**DATA MANAGEMNET PROJECT REPORT**

***SALES ANALYSIS OF SUPERSTORE***

Submitted by:

**Roshni Harshad Bhai Parmar**

**Education background = Data Analytics, Software Engineer**

**Data Analysis (Batch September) 2023**

**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to my teachers 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.

**INDEX**

1. Introduction…………………………………………………………………I

2. Scope of the Analysis………………………………………………………………………II

3. Source of dataset………………………………….………………………………………. III

5. Analysis on dataset (for each analysis) ………………………………...…………………VII

1. Introduction…………………………………………………………………...…VII
2. General Description……………………………………………………………. VIII
3. Specific Requirements, Function and Requirements……………………………IX
4. Analysis Result…………………………………….……………………………X
5. Visualization……………………………………………………………………. XI

6. List of Analysis with results……………………………………………………………XVII

7. Final Dashboard…………………………………………………………………………XXI

**INTRODUCTION**

**Data Analysis** is a process of inspecting, cleansing, transforming, and modelling 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 analyse 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 analyse 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:

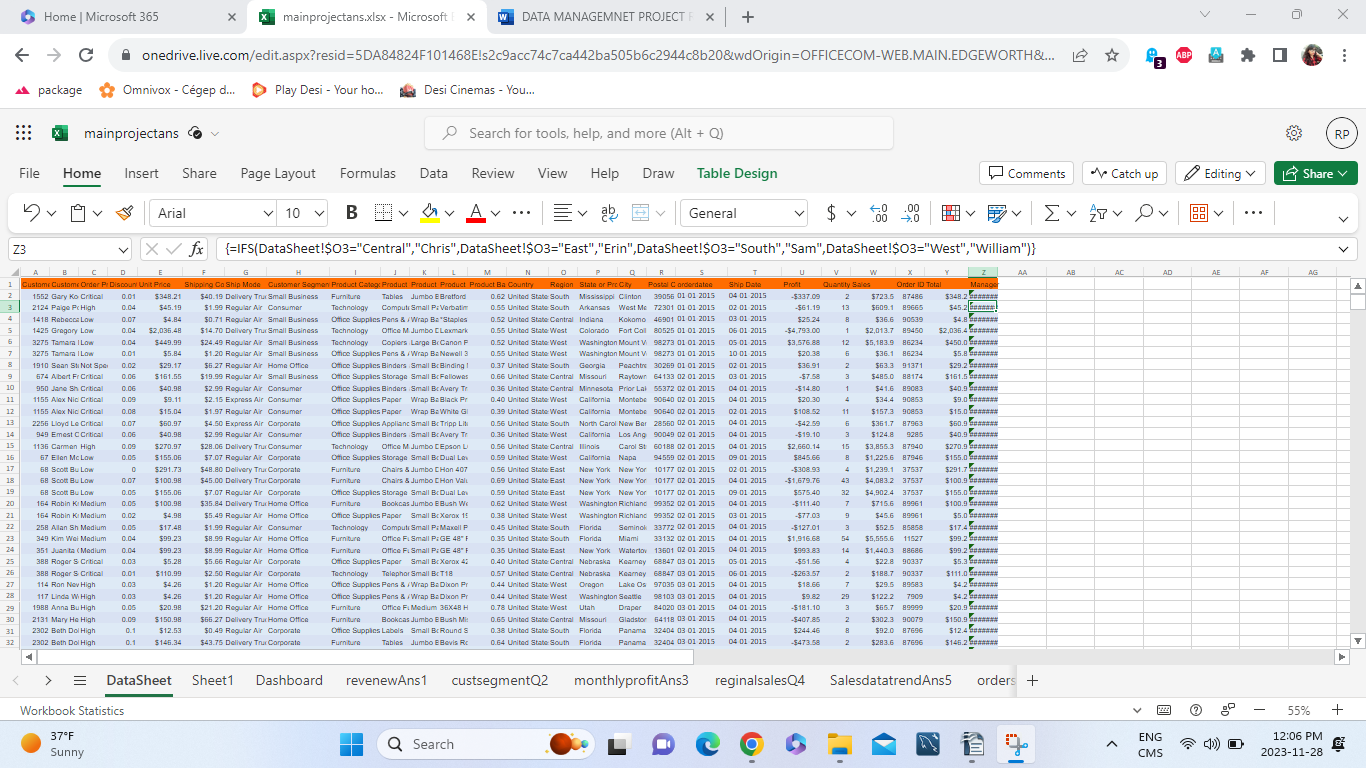
* Order ID – Id of the order created by the customer.
* Order Date and Ship Date – Date when the item was ordered and the date when the item was shipped to the person.
* Shipping Mode – Mode of shipping
* Product Category – Category of the product
* Product – The product that was ordered
* Sales – Quantity 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

