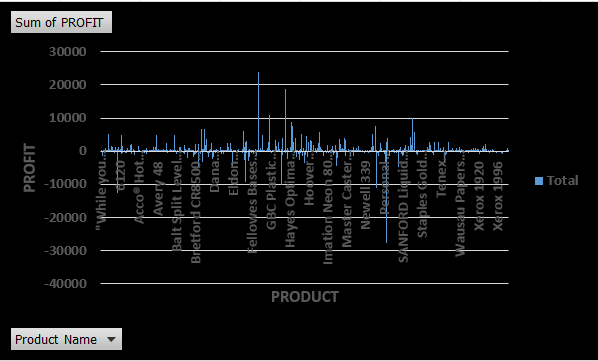
**Ans :-** The overall profitability of the business is 215023.3997.



****

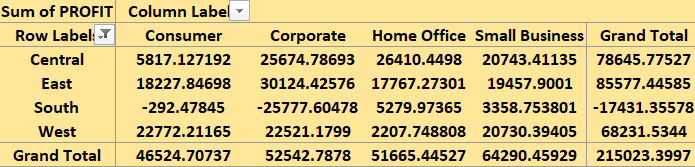
**B:-** Which products have the highest profit margins?

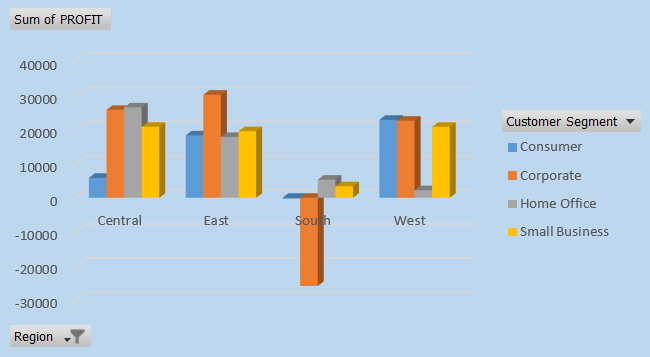
Ans:- Fellowes mobile numeric keypad ,Graphite(product name) product have the highest profit margins.



C:-Are there any customer segments or regions that are more profitable than others?

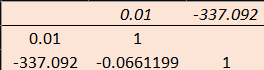
Ans:- East and corporate has more profit than others.





D:-Is there a correlation between discounts offered and profitability?

Ans:- Correlation between discounts offered and profitability is -0.0614



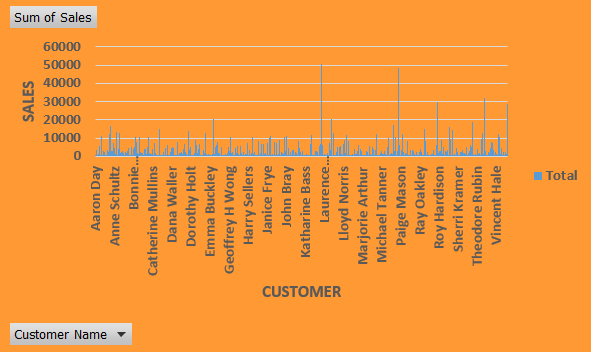
The formula of correlation is =CORREL(array1,array2)

I have taken value on this formula =correl(discount,profit)

**Customer Analysis:**

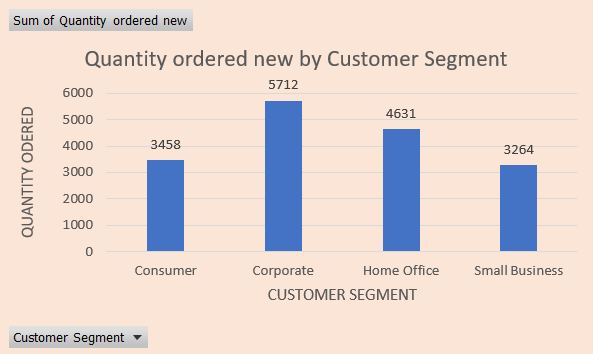
A:-Who are the top customers by total sales?

Ans:- Kristine Connolly is the top customer by total sales.



B:-What is the average order size for different customer segments?

Ans:-



C:-How does order priority impact sales or profitability?

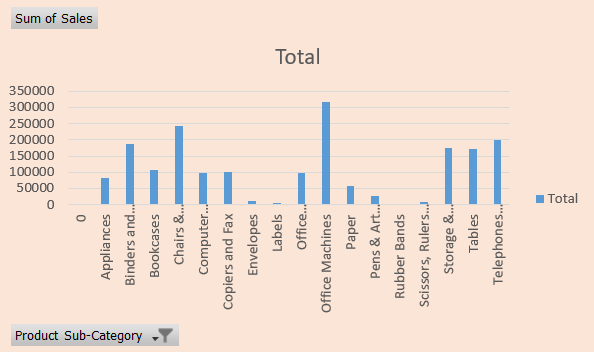
Ans:-



**Product Analysis**

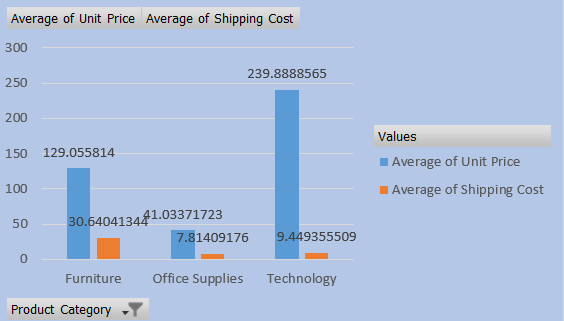
A:-Which product sub-categories are the best-selling?

Ans:-Office machines of sub-categories are the best-selling.



