**DATA MANAGEMNET PROJECT REPORT**

(Project Semester: August-December 2022)



*SALES ANALYSIS OF SUPERSTORE*

Submitted by

Nirbhay Singh

12019256

Programme and Section: B.Tech(CSE), KM076

Course Code: INT217

Under the Guidance of

**Savleen Kaur - 18306**

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

# CERTIFICATE

This is to certify that Nirbhay Singh bearing Registration no. 12019256 has completed INT217 project titled, **"SALES ANALYSIS OF SUPERSTORE"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

**Savleen Kaur**

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

Date:

# DECLARATION

I, Nirbhay Singh, student of B.Tech CSE 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.

Date:

Registration No.: 12019256                                               Nirbhay Singh

# ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher Mrs. Savleen Kaur who gave me the golden opportunity to do this wonderful project of analysis of the data of a superstore namely "*SALES ANALYSIS OF SUPERSTORE*" which also helped me in doing a lot of research and I came to know about so many new things. I am thankful to them. Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

# INDEX

# INTRODUCTION

**Data Analysis** is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

The Analytics team of a Super Store anywhere in the world would want to design a Sales and Performance dashboard to analyze the sales based on various product categories and other factors which have a role to play in the running of the store. The store managing head, or the owner wants to add user control for product category, so users can select a category and can see the trend month-wise and product-wise accordingly.

The Analytics team would also want to analyze various other things like how many days the store takes to ship the product, how many times a Customer orders a product, how much time is there between the first and second order of the customer etc.

The Super Store's database keeps track of the following data fields:

- Order ID – Id of the order created by the customer.
- Order Date and Ship Date – Date when the item was ordered and the date when the item was shipped to the person.
- Shipping Mode – Mode of shipping
- Product Category – Category of the product
- Product – The product that was ordered
- Sales – Quantity of the items ordered
- Discount – Discount value on the product
- Profit – How much profit was earned on that product
- Order Priority – Priority of the order determining delivery
- Customer Name – Name of the person who ordered the item.
- Customer ID – Unique ID for each Customer
- Region – Region where the sales was made.
- Order Month – Month when the order took place.
- Manager – Managing Head of Regions
- Customer Segment – Tells about customer type

# SCOPE OF ANALYSIS

The super store wants to see and analyze the sales trend month-wise and product-wise and work upon the lagging segments and outperforming employees accordingly. The Analytics team also wants to create analyze the database in depth to help the super store grow exponentially. The Analytics team wishes to answer the following objectives: -

1.  Sales, Quantity and Profit of each product category
2.  Segment Distribution of each product category
3.  Sales and Profit throughout months of a sales of each product category
4.  Regional Sales of each product category
5.  Overall Sales Trend throughout months of a sales year
6.  Distribution of Order Priority
7.  Customer Ordering Trend
8.  Analysis of work Performance of Regional Managers
9.  Comparison of sales and profit product category wise
10. Analyzing Shipping priority share.

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

# ETL PROCESS

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Precisely, ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL stands for Extract, Transform and Load.

Before ETL, the dataset looked like this. **This data is taken from Kaggle.**

| Name | | id | Order ID | priority | off | Unit Price | Shipping | |
|------|--|----|----------|----------|-----|-----------|----------|--|
| Kristine Connolly | | 553 | 359 | Medium | 0.08 | 124.49 | 51.94 | |
| Alexander O'Brien | | 3106 | 548 | Critical | 0.04 | 3.08 | 0.99 | |
| Alexander O'Brien | | 3106 | 548 | Critical | 0.02 | 6.48 | 5.9 | |
| Alexander O'Brien | | 3106 | 548 | Critical | 0.04 | 125.99 | 4.2 | |
| Maxine Collier Grady | | 1106 | 646 | High | 0.01 | 9.31 | 3.98 | |
| Geoffrey Saunders | | 2382 | 962 | Low | 0.06 | 122.99 | 19.99 | |
| Geoffrey Saunders | | 2382 | 962 | Low | 0.08 | 68.81 | 60 | |
| Kristine Connolly | | 553 | 2433 | Not Specified | 0.07 | 2036.48 | 14.7 | |
| Jenny Gold | | 699 | 3042 | Medium | 0.1 | 4.26 | 1.2 | |
| Ross Simpson | | 471 | 3138 | Not Specified | 0.07 | 179.99 | 19.99 | |
| Jacqueline Noble | | 342 | 3332 | Critical | 0.01 | 3.26 | 1.86 | |
| Caroline Johnston | | 102 | 3397 | Medium | 0.1 | 19.98 | 4 | |
| Caroline Johnston | | 102 | 3397 | Medium | 0.09 | 2.88 | 1.49 | |
| Wesley Waller | | 181 | 3585 | High | 0.07 | 1.68 | 1.57 | |
| Nina Horne Kelly | | 1733 | 3841 | Not Specified | 0.02 | 60.98 | 49 | |
| Nina Horne Kelly | | 1733 | 3841 | Not Specified | 0.02 | 1270.99 | 19.99 | |
| Andrew Gonzalez | | 2882 | 4839 | Critical | 0.05 | 6.48 | 8.73 | |
| Eleanor Swain | | 272 | 5509 | Low | 0.02 | 5.58 | 5.3 | |
| Eleanor Swain | | 272 | 5509 | Low | 0.03 | 40.89 | 18.98 | |
| Troy Cassidy | | 2431 | 5920 | High | 0.07 | 155.06 | 7.07 | |
| Louis Parrish | | 1193 | 5984 | Low | 0.03 | 10.64 | 5.16 | |
| Louis Parrish | | 1193 | 5984 | Low | 0.03 | 7.96 | 4.95 | |

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

**Step 1: Removing the blank cells from the dataset.**

For this, select the whole dataset. Go to Find and Select in the Home tab of excel. Select Go to Special from the drop-down menu and then tick the blank option. All the blank cells will

be selected. Then go to Delete option in the home tab again and select Delete Rows from the drop-down menu. This will remove any rows with blank cells.