**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.



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.**

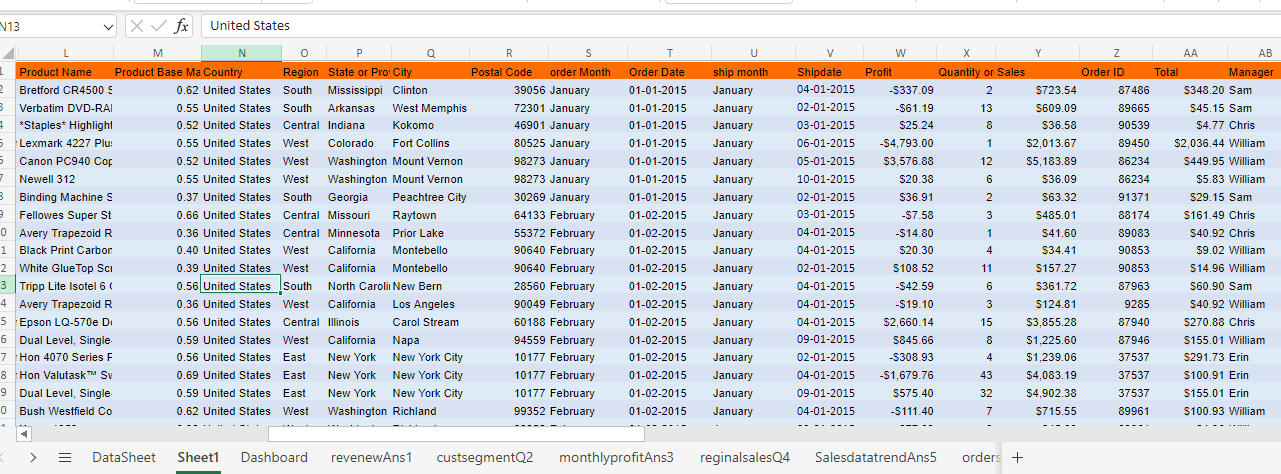
For this, select the whole dataset. Go to Find and Select in the Home tab of excel. Select Go to Special from the drop-down menu and then tick the blank option. All the blank cells will be selected. Then go to Delete option in the home tab again and select Delete Rows from the drop-down menu. This will remove any rows with blank cell

**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.

**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.



**Step 4: 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.

**Step 5: 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.

**SCOPE OF ANALYSIS**

The super store wants to see and analyse 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 analyse 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. Analysing Shipping priority share.