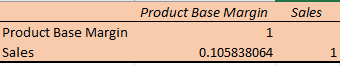
B:-What is the average unit price and shipping cost for each product category?

**Ans:-**



C:-Is there a relationship between product base margin and sales volume?

Ans:-

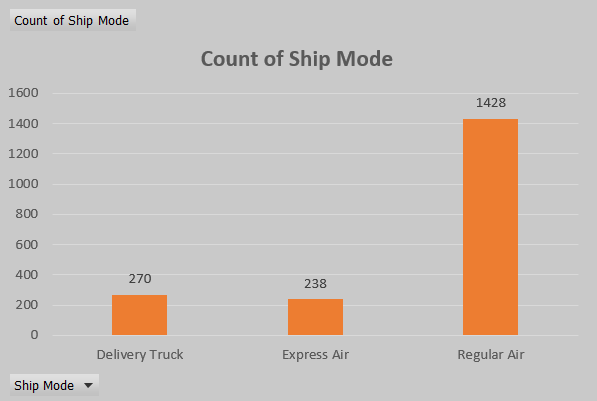


I have taken in formula of correlation =correl(product base margin,sales)

**Shipping Analysis:**

A:-Which shipping mode is most commonly used?

Ans:-

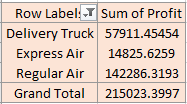


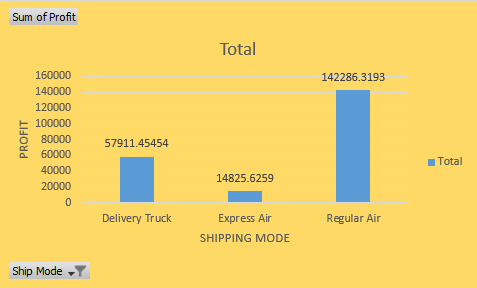
CONCLUSION:-

Regular air is the most commonly used in shipping mode.

B:-Does the choice of shipping mode affect profitability or customer satisfaction?

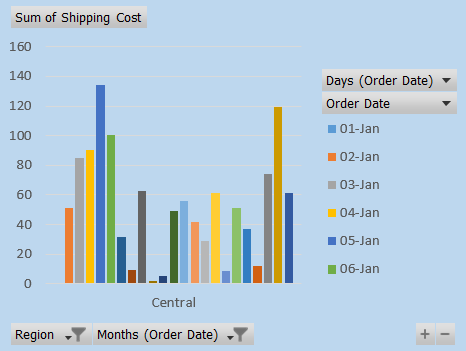
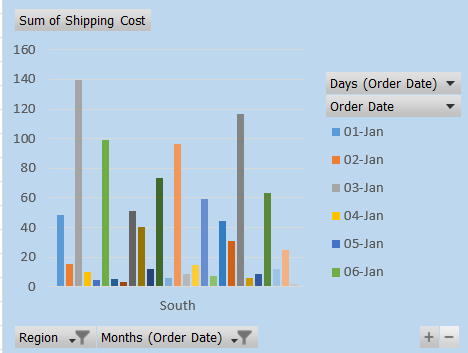
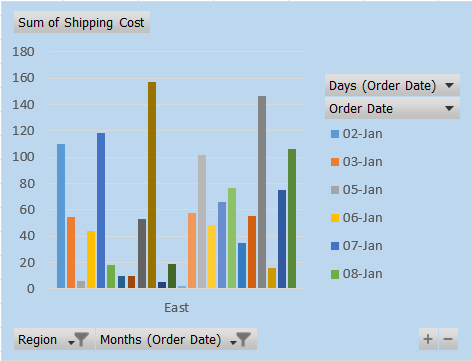
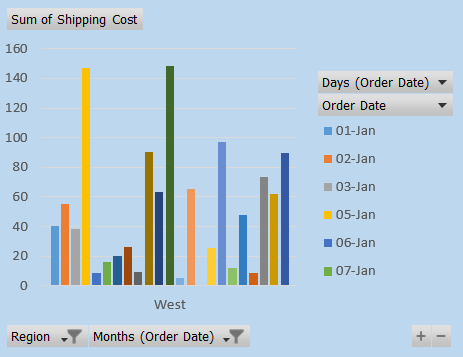
Ans:-





C:-Are there any trends in shipping costs over time or across regions?  
Ans:- yes, there are trends in shipping costs over time or across regions.

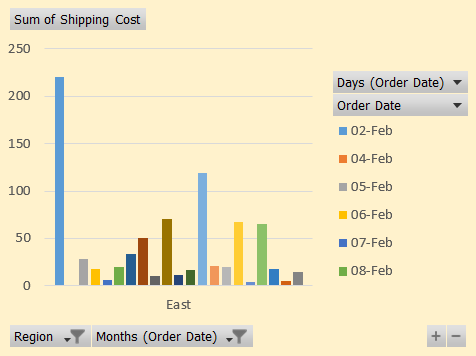
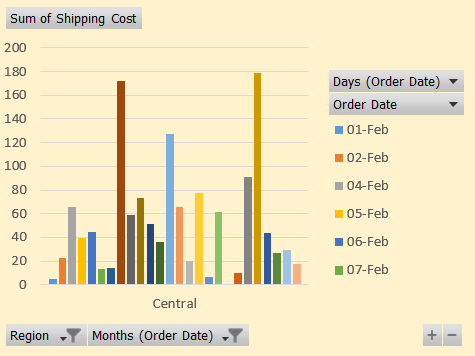
**JAN DATA**

