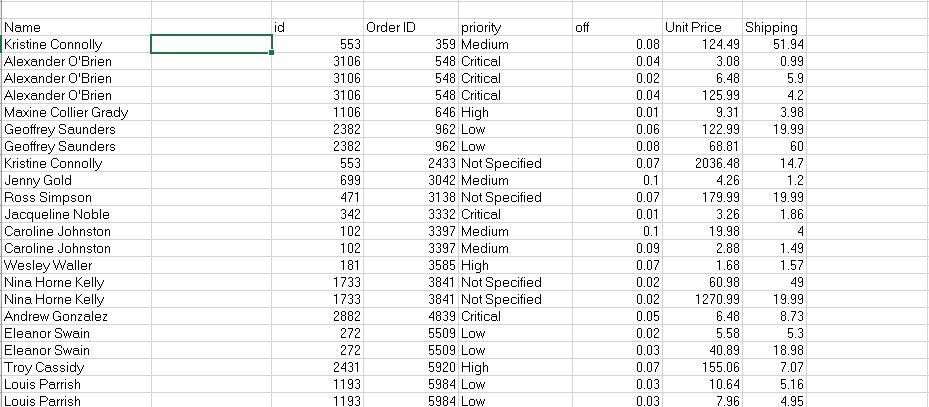
**ETL PROCESS**

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.**



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

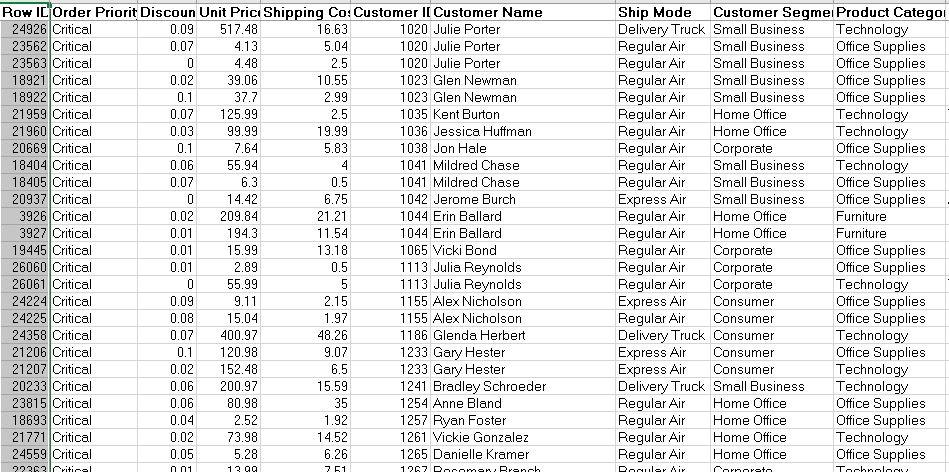
**Step 1: Removing the blank cells from the dataset.**

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.



**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.



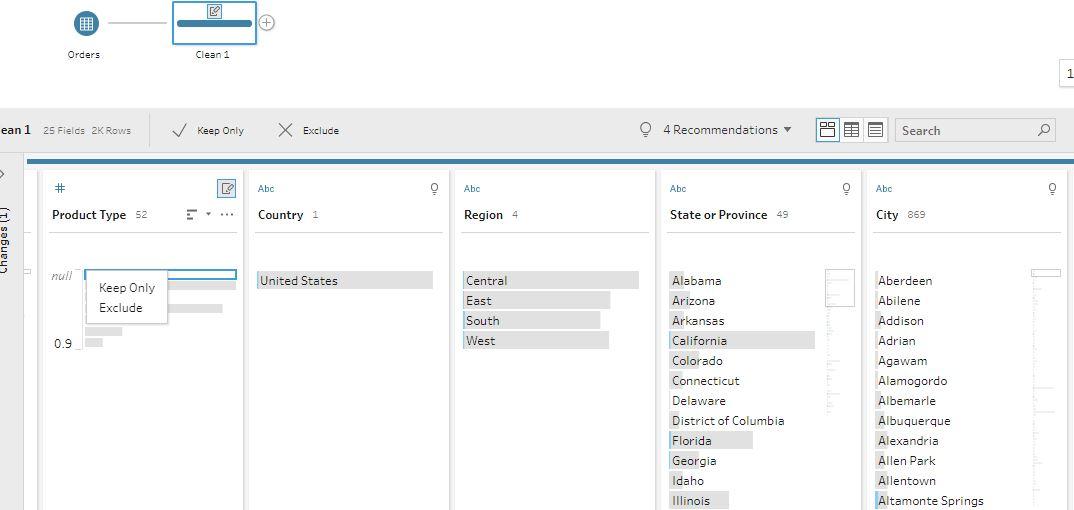
**Step 3: Giving proper and appropriate column names.**

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: 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.