| Order Pric | Discount | Unit | Shipping | Customer ID | Customer N | Ship Mode | | Customer Se | Product Category | Product | Product Container |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Critical | 0.06 | 9.5 | 7.29 | 11 | Marcus Dunk | Regular Air | | Home Office | Furniture | Office Fur | Small Pack |
| Critical | 0 | 4.4 | 4.99 | 15 | Timothy Ree | Regular Air | | Small Business | Office Supplies | Envelope | Small Box |
| Critical | 0.07 | ### | 8.73 | 53 | Sidney Russ | Delivery Truck | | Corporate | Technology | Office Ma | Jumbo Box |
| Critical | 0.06 | 8.6 | 6.14 | 123 | Shawn Stern | Regular Air | | Home Office | Office Supplies | Scissors, | Small Pack |
| Critical | 0.04 | 19 | 9.54 | 136 | Dale Gillespi | Regular Air | | Small Business | Office Supplies | Paper | Small Box |
| Critical | 0.09 | 11 | 3.37 | 136 | Dale Gillespi | Regular Air | | Small Business | Office Supplies | Scissors, | Small Pack |
| Critical | 0.03 | 23 | 11.54 | 142 | Brooke Wee | Regular Air | | Small Business | Office Supplies | Paper | Small Box |
| Critical | 0.05 | 11 | 3.37 | 144 | Marguerite M | Regular Air | | Small Business | Office Supplies | Scissors, | Small Pack |
| Critical | 0.09 | 33 | 5.5 | 151 | Geoffrey Zhu | Regular Air | | Home Office | Technology | Compute | Small Box |
| Critical | 0.09 | 2.9 | 0.7 | 152 | Kent Kerr | Regular Air | | Consumer | Office Supplies | Pens & A | Wrap Bag |
| Critical | 0.01 | 96 | 4.9 | 156 | Diana Xu | Regular Air | | Corporate | Technology | Telephor | Small Box |
| Critical | 0.05 | 1.9 | 1.49 | 171 | Christina Mat | Regular Air | | Corporate | Office Supplies | Binders a | Small Box |
| Critical | 0.02 | 50 | 19.99 | 181 | Wesley Wall | Regular Air | | Small Business | Technology | Compute | Small Box |
| Critical | 0.02 | 50 | 19.99 | 184 | Phillip Holme | Regular Air | | Small Business | Technology | Compute | Small Box |
| Critical | 0 | 162 | 19.99 | 197 | Samantha W | Regular Air | | Small Business | Office Supplies | Storage & | Small Box |
| Critical | 0 | 162 | 19.99 | 198 | Leroy Blanch | Regular Air | | Small Business | Office Supplies | Storage & | Small Box |
| Critical | 0.06 | 280 | 23.19 | 234 | Don Camero | Delivery Truck | | Small Business | Office Supplies | Appliance | Jumbo Drum |
| Critical | 0.02 | 2.6 | 1.3 | 250 | Brenda Nels | Express Air | | Corporate | Office Supplies | Pens & A | Wrap Bag |
| Critical | 0.02 | 66 | 3.9 | 250 | Brenda Nels | Regular Air | | Corporate | Technology | Telephor | Small Box |
| Critical | 0.03 | 8.3 | 2.64 | 256 | Irene Li | Regular Air | | Home Office | Office Supplies | Scissors, | Small Pack |
| Critical | 0.04 | 2 | 0.7 | 276 | Lucille Ranki | Express Air | | Corporate | Office Supplies | Rubber B | Wrap Bag |
| Critical | 0.03 | 56 | 5 | 282 | Vickie Andre | Regular Air | | Corporate | Technology | Telephor | Small Box |
| Critical | 0.09 | 28 | 1.99 | 288 | Patricia Cole | Regular Air | | Small Business | Technology | Compute | Small Pack |
| Critical | 0.08 | 66 | 4.99 | 288 | Patricia Cole | Express Air | | Small Business | Technology | Telephor | Small Box |
| Critical | 0.06 | 276 | 24.49 | 335 | Curtis O'Conr | Regular Air | | Corporate | Furniture | Chairs & | Large Box |
| Critical | 0.09 | 6.3 | 5.29 | 335 | Curtis O'Conr | Regular Air | | Corporate | Furniture | Office Fur | Small Box |
| Critical | 0.01 | 3.3 | 1.86 | 342 | Jacqueline N | Regular Air | | Corporate | Office Supplies | Pens & A | Wrap Bag |
| Critical | 0.03 | 15 | 27.75 | 343 | Lynn Epstein | Delivery Truck | | Corporate | Furniture | Tables | Jumbo Box |

## Step 2: Removing columns which are not properly defined or not crucial to our analysis.

For this we will columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete option in the home tag and then select Delete Columns from the drop-down menu.

| Row ID | Order Priorit | Discoun | Unit Price | Shipping Cos | Customer II | Customer Name | Ship Mode | Customer Segme | Product Catego |
|---|---|---|---|---|---|---|---|---|---|
| 24926 | Critical | 0.09 | 517.48 | 16.63 | 1020 | Julie Porter | Delivery Truck | Small Business | Technology |
| 23562 | Critical | 0.07 | 4.13 | 5.04 | 1020 | Julie Porter | Regular Air | Small Business | Office Supplies |
| 23563 | Critical | 0 | 4.48 | 2.5 | 1020 | Julie Porter | Regular Air | Small Business | Office Supplies |
| 18921 | Critical | 0.02 | 39.06 | 10.55 | 1023 | Glen Newman | Regular Air | Small Business | Office Supplies |
| 18922 | Critical | 0.1 | 37.7 | 2.99 | 1023 | Glen Newman | Regular Air | Small Business | Office Supplies |
| 21959 | Critical | 0.07 | 125.99 | 2.5 | 1035 | Kent Burton | Regular Air | Home Office | Technology |
| 21960 | Critical | 0.03 | 99.99 | 19.99 | 1036 | Jessica Huffman | Regular Air | Home Office | Technology |
| 20669 | Critical | 0.1 | 7.64 | 5.83 | 1038 | Jon Hale | Regular Air | Corporate | Office Supplies |
| 18404 | Critical | 0.06 | 55.94 | 4 | 1041 | Mildred Chase | Regular Air | Small Business | Technology |
| 18405 | Critical | 0.07 | 6.3 | 0.5 | 1041 | Mildred Chase | Regular Air | Small Business | Office Supplies |
| 20937 | Critical | 0 | 14.42 | 6.75 | 1042 | Jerome Burch | Express Air | Small Business | Office Supplies |
| 3926 | Critical | 0.02 | 209.84 | 21.21 | 1044 | Erin Ballard | Regular Air | Home Office | Furniture |
| 3927 | Critical | 0.01 | 194.3 | 11.54 | 1044 | Erin Ballard | Regular Air | Home Office | Furniture |
| 19445 | Critical | 0.01 | 15.99 | 13.18 | 1065 | Vicki Bond | Regular Air | Corporate | Office Supplies |
| 26060 | Critical | 0.01 | 2.89 | 0.5 | 1113 | Julia Reynolds | Regular Air | Corporate | Office Supplies |
| 26061 | Critical | 0 | 55.99 | 5 | 1113 | Julia Reynolds | Regular Air | Corporate | Technology |
| 24224 | Critical | 0.09 | 9.11 | 2.15 | 1155 | Alex Nicholson | Express Air | Consumer | Office Supplies |
| 24225 | Critical | 0.08 | 15.04 | 1.97 | 1155 | Alex Nicholson | Regular Air | Consumer | Office Supplies |
| 24358 | Critical | 0.07 | 400.97 | 48.26 | 1186 | Glenda Herbert | Delivery Truck | Consumer | Technology |
| 21206 | Critical | 0.1 | 120.98 | 9.07 | 1233 | Gary Hester | Express Air | Consumer | Office Supplies |
| 21207 | Critical | 0.02 | 152.48 | 6.5 | 1233 | Gary Hester | Express Air | Consumer | Technology |
| 20233 | Critical | 0.06 | 200.97 | 15.59 | 1241 | Bradley Schroeder | Delivery Truck | Small Business | Technology |
| 23815 | Critical | 0.06 | 80.98 | 35 | 1254 | Anne Bland | Regular Air | Home Office | Office Supplies |
| 18693 | Critical | 0.04 | 2.52 | 1.92 | 1257 | Ryan Foster | Regular Air | Home Office | Office Supplies |
| 21771 | Critical | 0.02 | 73.98 | 14.52 | 1261 | Vickie Gonzalez | Regular Air | Home Office | Technology |
| 24559 | Critical | 0.05 | 5.28 | 6.26 | 1265 | Danielle Kramer | Regular Air | Home Office | Office Supplies |
| 23363 | Critical | 0.01 | 13.99 | 7.51 | 1267 | Rosemary Branch | Regular Air | Corporate | Technology |