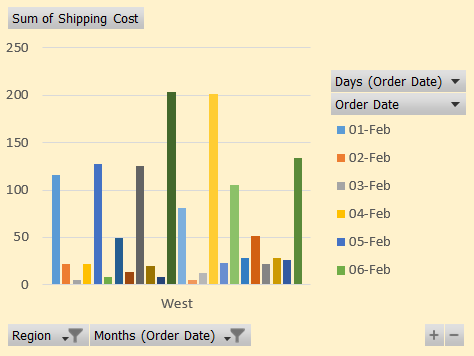
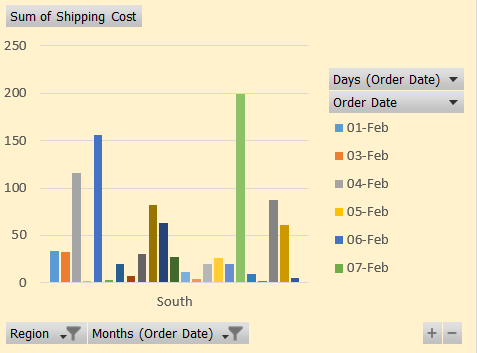
**  **

**CONCLUSION:**

**West has the highest shipping cost from all the data of jan.**

**FEB DATA**

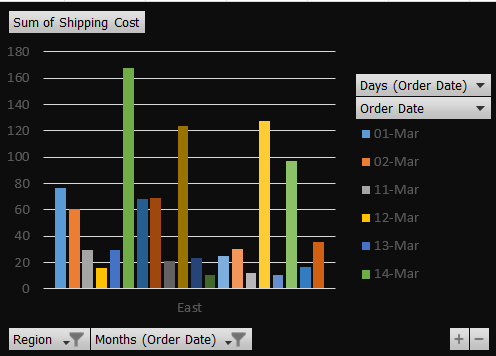
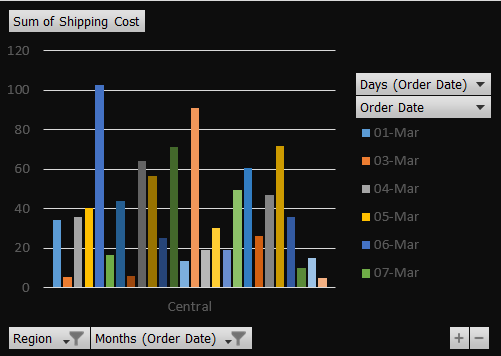
** **

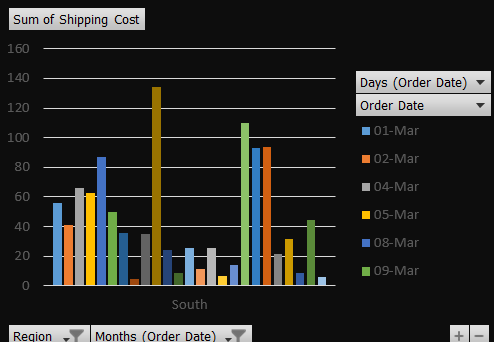
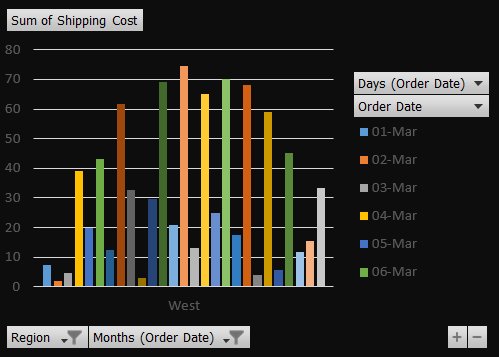
** **

**CONCLUSION:-**

**In the above graphs East has the highest shipping cost.**

**MAR DATA**

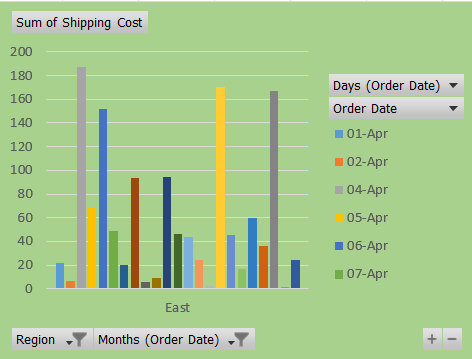
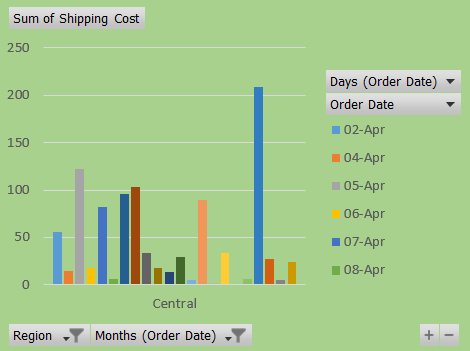
** **

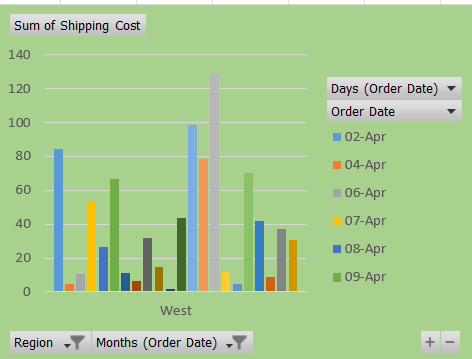
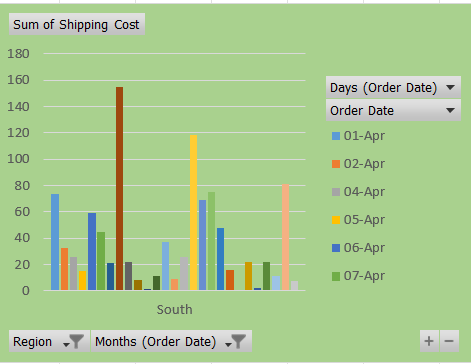
** **