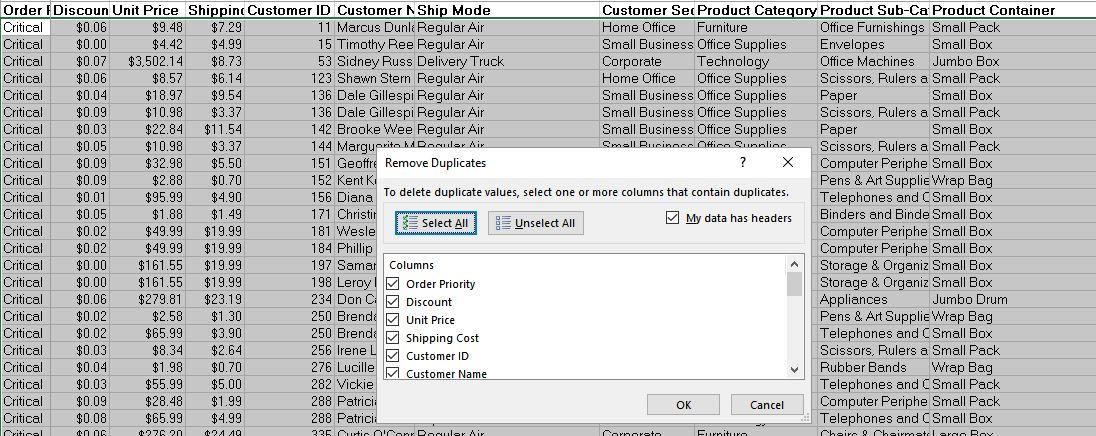
**Step 5: Improvising Proper Data Formatting**

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.



**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.



**ANALYSIS OF DATASET**

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

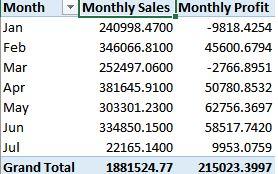
**Description:**

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**

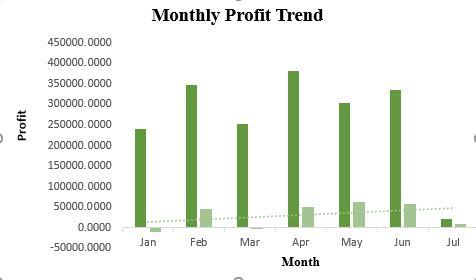
We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

**Results:**



**Visualization:**

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



1. **Segment Distribution of each product category:**

**Description:**

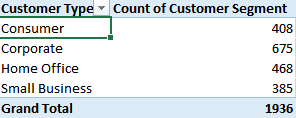
By knowing which segment of sales has themost number of sales and which has least we can identify factors which affect the sales and thereby improve our strategy of

making sales.

**Specific function and requirements**

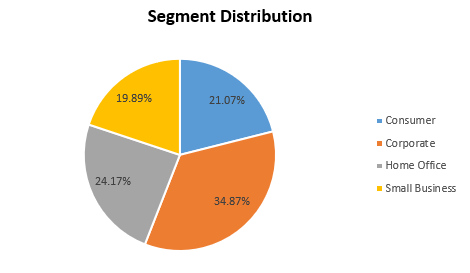
We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

**Results:**



**Visualization:**

We will use a pie chart to visualize the distribution.



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

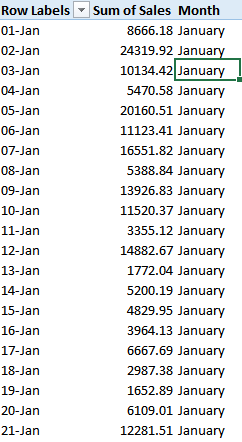
**Description:**

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:**

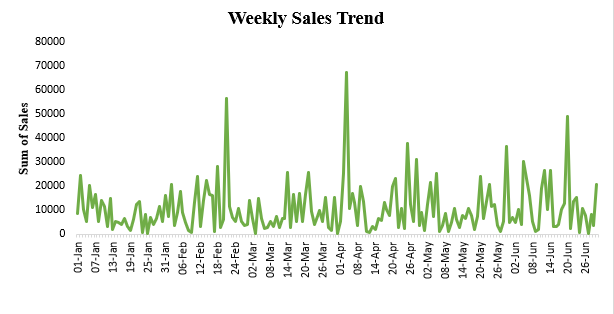
We have to create a pivot table. No specific functions are used. We then put the months and sum of sales in the columns.