## Step 3: Giving proper and appropriate column names.

The dataset does not have proper columns so our next step would be to giver proper column names to the columns wherever required.

| Order F | Discount | Unit Pric | Shipping | Customer ID | Customer N | Ship Mode | Customer Se | Product Category | Product Sub-Ca | Product Container |
|---------|----------|-----------|----------|-------------|------------|-----------|-------------|------------------|----------------|-------------------|
| Critical | 0.06 | 9.48 | 7.29 | 11 | Marcus Dunl | Regular Air | Home Office | Furniture | Office Furnishings | Small Pack |
| Critical | 0 | 4.42 | 4.99 | 15 | Timothy Ree | Regular Air | Small Business | Office Supplies | Envelopes | Small Box |
| Critical | 0.07 | 3502.14 | 8.73 | 53 | Sidney Russ | Delivery Truck | Corporate | Technology | Office Machines | Jumbo Box |
| Critical | 0.06 | 8.57 | 6.14 | 123 | Shawn Stern | Regular Air | Home Office | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | 0.04 | 18.97 | 9.54 | 136 | Dale Gillespi | Regular Air | Small Business | Office Supplies | Paper | Small Box |
| Critical | 0.09 | 10.98 | 3.37 | 136 | Dale Gillespi | Regular Air | Small Business | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | 0.03 | 22.84 | 11.54 | 142 | Brooke Wee | Regular Air | Small Business | Office Supplies | Paper | Small Box |
| Critical | 0.05 | 10.98 | 3.37 | 144 | Marguerite M | Regular Air | Small Business | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | 0.09 | 32.98 | 5.5 | 151 | Geoffrey Zhu | Regular Air | Home Office | Technology | Computer Periphe | Small Box |
| Critical | 0.09 | 2.88 | 0.7 | 152 | Kent Kerr | Regular Air | Consumer | Office Supplies | Pens & Art Supplie | Wrap Bag |
| Critical | 0.01 | 95.99 | 4.9 | 156 | Diana Xu | Regular Air | Corporate | Technology | Telephones and C | Small Box |
| Critical | 0.05 | 1.88 | 1.49 | 171 | Christina Mat | Regular Air | Corporate | Office Supplies | Binders and Binde | Small Box |
| Critical | 0.02 | 49.99 | 19.99 | 181 | Wesley Wall | Regular Air | Small Business | Technology | Computer Periphe | Small Box |
| Critical | 0.02 | 49.99 | 19.99 | 184 | Phillip Holme | Regular Air | Small Business | Technology | Computer Periphe | Small Box |
| Critical | 0 | 161.55 | 19.99 | 197 | Samantha W | Regular Air | Small Business | Office Supplies | Storage & Organiz | Small Box |
| Critical | 0 | 161.55 | 19.99 | 198 | Leroy Blanch | Regular Air | Small Business | Office Supplies | Storage & Organiz | Small Box |
| Critical | 0.06 | 279.81 | 23.19 | 234 | Don Camero | Delivery Truck | Small Business | Office Supplies | Appliances | Jumbo Drum |
| Critical | 0.02 | 2.58 | 1.3 | 250 | Brenda Nels | Express Air | Corporate | Office Supplies | Pens & Art Supplie | Wrap Bag |
| Critical | 0.02 | 65.99 | 3.9 | 250 | Brenda Nels | Regular Air | Corporate | Technology | Telephones and C | Small Box |
| Critical | 0.03 | 8.34 | 2.64 | 256 | Irene Li | Regular Air | Home Office | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | 0.04 | 1.98 | 0.7 | 276 | Lucille Ranki | Express Air | Corporate | Office Supplies | Rubber Bands | Wrap Bag |
| Critical | 0.03 | 55.99 | 5 | 282 | Vickie Andre | Regular Air | Corporate | Technology | Telephones and C | Small Box |
| Critical | 0.09 | 28.48 | 1.99 | 288 | Patricia Cole | Regular Air | Small Business | Technology | Computer Periphe | Small Pack |
| Critical | 0.08 | 65.99 | 4.99 | 288 | Patricia Cole | Express Air | Small Business | Technology | Telephones and C | Small Box |
| Critical | 0.06 | 276.2 | 24.49 | 335 | Curtis O'Conr | Regular Air | Corporate | Furniture | Chairs & Chairmats | Large Box |

**Step 4: Excluding the NULL values from the data.**

We'll be using Tableau prep for this work as it'll make the work simple and faster because we might not know how many null values could be there in this huge data set. Tableau helps us doing one step cleaning with ease.
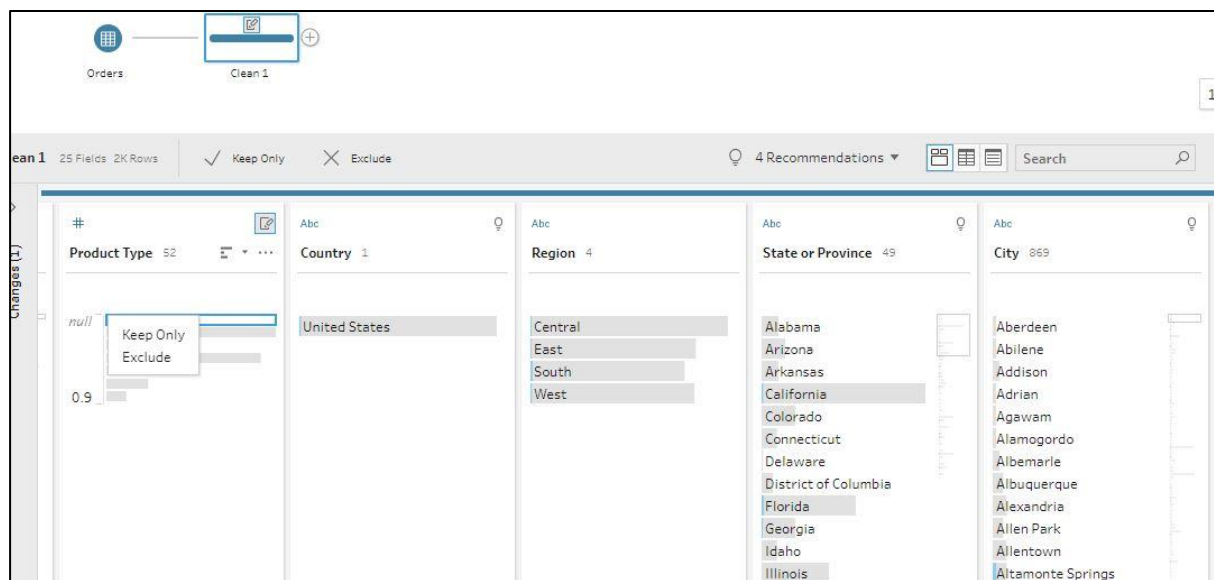
## Step 5: Improvising Proper Data Formatting

Without proper Data Formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format. For example, the dates should be in the date format and price and sales should be in currency format for better results.