**CONCLUSION:**

**As we can see from the above graphs it shows that the east has the highest shipping cost.**

**APR DATA**

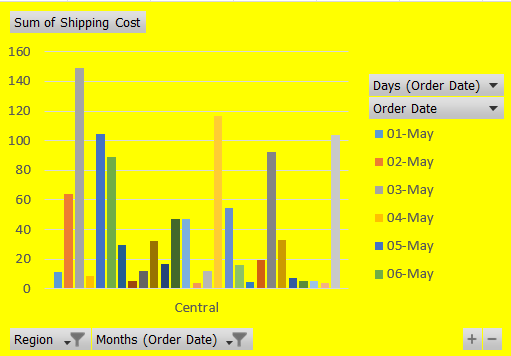
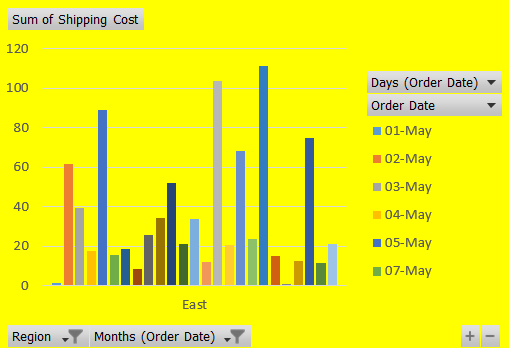
****

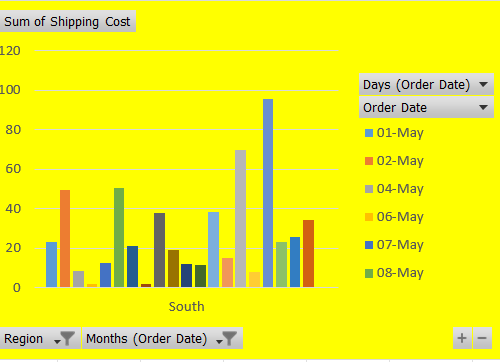
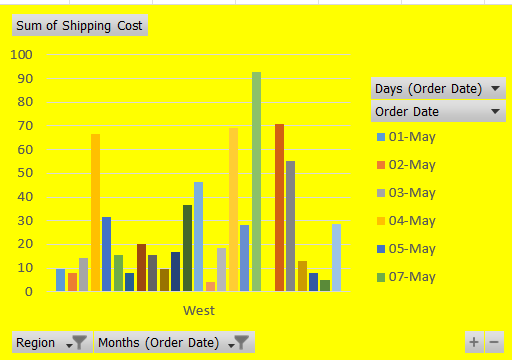
****

**CONCLUSION:**

**In this above graph central has the highest shipping cost.**

**MAY DATA**

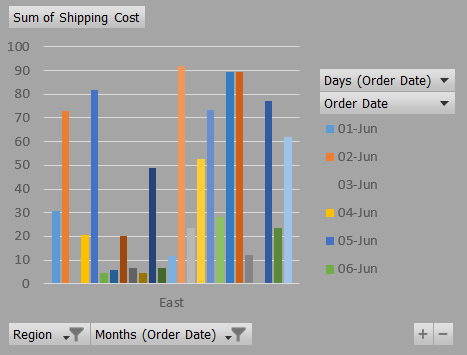
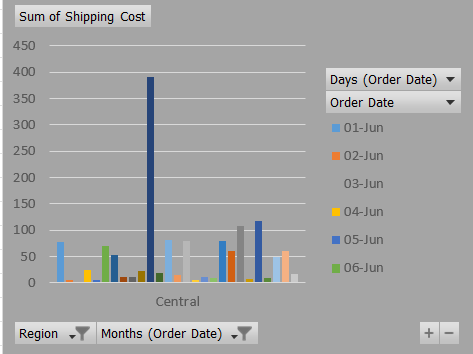
** **

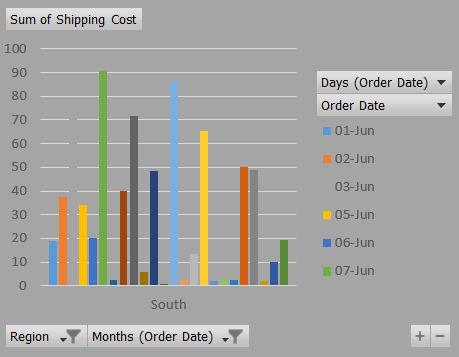
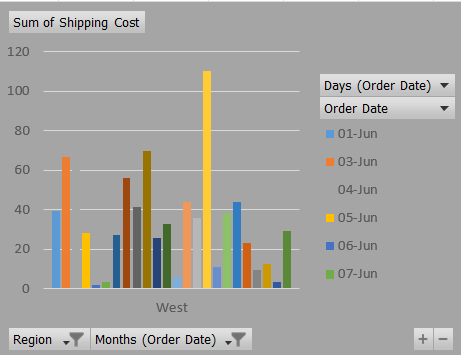
** **

**CONCLUSION:**

**The highest shipping cost in the central region.**

**JUN DATA**

** **

** **

**CONCLUSION:**

**It shows central has highest shipping cost.**

**OVERVIEW.**

**As you can see above graphs this data is the jan To jun month of shipping cost and as well as all the region:**

**There are four region represent**

**1.East**

**2.Central**

**3.South**

**4.West**

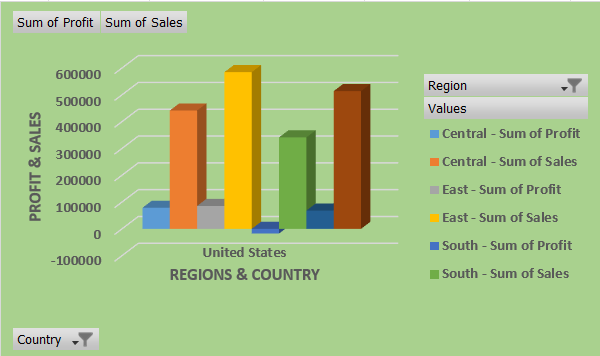
**South does not highest shipping cost in comparision to others.**

**Geographical Analysis:**

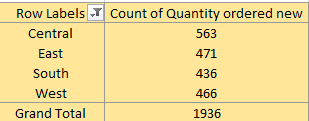
A:-Which regions or countries contribute the most to sales or profitability?