**Results:**



**Visualization:**

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



1. **Sales and their priority**

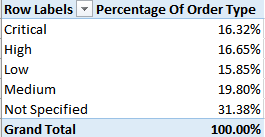
**Description:**

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**

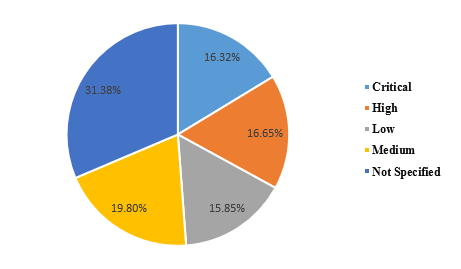
We have to create a pivot table. No specific functions are used. We then put the priority c and count of their respective sales in the columns of the pivot table.

**Results:**



**Visualization:**

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



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

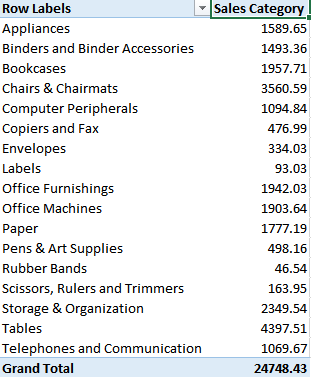
**Description:**

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:**

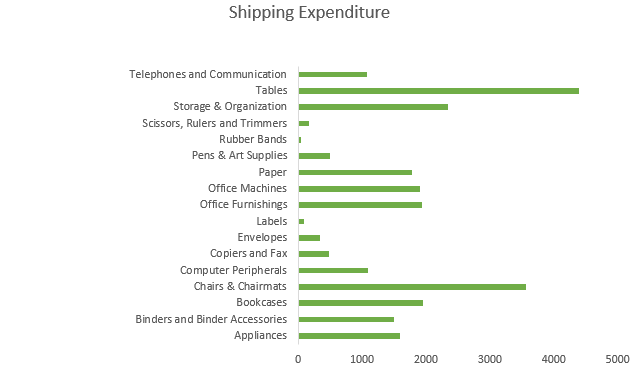
We have to create a pivot table. No specific functions are used. Product category is used as columns with summation of profit and sales of each product category.

**Results:**



**Visualization:**

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



1. **Employee Performance**

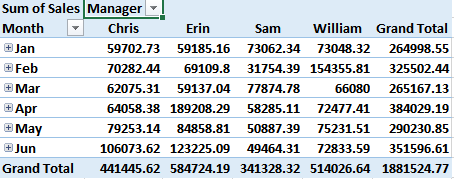
**Description:**

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,

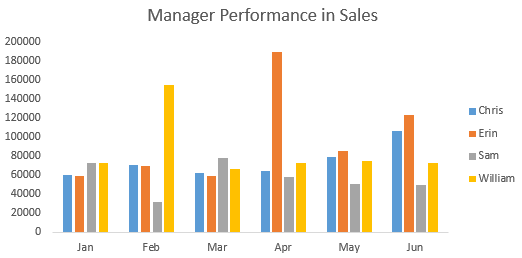
**Specific function and requirements:**

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

**Results:**



**Visualization:**



1. **Regional Sales Analysis**

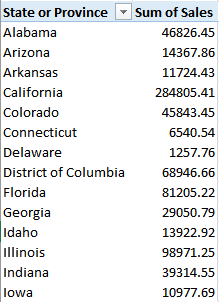
**Description:**

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.