| Order P | Discoun | Unit Price | Shipping | Customer ID | Customer N | Ship Mode | Customer Sec | Product Category | Product Sub-Ca | Product Cont |
|---|---|---|---|---|---|---|---|---|---|---|
| Critical | $0.06 | $9.48 | $7.29 | 11 | Marcus Dunl | Regular Air | Home Office | Furniture | Office Furnishings | Small Pack |
| Critical | $0.00 | $4.42 | $4.99 | 15 | Timothy Ree | Regular Air | Small Business | Office Supplies | Envelopes | Small Box |
| Critical | $0.07 | $3,502.14 | $8.73 | 53 | Sidney Russ | Delivery Truck | Corporate | Technology | Office Machines | Jumbo Box |
| Critical | $0.06 | $8.57 | $6.14 | 123 | Shawn Stern | Regular Air | Home Office | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | $0.04 | $18.97 | $9.54 | 136 | Dale Gillespi | Regular Air | Small Business | Office Supplies | Paper | Small Box |
| Critical | $0.09 | $10.98 | $3.37 | 136 | Dale Gillespi | Regular Air | Small Business | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | $0.03 | $22.84 | $11.54 | 142 | Brooke Wee | Regular Air | Small Business | Office Supplies | Paper | Small Box |
| Critical | $0.05 | $10.98 | $3.37 | 144 | Marguerite M | Regular Air | Small Business | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | $0.09 | $32.98 | $5.50 | 151 | Geoffrey Zhu | Regular Air | Home Office | Technology | Computer Periphe | Small Box |
| Critical | $0.09 | $2.88 | $0.70 | 152 | Kent Kerr | Regular Air | Consumer | Office Supplies | Pens & Art Supplie | Wrap Bag |
| Critical | $0.01 | $95.99 | $4.90 | 156 | Diana Xu | Regular Air | Corporate | Technology | Telephones and C | Small Box |
| Critical | $0.05 | $1.88 | $1.49 | 171 | Christina Mat | Regular Air | Corporate | Office Supplies | Binders and Binde | Small Box |
| Critical | $0.02 | $49.99 | $19.99 | 181 | Wesley Wall | Regular Air | Small Business | Technology | Computer Periphe | Small Box |
| Critical | $0.02 | $49.99 | $19.99 | 184 | Phillip Holme | Regular Air | Small Business | Technology | Computer Periphe | Small Box |
| Critical | $0.00 | $161.55 | $19.99 | 197 | Samantha W | Regular Air | Small Business | Office Supplies | Storage & Organiz | Small Box |
| Critical | $0.00 | $161.55 | $19.99 | 198 | Leroy Blanch | Regular Air | Small Business | Office Supplies | Storage & Organiz | Small Box |
| Critical | $0.06 | $279.81 | $23.19 | 234 | Don Camero | Delivery Truck | Small Business | Office Supplies | Appliances | Jumbo Drum |
| Critical | $0.02 | $2.58 | $1.30 | 250 | Brenda Nels | Express Air | Corporate | Office Supplies | Pens & Art Supplie | Wrap Bag |
| Critical | $0.02 | $65.99 | $3.90 | 250 | Brenda Nels | Regular Air | Corporate | Technology | Telephones and C | Small Box |
| Critical | $0.03 | $8.34 | $2.64 | 256 | Irene Li | Regular Air | Home Office | Office Supplies | Scissors, Rulers a | Small Pack |
| Critical | $0.04 | $1.98 | $0.70 | 276 | Lucille Ranki | Express Air | Corporate | Office Supplies | Rubber Bands | Wrap Bag |
| Critical | $0.03 | $55.99 | $5.00 | 282 | Vickie Andre | Regular Air | Corporate | Technology | Telephones and C | Small Pack |
| Critical | $0.09 | $28.48 | $1.99 | 288 | Patricia Cole | Regular Air | Small Business | Technology | Computer Periphe | Small Pack |

## Step 6: Removing Duplicate Values

It might be possible that our data may be containing duplicate values which may hinder in precise analysis. So, our last task in ETL will be removing duplicate values and making our data perfect for analysis.



VI

# ANALYSIS OF DATASET

1. **Monthly Sales and profit of each category**

   **Description:**

   By knowing about sales and profit over month we can know about the months which are more profitable for sales and hence customize our advertisement plan to increase the sales even more. After finding out the sales and profit we visualize the result with the help of a stacked bar graph.
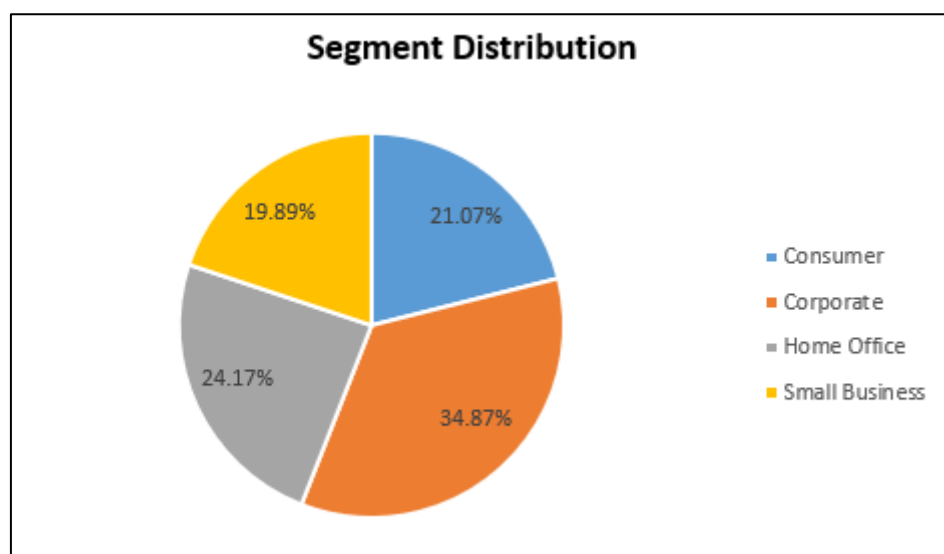
   **Specific function and requirements**

   We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

   **Results:**

   | Month | Monthly Sales | Monthly Profit |
   |---|---|---|
   | Jan | 240998.4700 | -9818.4254 |
   | Feb | 346066.8100 | 45600.6794 |
   | Mar | 252497.0600 | -2766.8951 |
   | Apr | 381645.9100 | 50780.8532 |
   | May | 303301.2300 | 62756.3697 |
   | Jun | 334850.1500 | 58517.7420 |
   | Jul | 22165.1400 | 9953.0759 |
   | Grand Total | 1881524.77 | 215023.3997 |

   **Visualization:**

   The results are then visualized in the form of a stacked bar graph for both profit and sales

## 2. Segment Distribution of each product category:

**Description:**

By knowing which segment of sales has the most number of sales and which has least we can identify factors which affect the sales and thereby improve our strategy of making sales.

**Specific function and requirements**

We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

**Results:**

| Customer Type | Count of Customer Segment |
|---|---|
| Consumer | 408 |
| Corporate | 675 |
| Home Office | 468 |
| Small Business | 385 |
| Grand Total | 1936 |

**Visualization:**

We will use a pie chart to visualize the distribution.



Segment Distribution

**3. Sales throughout months of a sales year.**

**Description:**

Monthly sales can help us identify which month is more profitable and helps identify the factor which helps us to do so. We can apply the identified the factors in other months to increase the sales.

**Specific function and requirements:**

We have to create a pivot table. No specific functions are used. We then put the months and sum of sales in the columns.