**ANALYSIS OF DATASET**

**Q1 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.

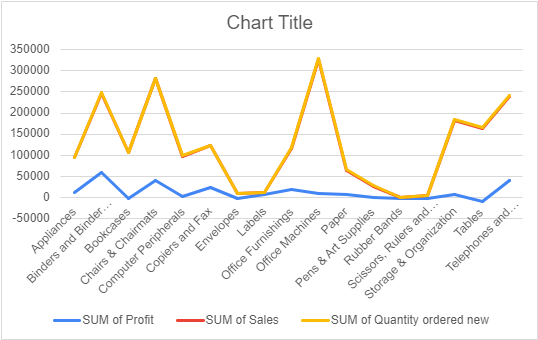
**Specific function and requirements**

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Product Sub-Category | SUM of Profit | SUM of Sales | SUM of Quantity ordered new |
| Appliances | 12594.8206 | 82201.15 | 1187 |
| Binders and Binder Accessories | 59296.38943 | 185928.14 | 2554 |
| Bookcases | -930.4384 | 107796.09 | 536 |
| Chairs & Chairmats | 39808.486 | 241200.42 | 1138 |
| Computer Peripherals | 1698.04398 | 96261.3 | 2481 |
| Copiers and Fax | 23990.20758 | 99069.48 | 181 |
| Envelopes | -1194.4125 | 10479.77 | 590 |
| Labels | 7028.1595 | 4914.82 | 903 |
| Office Furnishings | 18724.1191 | 98070.91 | 2356 |
| Office Machines | 8824.390564 | 318169.68 | 898 |
| Paper | 7769.31507 | 55813.92 | 3481 |
| Pens & Art Supplies | -257.6288 | 26071.61 | 2744 |
| Rubber Bands | -1544.826063 | 1789.43 | 501 |
| Scissors, Rulers and Trimmers | -1291.0959 | 6752.18 | 423 |
| Storage & Organization | 7803.9946 | 175730.05 | 1889 |
| Tables | -8086.639264 | 172511.33 | 882 |
| Telephones and Communication | 40790.51424 | 198764.49 | 2358 |
| Grand Total | 215023.3997 | 1881524.77 | 25102 |

**Visualization:**

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



Q2 **Segment Distribution of each product category:**

**Description:**

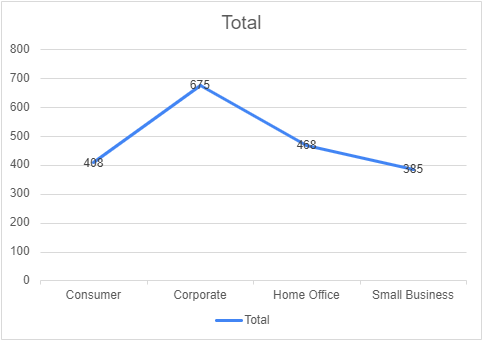
By knowing which segment of sales has themost number of sales and which has least we can identify factors which affect the sales and thereby improve our strategy of

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.

|  |  |
| --- | --- |
| Customer Segment | COUNTA of Product Category |
| Consumer | 408 |
| Corporate | 675 |
| Home Office | 468 |
| Small Business | 385 |
| Grand Total | 1936 |



Segment Distribution Visulization

**Q3 Sales throughout**  **months of a sales year.**

**Description:**

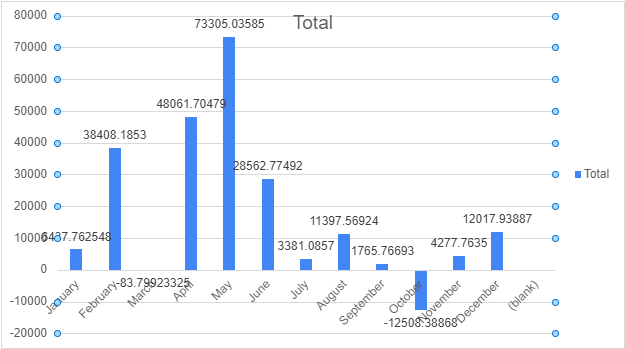
Monthly sales can help us identify which month is more profitable and helps identify the factor which helps us to do so. We can apply the identified the factors in other months to increase the sales.

**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.

|  |  |
| --- | --- |
| order Month | Sum of Profit |
| January | 6437.762548 |
| February | 38408.1853 |
| March | -83.79923325 |
| April | 48061.70479 |
| May | 73305.03585 |
| June | 28562.77492 |
| July | 3381.0857 |
| August | 11397.56924 |
| September | 1765.76693 |
| October | -12508.38868 |
| November | 4277.7635 |
| December | 12017.93887 |
| (blank) |  |
| Grand Total | 215023.3997 |

Visitation Result :