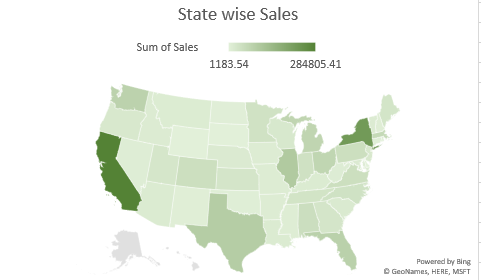
**Specific function and requirements:**

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

**Results:**



**Visualization:**



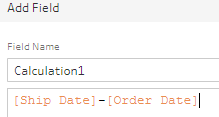
1. **Days to Ship a Product**

**Description:**

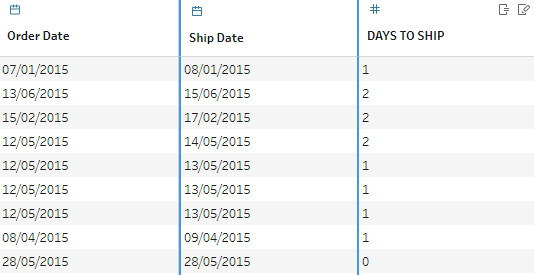
In this we analyze how much time the super store is taking to ship a product after successful placement of order by the user. This can help us to improve the customer service and improve the service quality provided to the customer.

**Specific function and requirements:**

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



**Results:**



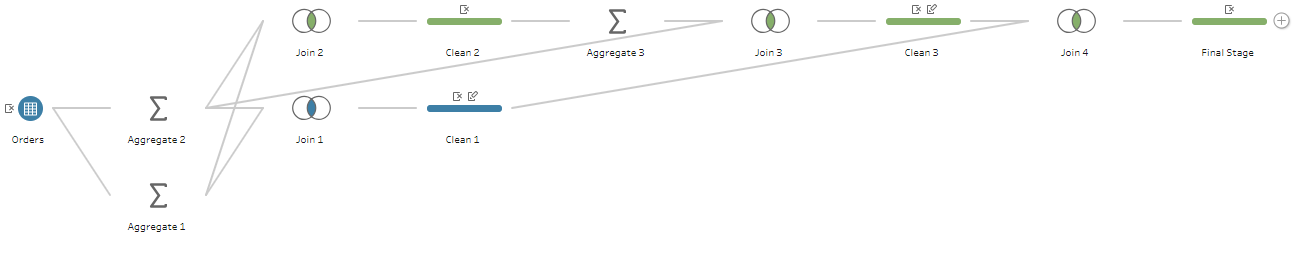
1. **Customer Ordering Trend**

**Description:**

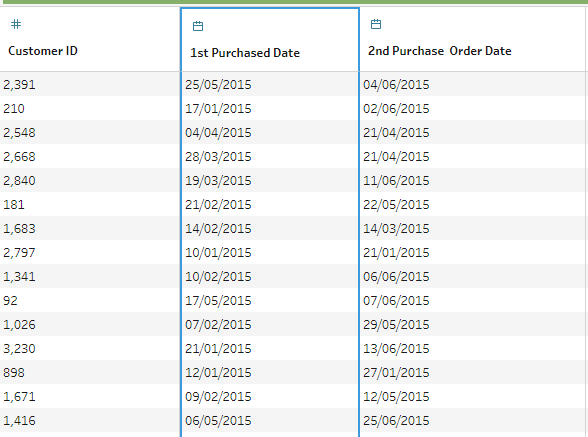
In this we analyze after how much time does a particular revisits us and places the order again. We can create offers and Discounts accordingly and increase the customer engagement for frequent visits. This will ultimately help us to improve the customer service and improve the service quality provided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

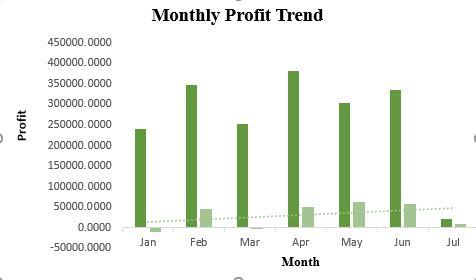


**Results:**



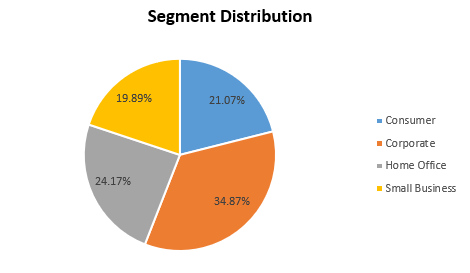
**ANALYSIS RESULTS**

1. **Monthly Profit Trend**



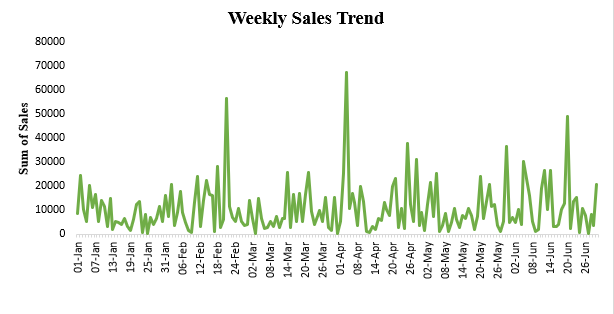
Sales was high in the Jan but still resulted in negative profit i.e. loss. Still the super store managed to work well and increase the profit exponentially by the end of Jun.