**Results:**

| Row Labels | Sum of Sales | Month |
|---|---|---|
| 01-Jan | 8666.18 | January |
| 02-Jan | 24319.92 | January |
| 03-Jan | 10134.42 | January |
| 04-Jan | 5470.58 | January |
| 05-Jan | 20160.51 | January |
| 06-Jan | 11123.41 | January |
| 07-Jan | 16551.82 | January |
| 08-Jan | 5388.84 | January |
| 09-Jan | 13926.83 | January |
| 10-Jan | 11520.37 | January |
| 11-Jan | 3355.12 | January |
| 12-Jan | 14882.67 | January |
| 13-Jan | 1772.04 | January |
| 14-Jan | 5200.19 | January |
| 15-Jan | 4829.95 | January |
| 16-Jan | 3964.13 | January |
| 17-Jan | 6667.69 | January |
| 18-Jan | 2987.38 | January |
| 19-Jan | 1652.89 | January |
| 20-Jan | 6109.01 | January |
| 21-Jan | 12281.51 | January |

**Visualization:**

The results are visualized with the help of line graph with a trend line displaying the trend of sales over months.

**Weekly Sales Trend**
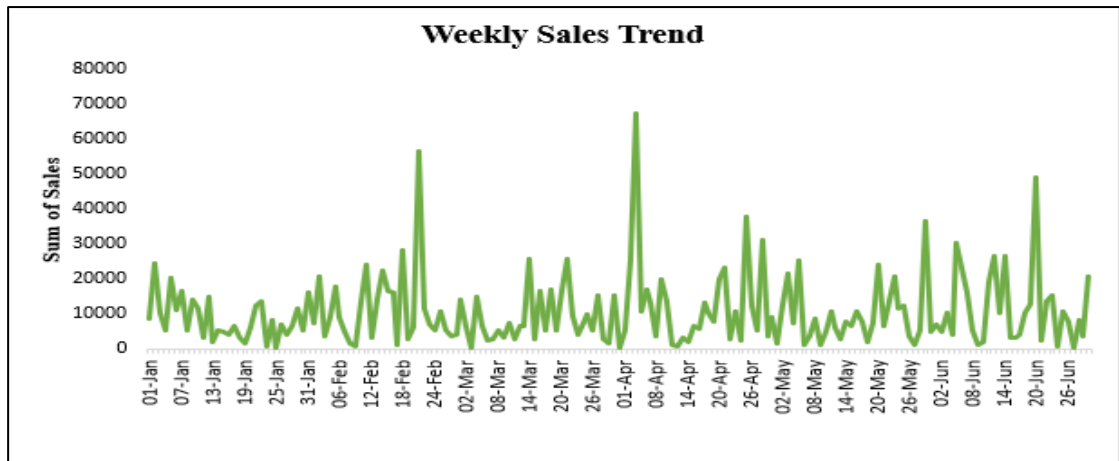
## 4. Sales and their priority

**Description:**

Every sale is going to have an order priority associated with it. Greater the priority, faster the item would be shipped out and received.
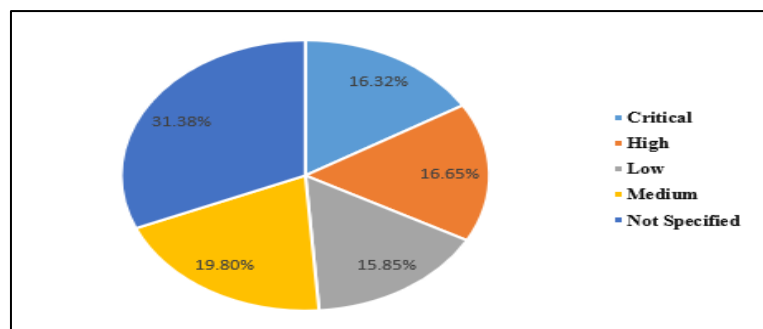
**Specific function and requirements**

We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

**Results:**

| Row Labels | Percentage Of Order Type |
|---|---|
| Critical | 16.32% |
| High | 16.65% |
| Low | 15.85% |
| Medium | 19.80% |
| Not Specified | 31.38% |
| Grand Total | 100.00% |

**Visualization:**

We visualize the above results with the help a pie chart created using pivot charts.



X

**5. Comparison of sales of each product category**

**Description:**

By comparing sales of each product category side by side, we can come to know what kind of products are sold the most and which the least. This information can help us target customers more effectively to improve the sales and thus by increasing profits which is the main goal of any organization.

**Specific function and requirements:**

We have to create a pivot table. No specific functions are used. Product category is used as columns with summation of profit and sales of each product category.

**Results:**

| Row Labels | Sales Category |
|---|---|
| Appliances | 1589.65 |
| Binders and Binder Accessories | 1493.36 |
| Bookcases | 1957.71 |
| Chairs & Chairmats | 3560.59 |
| Computer Peripherals | 1094.84 |
| Copiers and Fax | 476.99 |
| Envelopes | 334.03 |
| Labels | 93.03 |
| Office Furnishings | 1942.03 |
| Office Machines | 1903.64 |
| Paper | 1777.19 |
| Pens & Art Supplies | 498.16 |
| Rubber Bands | 46.54 |
| Scissors, Rulers and Trimmers | 163.95 |
| Storage & Organization | 2349.54 |
| Tables | 4397.51 |
| Telephones and Communication | 1069.67 |
| **Grand Total** | **24748.43** |

**Visualization:**

The results are visualized in the form of stacked bar graph.

Shipping Expenditure
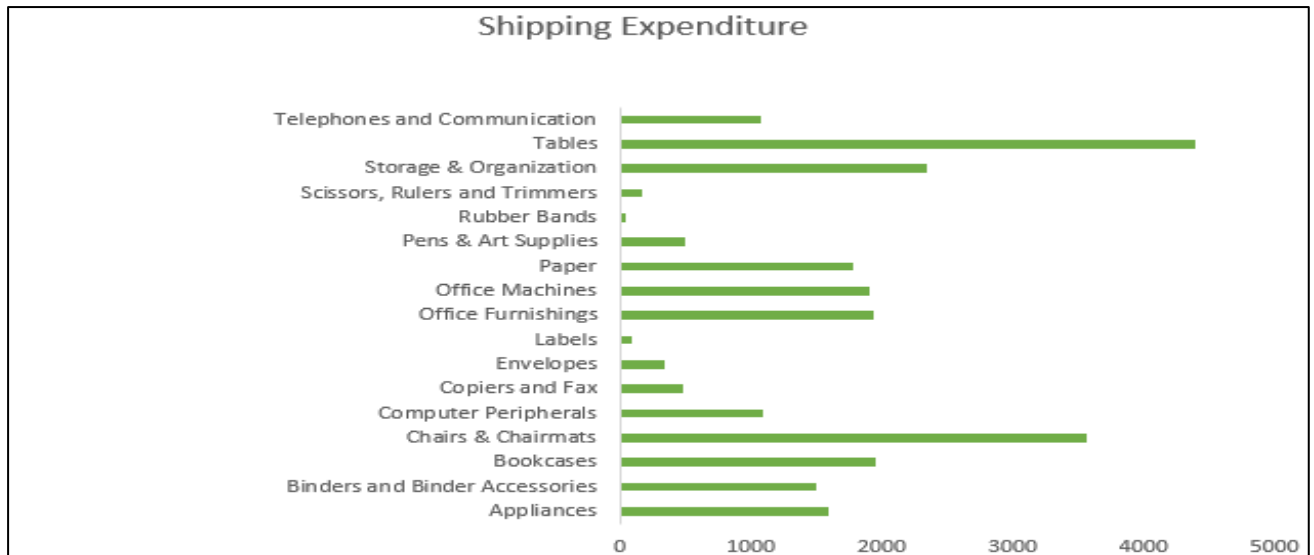
## 6. Employee Performance