Q4 **Sales and their priority State vise**

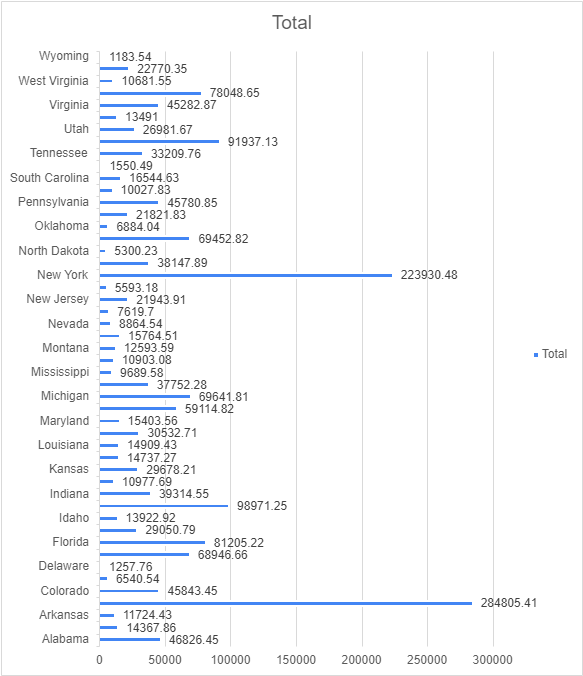
**Description:**

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.

|  |  |
| --- | --- |
| 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 |
| Kansas | 29678.21 |
| Kentucky | 14737.27 |
| Louisiana | 14909.43 |
| Maine | 30532.71 |
| Maryland | 15403.56 |
| Massachusetts | 59114.82 |
| Michigan | 69641.81 |
| Minnesota | 37752.28 |
| Mississippi | 9689.58 |
| Missouri | 10903.08 |
| Montana | 12593.59 |
| Nebraska | 15764.51 |
| Nevada | 8864.54 |
| New Hampshire | 7619.7 |
| New Jersey | 21943.91 |
| New Mexico | 5593.18 |
| New York | 223930.48 |
| North Carolina | 38147.89 |
| North Dakota | 5300.23 |
| Ohio | 69452.82 |
| Oklahoma | 6884.04 |
| Oregon | 21821.83 |
| Pennsylvania | 45780.85 |
| Rhode Island | 10027.83 |
| South Carolina | 16544.63 |
| South Dakota | 1550.49 |
| Tennessee | 33209.76 |
| Texas | 91937.13 |
| Utah | 26981.67 |
| Vermont | 13491 |
| Virginia | 45282.87 |
| Washington | 78048.65 |
| West Virginia | 10681.55 |
| Wisconsin | 22770.35 |
| Wyoming | 1183.54 |
| Grand Total | 1881524.77 |