Ans:- East contribute the highest sales and profit also in comparision to other regions.

In this data only one country has given that is the united states.



B:-Is there a correlation between geographic location and order frequency?

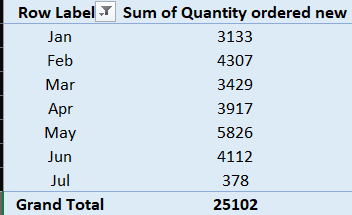


**Order Analysis:**

A:-What is the average order value?

Ans:- The average of quantity ordered new is 12.96591

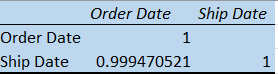
B:-How does order volume vary by month or season?



the main dataset season is not given. that’s why I have taken a quantity ordered new and months.

C:-Are there any correlations between order date and shipping date?

Ans:- The correlation between order date and shipping date is 0.999471



The formula of correlation :- =correl(array1,array2)

I got the correlation from this formula= correl(order date, ship date)

**Managerial Insights:**

A:-Based on the analysis, what recommendations can be made to improve profitability?

Ans:-In my point of view if Sales increases at a suitable price so the profit automatically improves.

We have to take the step for customer satisfaction if the customer is fully satisfy.make some pricing strategy so they will buy more products.

**B:-** Are there any operational inefficiencies that can be addressed?

Ans:-yes, there are some operational inefficiencies that can be addressed.

1. **COMMUNICATION:-** poor communication channels or lack or clarity in communication can lead to delays, misunderstandings, and mistakes.
2. **POOR RESOURCE ALLOCATION:-** Misallocation of resources, whether it be human, financial, or material resources, can lead to inefficiencies.
3. **WORKFLOW BOTTLENECKS:-** identify points in the workflow where tasks tend to pile up or slow down.

C:-How can the business better tailor its offerings to meet customer needs?

Ans:-There are several strategies to help a business better tailor its offerings:

* + **MARKET RESEARCH**:- Conduct comprehensive market research to understand customer demographics, preferences, and purchasing behaviors. Use surveys, focus groups, and data analytics to gather insights into what customers want and need.
  + **CUSTOMER SEGMENTATION**:-Segment customers based on shared characteristics such as demographics, purchasing behavior, and preferences. Tailor offerings to address the specific needs and preferences of each segment.
  + **FEEDBACK COLLECTION**:- Actively solicit feedback from customers through surveys, reviews, and direct communication channels. Use this feedback to identify areas for improvement and make adjustments to products, services, and customer experiences accordingly.
  + **CUSTOMIZATION OPTIONS**:-Offer customization options that allow customers to tailor products or service to their specific needs and preferences. Provide flexibility in features, pricing, and packaging to accommodate varying customer requirements.
  + **CONTINUOUS IMPROVEMENT**:- Commit to continuous improvement by regularly evaluating and refining offerings based on customer feedback, market dynamics, and competitive pressures. Stay agile and adaptable to evoluing customer preferences and market conditions.

**Pivot Table:**

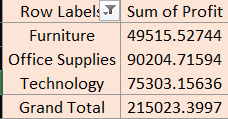
1.Which are the top 10 cities in terms of profit?

Ans:-There are top 10 cities in terms of profit.

VINCENNES , THORNTON, APPLETON, BANJOR, CLARKVILLE , HACKENSACK, LOUISVILLE, BALDWIN SPRINGFIELD , WINTER HAVEN.

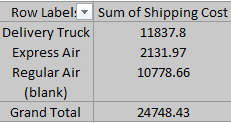
2.What are the most revenue-generating products in terms of profit?

Ans:- Office suppliers are the most revenue-generating products in terms of profit.



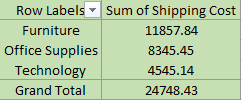
3.How does shipping expense vary for different transportation methods?

Ans:-



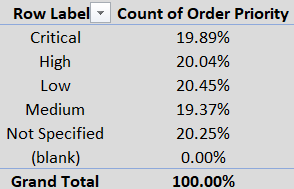
4.How do shipping charges vary for different products?

Ans:-



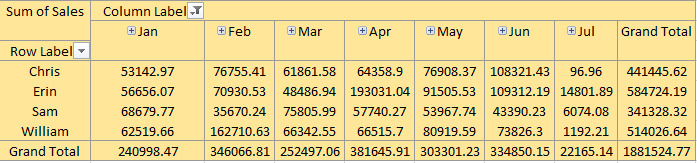
5.What is the percentage of orders based on priority?

Ans:-



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

Ans:-



7.How do sales vary across different regions?

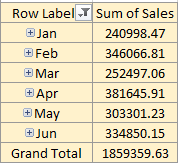
Ans:- There are 4 regions available in data and they have different – different sales.

|  |  |
| --- | --- |
| Row Labels | Sum of Sales |
| Central | 441445.62 |
| East | 584724.19 |
| South | 341328.32 |
| West | 514026.64 |
| Grand Total | 1881524.77 |

Highest sales are in the East region. then the West has the second highest sale and third sale in the central and then last sale in the South.