**Description:**

In this we analyze which regional manager is doing well and which one is performing the least. It'll help us giving them incentives, promoting them and training them for better performance,
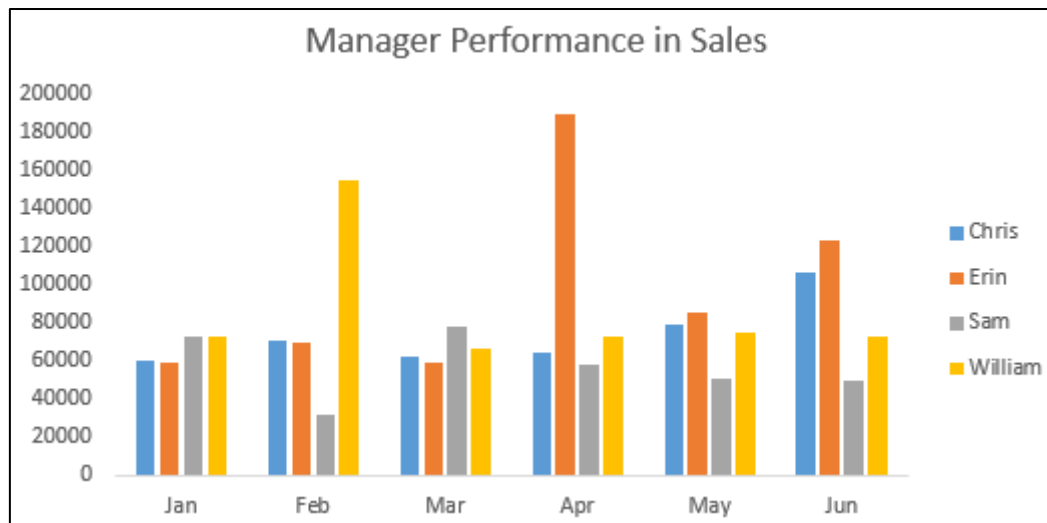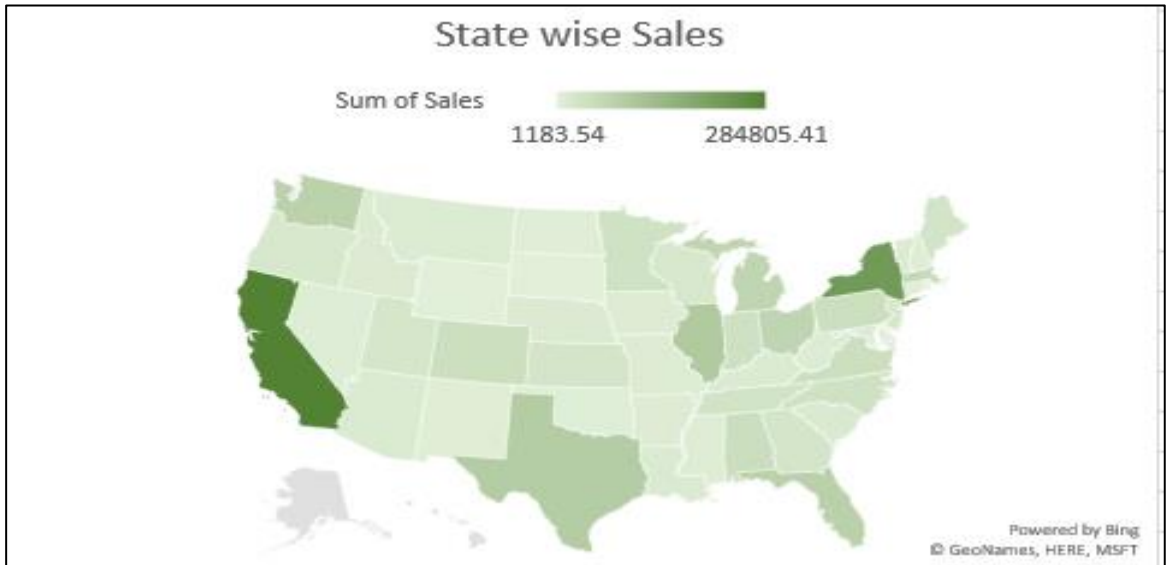
**Specific function and requirements:**

We have to create a pivot table. No specific functions are used.

**Results:**

| Sum of Sales | Manager | | | | |
|---|---|---|---|---|---|
| Month | Chris | Erin | Sam | William | Grand Total |
| ⊞ Jan | 59702.73 | 59185.16 | 73062.34 | 73048.32 | 264998.55 |
| ⊞ Feb | 70282.44 | 69109.8 | 31754.39 | 154355.81 | 325502.44 |
| ⊞ Mar | 62075.31 | 59137.04 | 77874.78 | 66080 | 265167.13 |
| ⊞ Apr | 64058.38 | 189208.29 | 58285.11 | 72477.41 | 384029.19 |
| ⊞ May | 79253.14 | 84858.81 | 50887.39 | 75231.51 | 290230.85 |
| ⊞ Jun | 106073.62 | 123225.09 | 49464.31 | 72833.59 | 351596.61 |
| Grand Total | 441445.62 | 584724.19 | 341328.32 | 514026.64 | 1881524.77 |

**Visualization:**



## 7. Regional Sales Analysis

**Description:**

In this we analyze which particular region is having most amount of sales and which is least. Furthermore we can look upon the factors which might be impacting the sales and we can look upon them to increase the sales and invest in the areas of maximum sales.

**Specific function and requirements:**

We have to create a pivot table. No specific functions are used.

**Results:**

| State or Province | Sum of Sales |
|---|---|
| Alabama | 46826.45 |
| Arizona | 14367.86 |
| Arkansas | 11724.43 |
| California | 284805.41 |
| Colorado | 45843.45 |
| Connecticut | 6540.54 |
| Delaware | 1257.76 |
| District of Columbia | 68946.66 |
| Florida | 81205.22 |
| Georgia | 29050.79 |
| Idaho | 13922.92 |
| Illinois | 98971.25 |
| Indiana | 39314.55 |
| Iowa | 10977.69 |

**Visualization:**



State wise Sales

Sum of Sales

1183.54          284805.41

Powered by Bing
© GeoNames, HERE, MSFT

## 8. Days to Ship a Product

**Description:**

In this we analyze how much time the super store is taking to ship a product after successful placement of order by the user. This can help us to improve the customer service and improve the service quality provided to the customer.

**Specific function and requirements:**

We are using Tableau prep for this and simply creating a calculated field .



Add Field

Field Name

Calculation1

[Ship Date]-[Order Date]

**Results:**

| 📅 Order Date | 📅 Ship Date | # DAYS TO SHIP |
|---|---|---|
| 07/01/2015 | 08/01/2015 | 1 |
| 13/06/2015 | 15/06/2015 | 2 |
| 15/02/2015 | 17/02/2015 | 2 |
| 12/05/2015 | 14/05/2015 | 2 |
| 12/05/2015 | 13/05/2015 | 1 |
| 12/05/2015 | 13/05/2015 | 1 |
| 12/05/2015 | 13/05/2015 | 1 |
| 08/04/2015 | 09/04/2015 | 1 |
| 28/05/2015 | 28/05/2015 | 0 |

## 9. Customer Ordering Trend

**Description:**

In this we analyze  after how much time does a particular revisits us and places the order again. We can create offers and Discounts accordingly and increase the customer engagement for frequent visits. This will ultimately help us to improve the customer service and improve the service quality provided to the customer.