Q5 **Comparison of sales data trend Datevise**

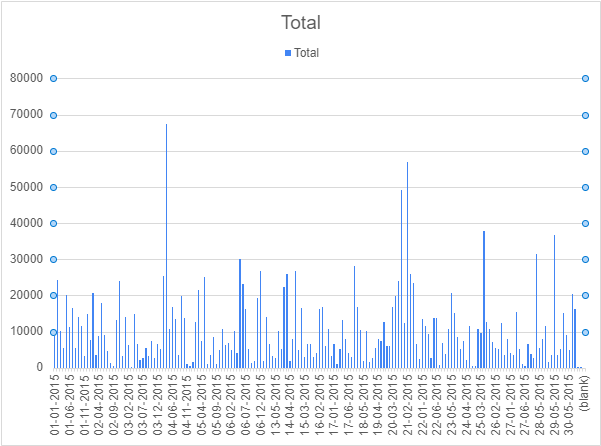
**Description:**

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

|  |  |
| --- | --- |
| Order Date | Sum of Sales |
| 01-01-2015 | 8666.18 |
| 01-02-2015 | 24319.92 |
| 01-03-2015 | 10134.42 |
| 01-04-2015 | 5470.58 |
| 01-05-2015 | 20160.51 |
| 01-06-2015 | 11123.41 |
| 01-07-2015 | 16551.82 |
| 01-08-2015 | 5388.84 |
| 01-09-2015 | 13926.83 |
| 01-10-2015 | 11520.37 |
| 01-11-2015 | 3355.12 |
| 01-12-2015 | 14882.67 |
| 02-01-2015 | 7529.03 |
| 02-02-2015 | 20691.02 |
| 02-03-2015 | 3485.71 |
| 02-04-2015 | 8893.91 |
| 02-05-2015 | 17906.15 |
| 02-06-2015 | 9149.27 |
| 02-07-2015 | 4720.29 |
| 02-08-2015 | 1404.04 |
| 02-09-2015 | 570.14 |
| 02-10-2015 | 13160.58 |
| 02-11-2015 | 23856.93 |
| 02-12-2015 | 3233.9 |
| 03-01-2015 | 14138.97 |
| 03-02-2015 | 6407.95 |
| 03-03-2015 | 262.78 |
| 03-04-2015 | 14738.71 |
| 03-05-2015 | 6541.13 |
| 03-06-2015 | 2218.25 |
| 03-07-2015 | 2711.04 |
| 03-08-2015 | 5373.52 |
| 03-09-2015 | 3255.53 |
| 03-10-2015 | 7421.79 |
| 03-11-2015 | 2619.37 |
| 03-12-2015 | 6679.66 |
| 04-01-2015 | 5203.52 |
| 04-02-2015 | 25267.5 |
| 04-04-2015 | 67267.77 |
| 04-05-2015 | 10775.67 |
| 04-06-2015 | 16890.22 |
| 04-07-2015 | 13355.85 |
| 04-08-2015 | 3637.02 |
| 04-09-2015 | 19709.51 |
| 04-10-2015 | 13636.13 |
| 04-11-2015 | 954.94 |
| 04-12-2015 | 490 |
| 05-01-2015 | 1701.6 |
| 05-02-2015 | 12530.77 |
| 05-03-2015 | 21562.16 |
| 05-04-2015 | 7375.22 |
| 05-05-2015 | 25136.13 |
| 05-06-2015 | 992.07 |
| 05-07-2015 | 3652.49 |
| 05-08-2015 | 8591.78 |
| 05-09-2015 | 909.09 |
| 05-10-2015 | 4957.99 |
| 05-11-2015 | 10693.7 |
| 05-12-2015 | 6177.17 |
| 06-01-2015 | 6929.82 |
| 06-02-2015 | 4790.69 |
| 06-03-2015 | 10096.13 |
| 06-04-2015 | 4178.57 |
| 06-05-2015 | 30105.89 |
| 06-06-2015 | 23121.02 |
| 06-07-2015 | 16260.81 |
| 06-08-2015 | 5137.93 |
| 06-09-2015 | 1229.07 |
| 06-10-2015 | 1807.72 |
| 06-11-2015 | 19228.73 |
| 06-12-2015 | 26689.09 |
| 13-01-2015 | 1762.81 |
| 13-02-2015 | 14090.4 |
| 13-03-2015 | 6648.49 |
| 13-04-2015 | 3318.93 |
| 13-05-2015 | 2629.94 |
| 13-06-2015 | 10284.63 |
| 14-01-2015 | 5200.19 |
| 14-02-2015 | 22401.48 |
| 14-03-2015 | 25783.09 |
| 14-04-2015 | 1809.72 |
| 14-05-2015 | 7935.95 |
| 14-06-2015 | 26719.23 |
| 15-01-2015 | 4829.95 |
| 15-02-2015 | 16533.18 |
| 15-03-2015 | 2943.82 |
| 15-04-2015 | 6553.25 |
| 15-05-2015 | 6688.99 |
| 15-06-2015 | 3011.6 |
| 16-01-2015 | 3964.13 |
| 16-02-2015 | 16120.57 |
| 16-03-2015 | 16728.53 |
| 16-04-2015 | 5886.63 |
| 16-05-2015 | 10835.31 |
| 16-06-2015 | 3219.89 |
| 17-01-2015 | 6667.69 |
| 17-02-2015 | 1076.39 |
| 17-03-2015 | 5244.52 |
| 17-04-2015 | 13273.29 |
| 17-05-2015 | 7930.9 |
| 17-06-2015 | 3968.53 |
| 18-01-2015 | 2987.38 |
| 18-02-2015 | 28264.33 |
| 18-03-2015 | 16840.65 |
| 18-04-2015 | 10317.05 |
| 18-05-2015 | 1916.8 |
| 18-06-2015 | 10185.36 |
| 19-01-2015 | 1652.89 |
| 19-02-2015 | 2676.23 |
| 19-03-2015 | 5402.48 |
| 19-04-2015 | 7818.3 |
| 19-05-2015 | 7445.03 |
| 19-06-2015 | 12684.68 |
| 20-01-2015 | 6109.01 |
| 20-02-2015 | 5978.24 |
| 20-03-2015 | 16675.58 |
| 20-04-2015 | 19795.49 |
| 20-05-2015 | 24061.33 |
| 20-06-2015 | 49028.22 |
| 21-01-2015 | 12281.51 |
| 21-02-2015 | 56764.54 |
| 21-03-2015 | 25857.3 |
| 21-04-2015 | 23413.18 |
| 21-05-2015 | 6585.4 |
| 21-06-2015 | 2367.99 |
| 22-01-2015 | 13513.64 |
| 22-02-2015 | 11464.77 |
| 22-03-2015 | 9403.56 |
| 22-04-2015 | 2565.9 |
| 22-05-2015 | 13664.44 |
| 22-06-2015 | 13650.95 |
| 23-01-2015 | 765.9 |
| 23-02-2015 | 6812.64 |
| 23-03-2015 | 3850.97 |
| 23-04-2015 | 10621.21 |
| 23-05-2015 | 20716.77 |
| 23-06-2015 | 15196.84 |
| 24-01-2015 | 8376.58 |
| 24-02-2015 | 5160.96 |
| 24-03-2015 | 7449.45 |
| 24-04-2015 | 2152.53 |
| 24-05-2015 | 11562.93 |
| 24-06-2015 | 572.28 |
| 25-01-2015 | 445.25 |
| 25-02-2015 | 10752.13 |
| 25-03-2015 | 9686.07 |
| 25-04-2015 | 37768.2 |
| 25-05-2015 | 12568.22 |
| 25-06-2015 | 10768.77 |
| 26-01-2015 | 7035.73 |
| 26-02-2015 | 5425.5 |
| 26-03-2015 | 5237.6 |
| 26-04-2015 | 12354.36 |
| 26-05-2015 | 3439.78 |
| 26-06-2015 | 7888.86 |
| 27-01-2015 | 4036.39 |
| 27-02-2015 | 3537.5 |
| 27-03-2015 | 15468.49 |
| 27-04-2015 | 5296.63 |
| 27-05-2015 | 1071.55 |
| 27-06-2015 | 338.41 |
| 28-01-2015 | 6617.65 |
| 28-02-2015 | 3842.61 |
| 28-03-2015 | 2783.93 |
| 28-04-2015 | 31319 |
| 28-05-2015 | 5378.71 |
| 28-06-2015 | 7990.09 |
| 29-01-2015 | 11666.61 |
| 29-03-2015 | 1516.49 |
| 29-04-2015 | 3561.59 |
| 29-05-2015 | 36735.8 |
| 29-06-2015 | 3635.91 |
| 30-01-2015 | 5276.23 |
| 30-03-2015 | 15175.7 |
| 30-04-2015 | 9015.8 |
| 30-05-2015 | 4782.83 |
| 30-06-2015 | 20508.9 |
| 31-01-2015 | 16299.11 |
| 31-03-2015 | 101.71 |
| 13/01/2015 | 9.23 |
| (blank) |  |
| Grand Total | 1881524.77 |