1. **Segment Distribution of each product category**



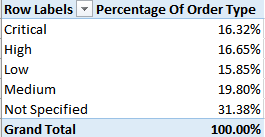
It is clear that Corporate Sector is our valuable customer followed by Home Office. We can offer them special discounts and can have tie ups to increase the engagement. Further, Small Business is lagging in our partnership and we can offer them new offers for more sales.

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

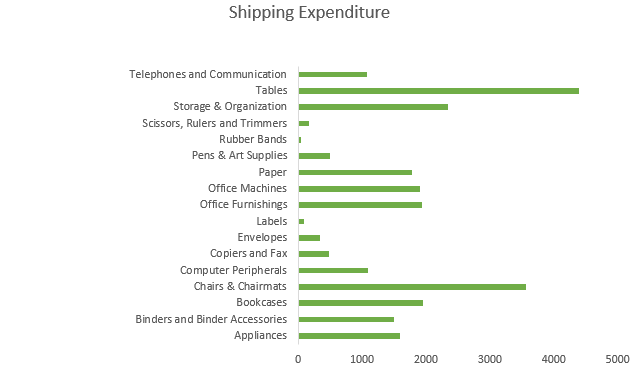


Sales were at peak once in mid Feb and again in Starting of the April followed by June End. We can create offers for other times as well to increase the sales growth.

1. **Sales and their priority**

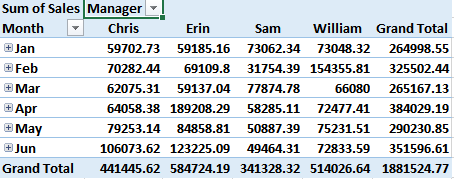


We are having more of the orders without any priority followed by medium priority. We can enhance this by introducing faster services ultimately increasing the revenue.

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

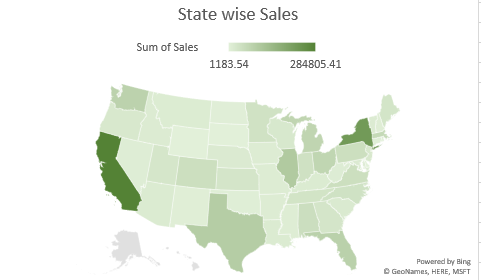
It is clear that tables are our best selling products followed by chairs and chair mats. We can work upon the one’s not performing well to increase their sales also.

1. **Employee Performance**



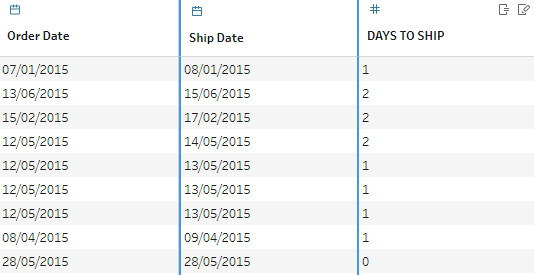
Erin was our Best employee for this quarter with maximum sales whereas Sam lagged behind everyone with a huge margin and needs to perform well in the other half of the year.

1. **Regional Sales**



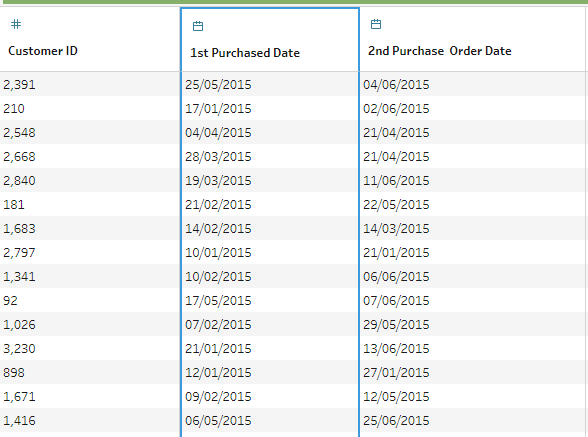
We performed the best in California followed by New York and Texas. We might think of opening a store in those places in future.

1. **Days to Ship Products**



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

1. **Customer ordering Trend**



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**