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?

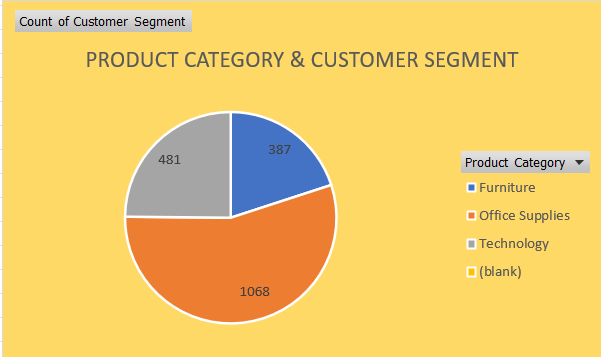
Ans:-

|  |  |
| --- | --- |
| **Row Labels** | **Sum of Profit** |
| **Jan** | **-9818.425395** |
| **Feb** | **45600.67942** |
| **Mar** | **-2766.895053** |
| **Apr** | **50780.85316** |
| **May** | **62756.3697** |
| **Jun** | **58517.74201** |
| **Grand Total** | **205070.3238** |

MAY has the highest monthly profit throughout the first half of the year.

10:-How can customers be analyzed based on segments?

Ans:-



**Pivot Charts:**

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

Ans:-



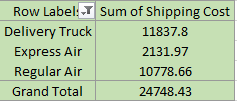
2.What visualizations can be used to display the most revenue-generating products?

Ans:-



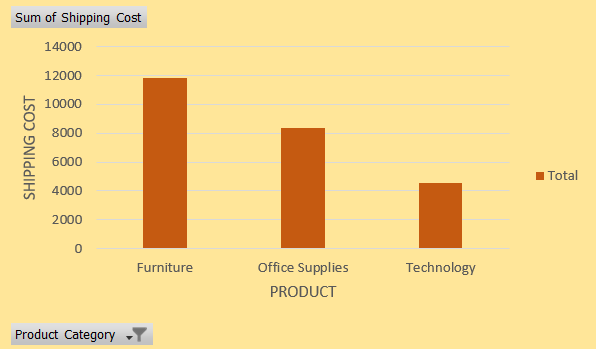
3. How can we visualize the shipping expense for different transport methods?

**Ans:-**

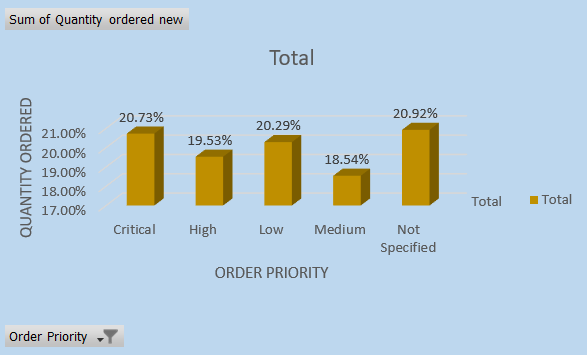
****

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

Ans:-

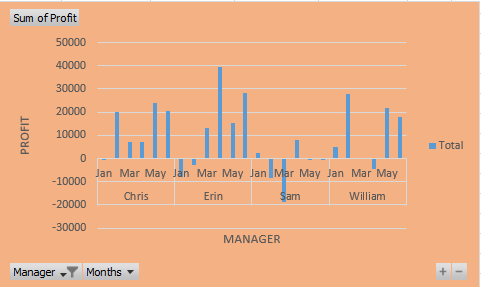


5.How can we visually display the percentage of orders based on priority?



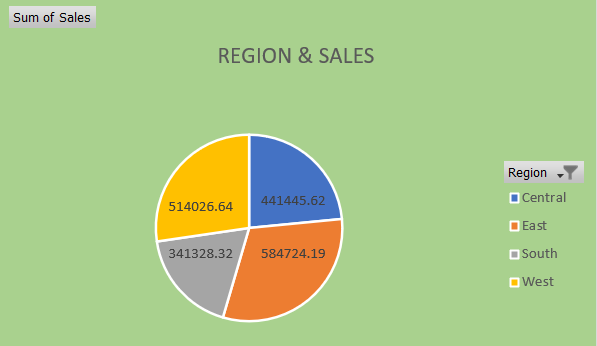
6.What type of chart or graph can effectively showcase the performance of managers for each month?

Ans:-



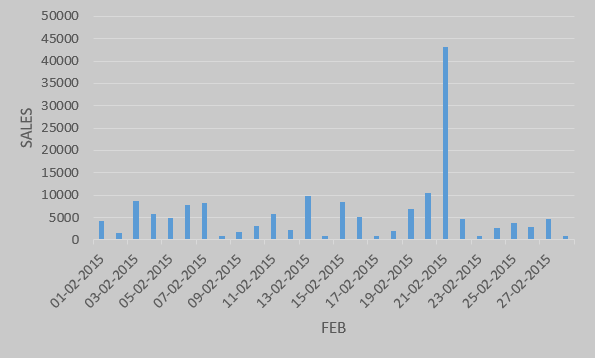
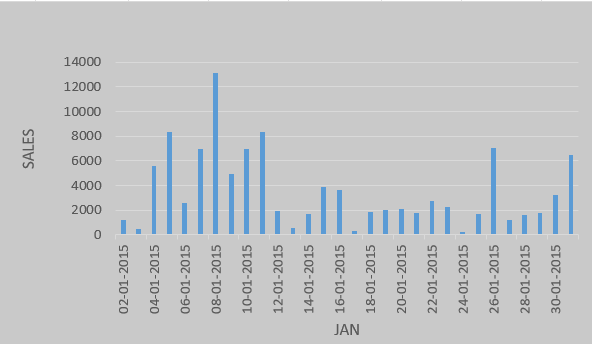
7.How can regional sales be displayed on the dashboard for easy interpretation?