Q6 **Sales and their priority**

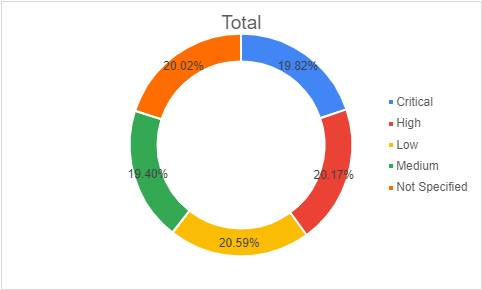
**Description:**

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.

|  |  |
| --- | --- |
| Order Priority | SUM of Product Base Margin |
| Critical | 19.82% |
| High | 20.17% |
| Low | 20.59% |
| Medium | 19.40% |
| Not Specified | 20.02% |
| Grand Total | 100.00% |



Q7 **Comparison of sales of each product category**

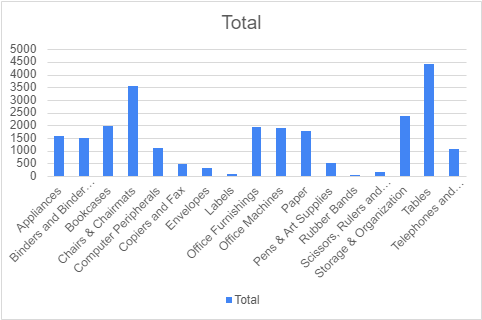
**Description:**

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.