**Specific function and requirements:**

We are using Tableau prep for this and applying aggregate function to extract the first order date and the second order date and then finally joining them in a single field.
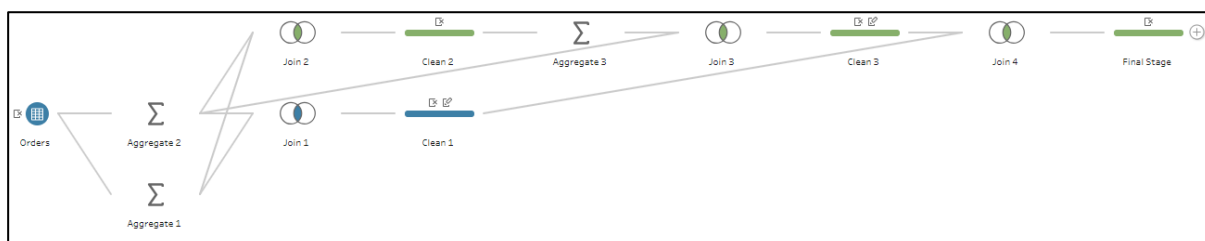
**Results:**

| Customer ID | 1st Purchased Date | 2nd Purchase Order Date |
|---|---|---|
| 2,391 | 25/05/2015 | 04/06/2015 |
| 210 | 17/01/2015 | 02/06/2015 |
| 2,548 | 04/04/2015 | 21/04/2015 |
| 2,668 | 28/03/2015 | 21/04/2015 |
| 2,840 | 19/03/2015 | 11/06/2015 |
| 181 | 21/02/2015 | 22/05/2015 |
| 1,683 | 14/02/2015 | 14/03/2015 |
| 2,797 | 10/01/2015 | 21/01/2015 |
| 1,341 | 10/02/2015 | 06/06/2015 |
| 92 | 17/05/2015 | 07/06/2015 |
| 1,026 | 07/02/2015 | 29/05/2015 |
| 3,230 | 21/01/2015 | 13/06/2015 |
| 898 | 12/01/2015 | 27/01/2015 |
| 1,671 | 09/02/2015 | 12/05/2015 |
| 1,416 | 06/05/2015 | 25/06/2015 |

# ANALYSIS RESULTS

### 1. Monthly Profit Trend



Sales was high in the Jan but still resulted in negative profit i.e. loss. Still the super store managed to work well and increase the profit exponentially by the end of Jun.

### 2. Segment Distribution of each product category
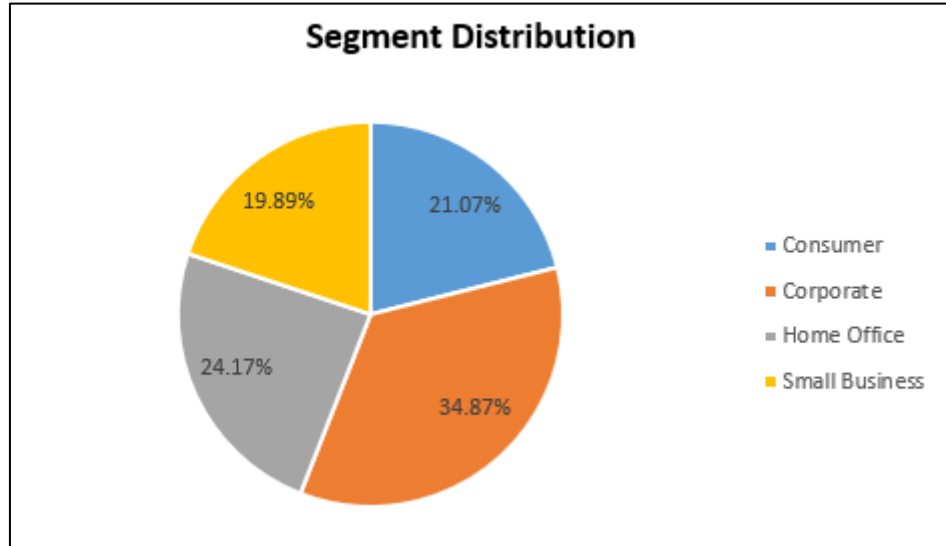


It is clear that Corporate Sector is our valuable customer followed by Home Office. We can offer them special discounts and can have tie ups to increase the engagement. Further, Small Business is lagging in our partnership and we can offer them new offers for more sales.

**3. Sales throughout months of a sales year.**
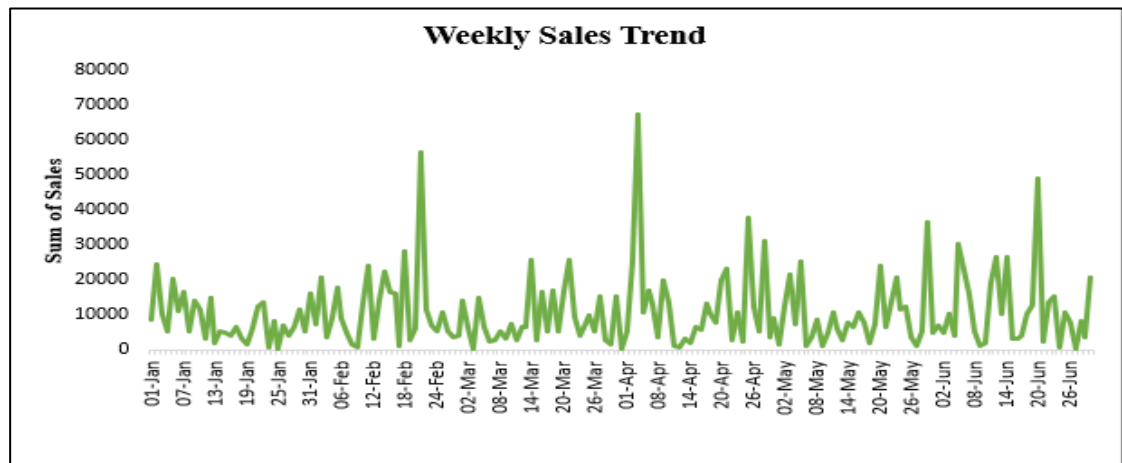


Sales were at peak once in mid Feb and again in Starting of the April followed by June End. We can create offers for other times as well to increase the sales growth.

**4. Sales and their priority**

| Row Labels | Percentage Of Order Type |
|---|---|
| Critical | 16.32% |
| High | 16.65% |
| Low | 15.85% |
| Medium | 19.80% |
| Not Specified | 31.38% |
| Grand Total | 100.00% |

We are having more of the orders without any priority followed by medium priority. We can enhance this by introducing faster services ultimately increasing the revenue.