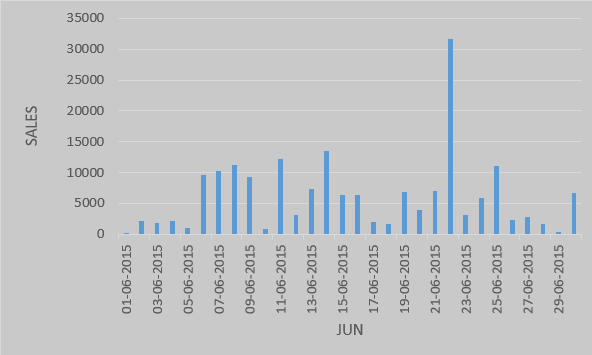
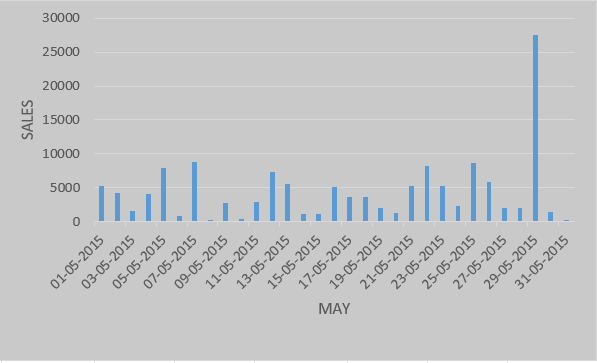
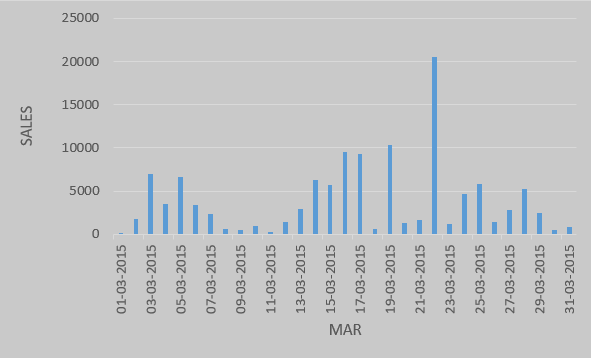
Ans:-From the pie chart we can easily interpret regional sales be displayed on the dashboard..



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

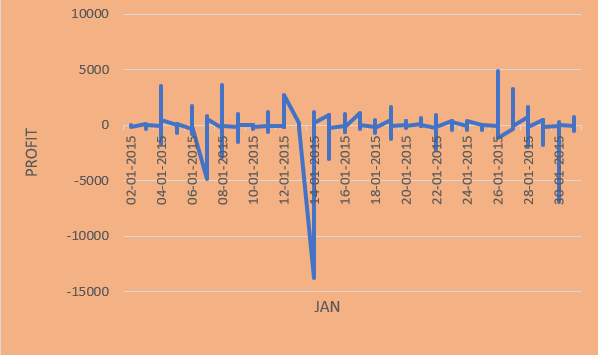
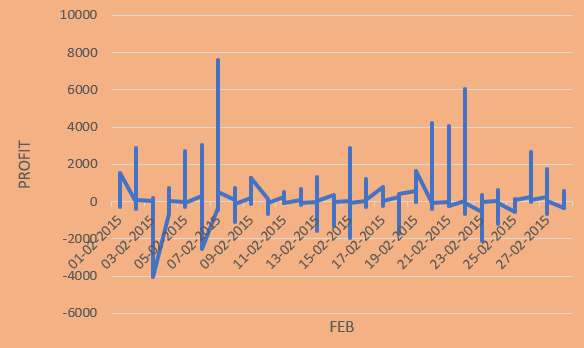
Ans:-

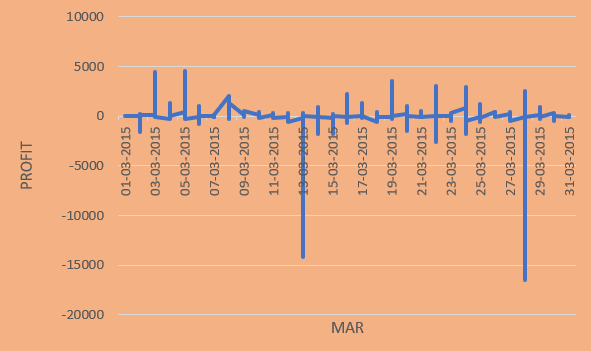
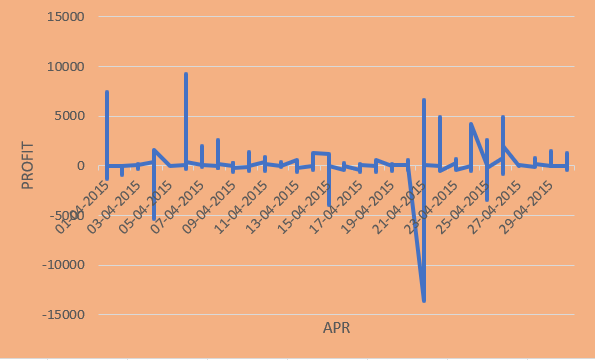