|  |  |
| --- | --- |
| Product Sub-Category | SUM of Shipping Cost |
| 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 |



Q8 **Employee Performance**

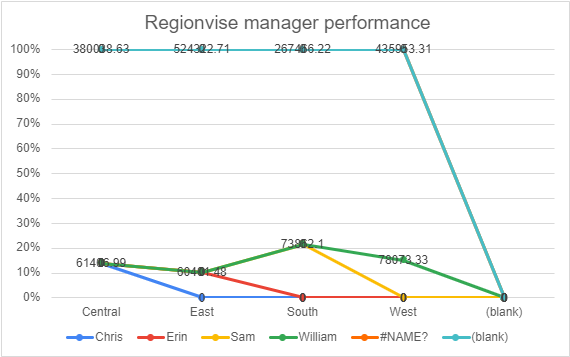
**Description:**

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.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sum of Sales | Manager |  |  |  |  |  |  |
| Region | Chris | Erin | Sam | William | #NAME? | (blank) | Grand Total |
| Central | 61406.99 |  |  |  | 380038.63 |  | 441445.62 |
| East |  | 60401.48 |  |  | 524322.71 |  | 584724.19 |
| South |  |  | 73862.1 |  | 267466.22 |  | 341328.32 |
| West |  |  |  | 78073.33 | 435953.31 |  | 514026.64 |
| (blank) |  |  |  |  |  |  |  |
| Grand Total | 61406.99 | 60401.48 | 73862.1 | 78073.33 | 1607780.87 |  | 1881524.77 |



**Q9** **Comparing Sales Analysis**

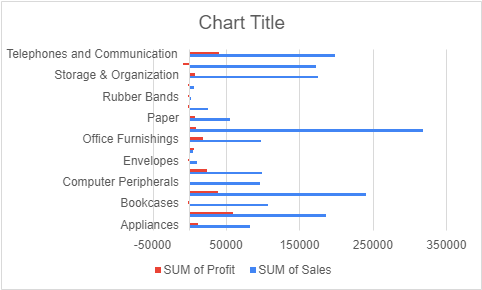
**Description:**

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.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Product Sub-Category | SUM of Sales | SUM of Profit |
| Appliances | 82201.15 | 12594.8206 |
| Binders and Binder Accessories | 185928.14 | 59296.38943 |
| Bookcases | 107796.09 | -930.4384 |
| Chairs & Chairmats | 241200.42 | 39808.486 |
| Computer Peripherals | 96261.3 | 1698.04398 |
| Copiers and Fax | 99069.48 | 23990.20758 |
| Envelopes | 10479.77 | -1194.4125 |
| Labels | 4914.82 | 7028.1595 |
| Office Furnishings | 98070.91 | 18724.1191 |
| Office Machines | 318169.68 | 8824.390564 |
| Paper | 55813.92 | 7769.31507 |
| Pens & Art Supplies | 26071.61 | -257.6288 |
| Rubber Bands | 1789.43 | -1544.826063 |
| Scissors, Rulers and Trimmers | 6752.18 | -1291.0959 |
| Storage & Organization | 175730.05 | 7803.9946 |
| Tables | 172511.33 | -8086.639264 |
| Telephones and Communication | 198764.49 | 40790.51424 |
| Grand Total | 1881524.77 | 215023.3997 |



**Q10 Sales ship mode and their priority**

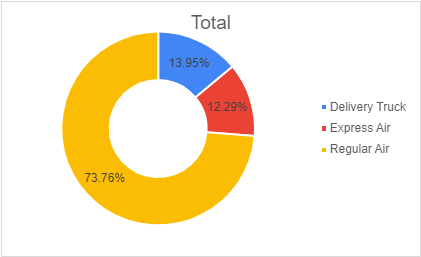
**Description:**

In this we analyse which particular region is having most number of sales and which is having priority. Furthermore, we can look upon the factors which might be impacting the mode, 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.

|  |  |
| --- | --- |
| Ship Mode | COUNTA of Order Priority |
| Delivery Truck | 13.95% |
| Express Air | 12.29% |
| Regular Air | 73.76% |
| Grand Total | 100.00% |



**FINAL DASHBOARD**