## 5. Comparison of sales of each product category
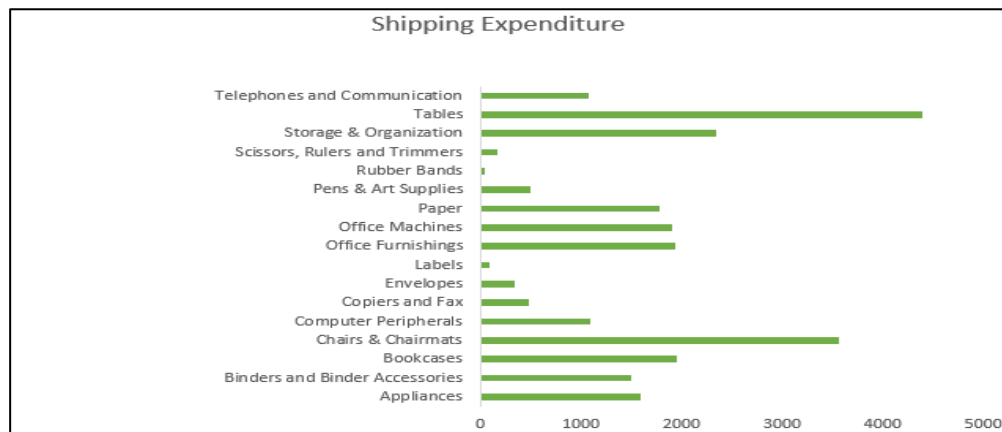


It is clear that tables are our best selling products followed by chairs and chair mats. We can work upon the one's not performing well to increase their sales also.

## 6. Employee Performance

| Sum of Sales | Manager ▼ | | | | |
|---|---|---|---|---|---|
| Month ▼ | Chris | Erin | Sam | William | Grand Total |
| ⊞ Jan | 59702.73 | 59185.16 | 73062.34 | 73048.32 | 264998.55 |
| ⊞ Feb | 70282.44 | 69109.8 | 31754.39 | 154355.81 | 325502.44 |
| ⊞ Mar | 62075.31 | 59137.04 | 77874.78 | 66080 | 265167.13 |
| ⊞ Apr | 64058.38 | 189208.29 | 58285.11 | 72477.41 | 384029.19 |
| ⊞ May | 79253.14 | 84858.81 | 50887.39 | 75231.51 | 290230.85 |
| ⊞ Jun | 106073.62 | 123225.09 | 49464.31 | 72833.59 | 351596.61 |
| Grand Total | 441445.62 | 584724.19 | 341328.32 | 514026.64 | 1881524.77 |

Erin was our Best employee for this quarter with maximum sales whereas Sam lagged behind everyone with a huge margin and needs to perform well in the other half of the year.

## 7. Regional Sales

We performed the best in California followed by New York and Texas. We might think of opening a store in those places in future.

## 8. Days to Ship Products

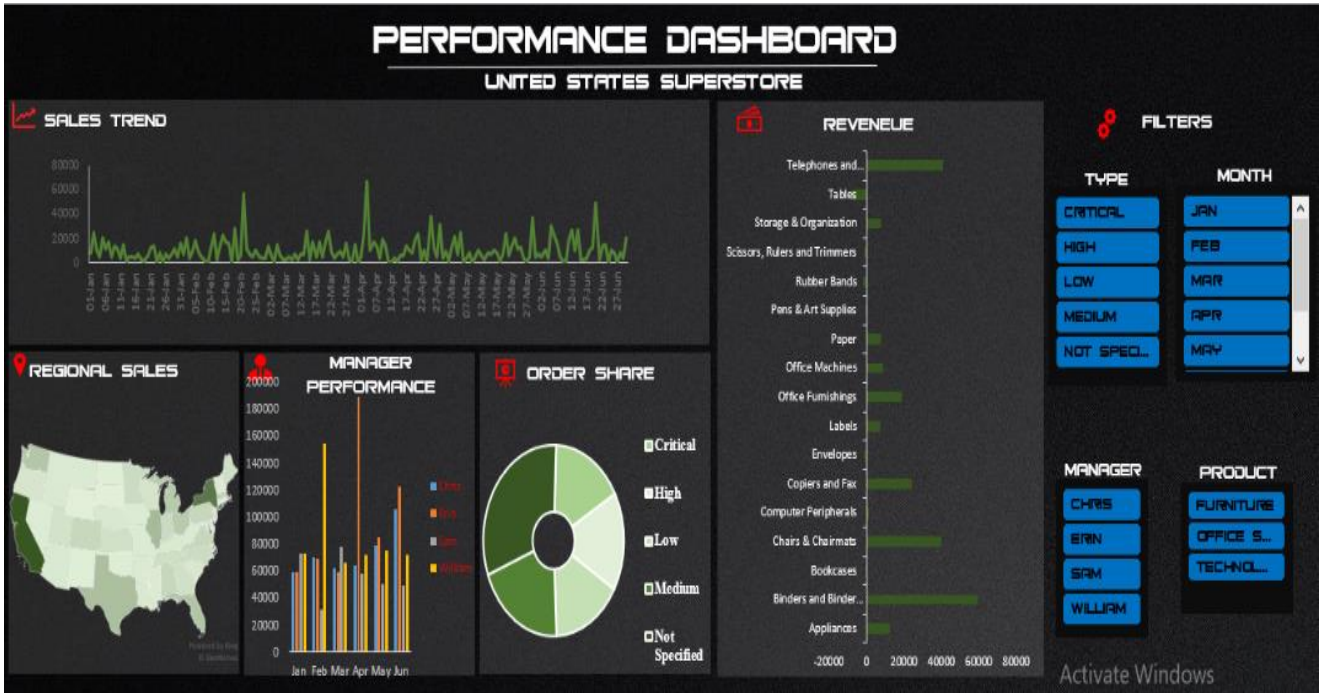| Order Date | Ship Date | DAYS TO SHIP |
|---|---|---|
| 07/01/2015 | 08/01/2015 | 1 |
| 13/06/2015 | 15/06/2015 | 2 |
| 15/02/2015 | 17/02/2015 | 2 |
| 12/05/2015 | 14/05/2015 | 2 |
| 12/05/2015 | 13/05/2015 | 1 |
| 12/05/2015 | 13/05/2015 | 1 |
| 12/05/2015 | 13/05/2015 | 1 |
| 08/04/2015 | 09/04/2015 | 1 |
| 28/05/2015 | 28/05/2015 | 0 |

The products are shipped within 2 days of the order date according to their priority.

## 9. Customer ordering Trend

| Customer ID | 1st Purchased Date | 2nd Purchase Order Date |
|---|---|---|
| 2,391 | 25/05/2015 | 04/06/2015 |
| 210 | 17/01/2015 | 02/06/2015 |
| 2,548 | 04/04/2015 | 21/04/2015 |
| 2,668 | 28/03/2015 | 21/04/2015 |
| 2,840 | 19/03/2015 | 11/06/2015 |
| 181 | 21/02/2015 | 22/05/2015 |
| 1,683 | 14/02/2015 | 14/03/2015 |
| 2,797 | 10/01/2015 | 21/01/2015 |
| 1,341 | 10/02/2015 | 06/06/2015 |
| 92 | 17/05/2015 | 07/06/2015 |
| 1,026 | 07/02/2015 | 29/05/2015 |
| 3,230 | 21/01/2015 | 13/06/2015 |
| 898 | 12/01/2015 | 27/01/2015 |
| 1,671 | 09/02/2015 | 12/05/2015 |
| 1,416 | 06/05/2015 | 25/06/2015 |

It can be seen that the frequency of Customer is quite low and needs to be improved in order to maintain a good profit and growth of the Super Store.

# FINAL DASHBOARD

# REFERENCES AND BIBLIOGRAPHY

- Youtube (https://www.youtube.com/user/MyOnlineTrainingHub)
- Analytics Vidhiya (https://www.analyticsvidhya.com/resources-excel/)
- Kaggle (https://www.kaggle.com/shobanama/superstore)