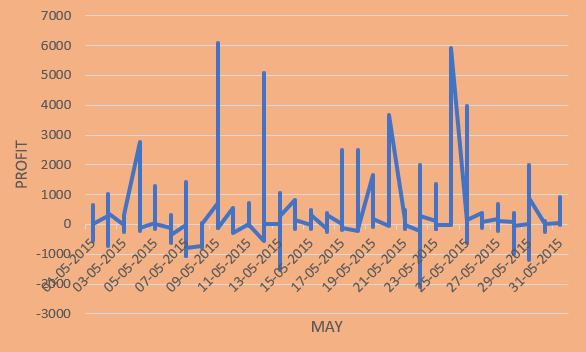
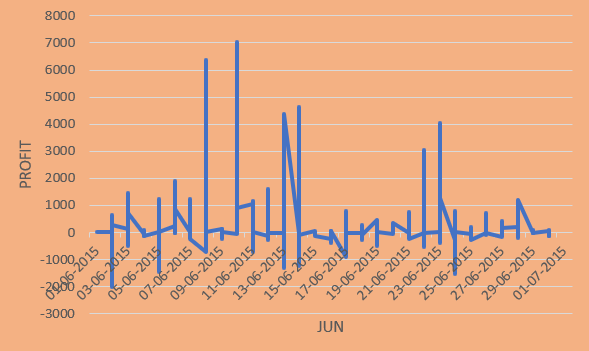


9.How can we visualize monthly profit variations on the dashboard?

Ans:-

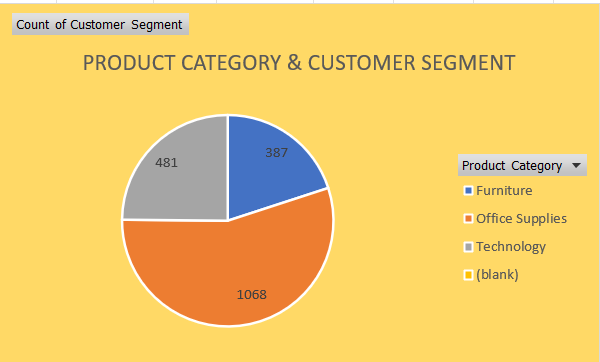
 

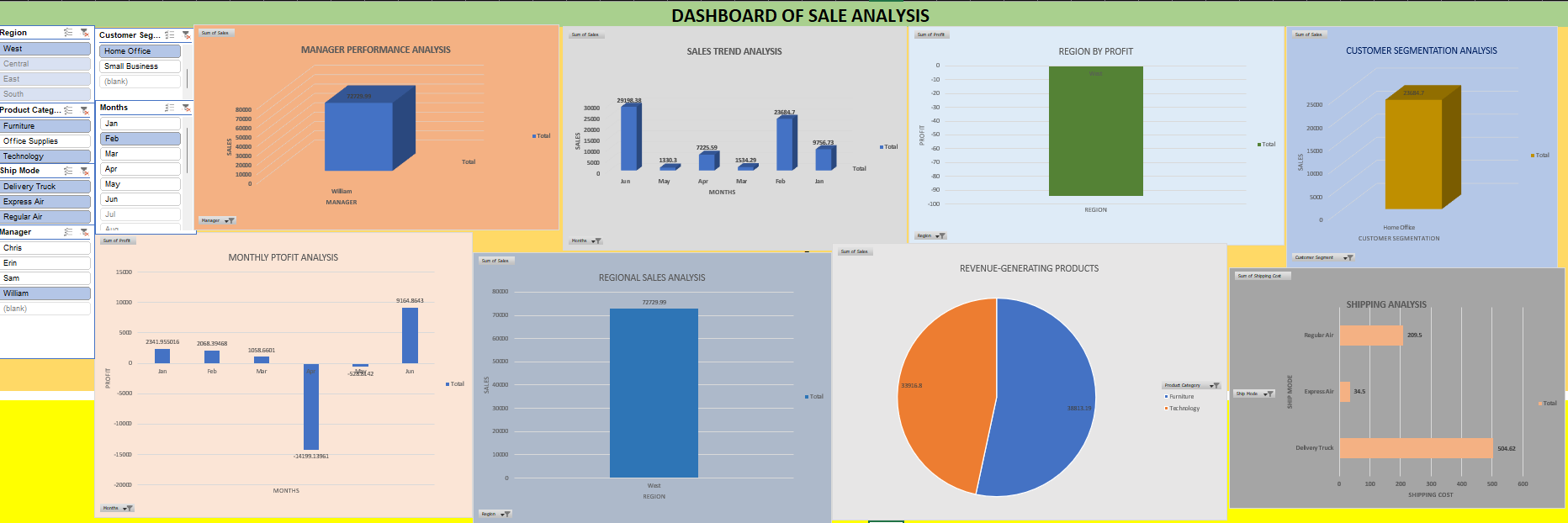
10.What visualizations can be used to analyze customers based on segments?

Ans:-



**DashBoard:**

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.



I have created this dashboard by combining all the results obtained in

the data analysis, then connecting the required slicers, adding images, objects, and style to the dashboard.

**SUMMARY OF REPORT**

In this report I have discussed in detail my project, its Tables, Chats and some compliment.I have explained Sales data and Statistics Dashboard using raw data using Excel. data using several excel features like pivot table, filtering, sorting, formulas to perform data analysis and deduce important results then representing them graphically using charts.

This report is a collection of data gathered from my point-of-sale system. A sales company should know about their customers, Understanding customers' wants, giving them custom solutions, and earning their trust are all important parts of the process. Sales matter when it comes to making money and growing a business. It is important to balance meeting customers’ needs and moving the company's mission forward.

I have also attached the preview of the dashboard, and all the objectives in this report.

**REFRENCES**

Sales analysis (Data Set)- <https://docs.google.com/spreadsheets/d/1B7w444UZ1eWQtnE3Z5Cmozipr3RkOAhe/edit?usp=drive_link&ouid=109618786931407309161&rtpof=true&sd=true>

**THANKING YOU TO PROVIDING ME THIS PROJECT**

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |