

School of System and Enterprise
EM:624 Informatics for Engineering Management
Wall-Mart Data Analysis
Submitted by: Devansh Modi (10411940)
Guided by: Dr. Carlo Lipizzi

Overview of Research Question

Imagine you being appointed as a CEO of a big Consumer retail outlet and every time you take decisions for the company based on the data available to you. But, some questions still remain to be dealt with such as, what is the data about? What inference can be made? What are the key important parameters affecting or attracting the customers? The answers to all those questions are quite easy if one uses proper techniques and set target to particular parameters affecting customers.

In the following project I am trying to do a Data Analysis on a big Consumer Retail outlet “Wall Mart”. By examining the Data set available, the project would do analysis on certain parameters mentioned below

- To analyze which Region has the highest number of sales based on that it would give us a specific market of interest where we can focus more and make them satisfied which in turn improve our sales.
- To analyze which category placed most orders this would help us keeping products in those category sufficient so at any time we don't run out products and it would also prevent us to stock up unnecessary products which could fetch loss to the company.
- To analyze which State has the highest number of sales so that we can know our market accordingly and provide necessary service to keep those customers satisfied.
- To analyze which City places highest order which can help us to improve our market and give them offers and discount at regular intervals to keep them busy buying our products.
- To analyze which Segment places more order so that we can improve our products inventory accordingly.
- To analyze which manufacturer has the highest number of sales by this we would get to know that customers are more satisfied with what kind of Manufacturer!
- To analyze out the time required to ship the order according to various States so that we can try to deliver the product without keeping customers waiting.
- To analyze which are the sub categories that customer had ordered more this analysis would help us determine the daily needs of customer and if you keep those things stocked you would never run out of business and can keep your customers with you.

Data Description

URL: Sample Data from Tableau- <https://community.tableau.com/docs/DOC-1236>

SOURCE OF DATA: Walmart multinational retail corporation

DATASET:

The Data is collection of 17 variables and there are total 9994 records means the Data base contains records of 9994 customers which is framed by the date when they have place the order, Category of the product being ordered and other parameters.

- Category of the product such as Furniture, office supplies, Technology, etc.
- City is from the where the customer belongs and we need to ship the parcel to that particular city
- Order ID the reference ID of the order which had been placed by the customer. It is given for future aspects such as Returns, Tracking shipment, etc.
- Order Date when the order was placed by the customer.
- Ship Date when it was out from the distribution Center to be delivered
- State is referred as in which state the shipment is to be delivered
- Region is referred as from which part of country the order is placed like for example North, South, North east, etc.
- Discount is referred as up to how much discount the customer got when he placed the order
- Profit is the amount which is earned per order.
- Sales is the value of product per order which the customer paid after getting discount.
- Ship mode refers to the class of mail in which shipment was sent Ex. First Class, Second Class, etc.
- Product name defines the name of the product ordered by the customer.
- Segment is divided into Consumer, Office supplies and Home
- Postal code refers to particular area in that particular city from where the customer belongs.
- Country is from where the customer belongs to.
- Manufacturer is the parent company of the product being ordered who made it.
- Discount is referred as how much discount does the customer got on that particular order.
- Profit ratio is what percent of profit per sale you got for that particular order.
- Quantity is referred as how of those product the customer ordered.

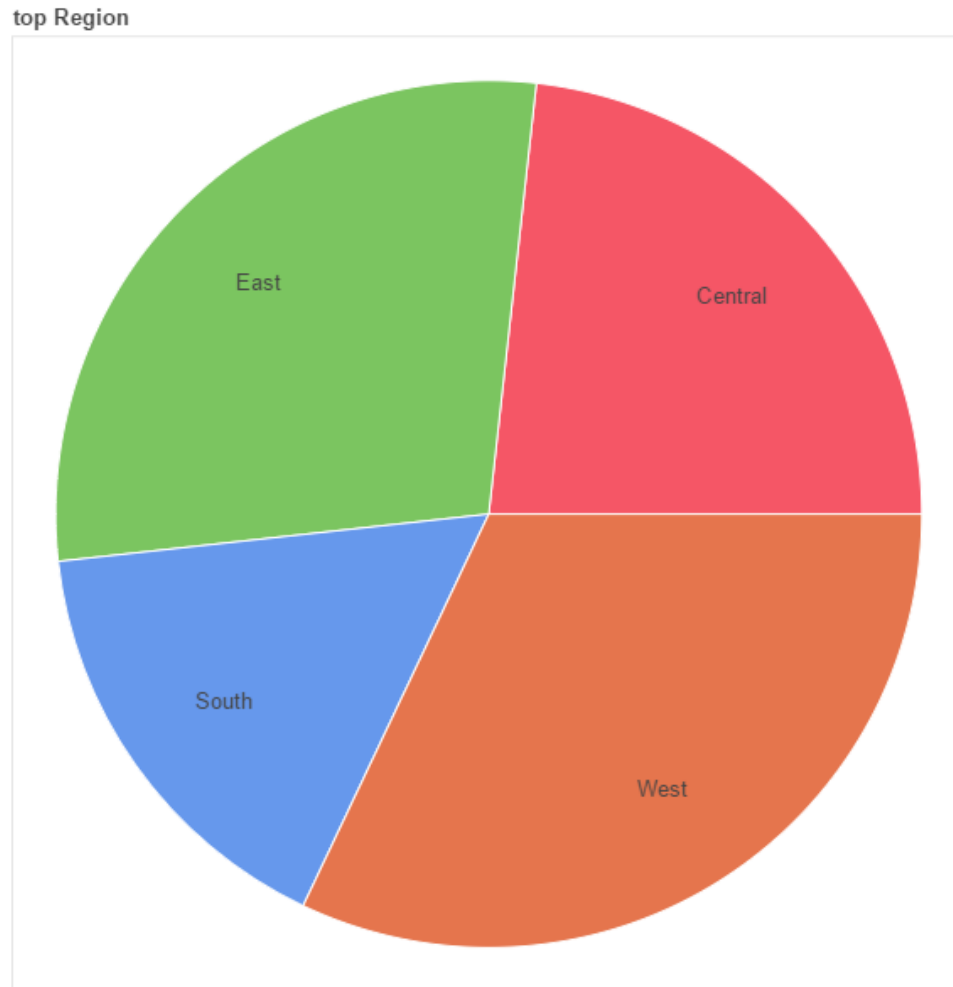
Data Preparation

- 1) The data was read from the file and was transformed as Data Frame.
- 2) The following variables were ignored as they were not use for analysis
 - Country
 - Customer name
 - Order-ID
 - Postal Code
 - Product Name
 - Ship Mode
 - Number
 - Discount
- 3) There was not that much of cleaning the data as it was simple and easy to handle
- 4) Added one column "Duration" which is the difference between the order date and shipped date.

Methodology

- 1) First using pandas, I read the .csv file and named the variable as w1
- 2) Made the list for the following variables
"City"," State"," Segment"," Category"," Subcategory"," region"," manufacturer"
- 3) Then imported Counter and looked for the most common for all the above list which gave me the count of which Top_15 cities which placed more order and the same for all the other parameters.
- 4) Then at last I plotted the graph for all the list mentioned and analyzed what it meant as well as it gave me options what can be done to increase sales and make customers satisfied.
- 5) I faced a lot of problem during plotting because the list was large and every time the error I faced was with respect to undefined label. Due to which it always went out of range.

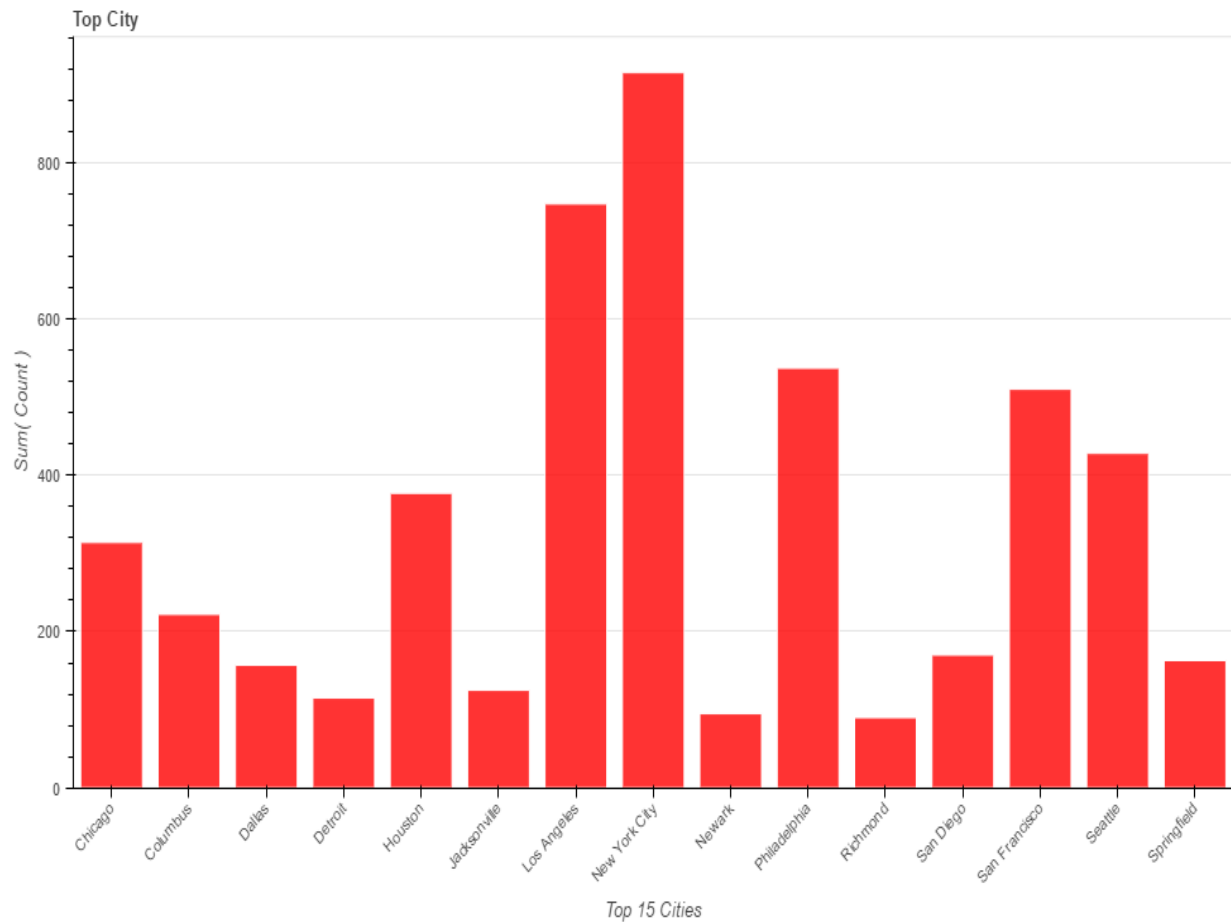
Results



[Figure 1 – Region having highest orders]

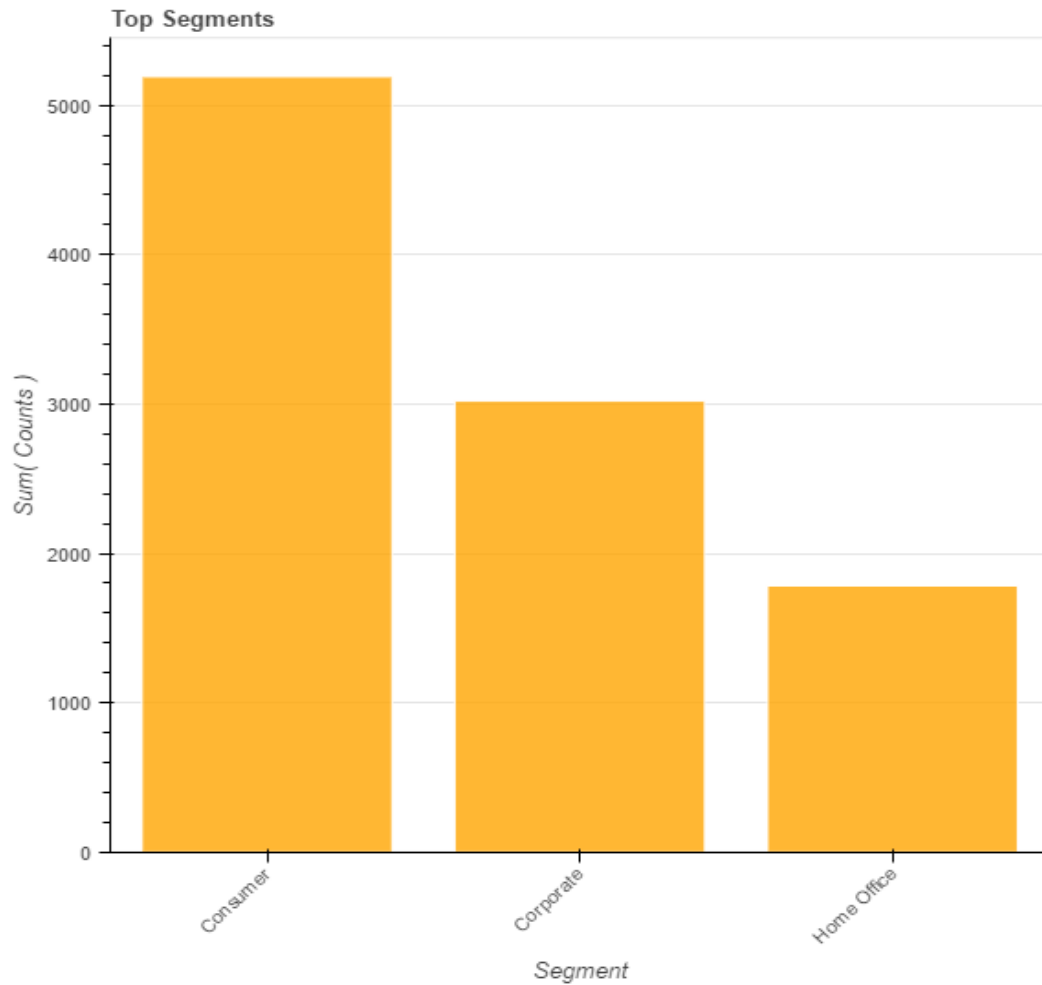
The above graph represents Pie chart of the Regions from where more order was placed. From the chart it can be shown that Store is getting more orders from the West followed by the East region after that we can see the Central and the south region.

Now I know that west region places highest number of order or have highest number of sales but we can't predict the market based on region. So I tried to analyze sales for the top 15 Cities which is shown in the next figure.



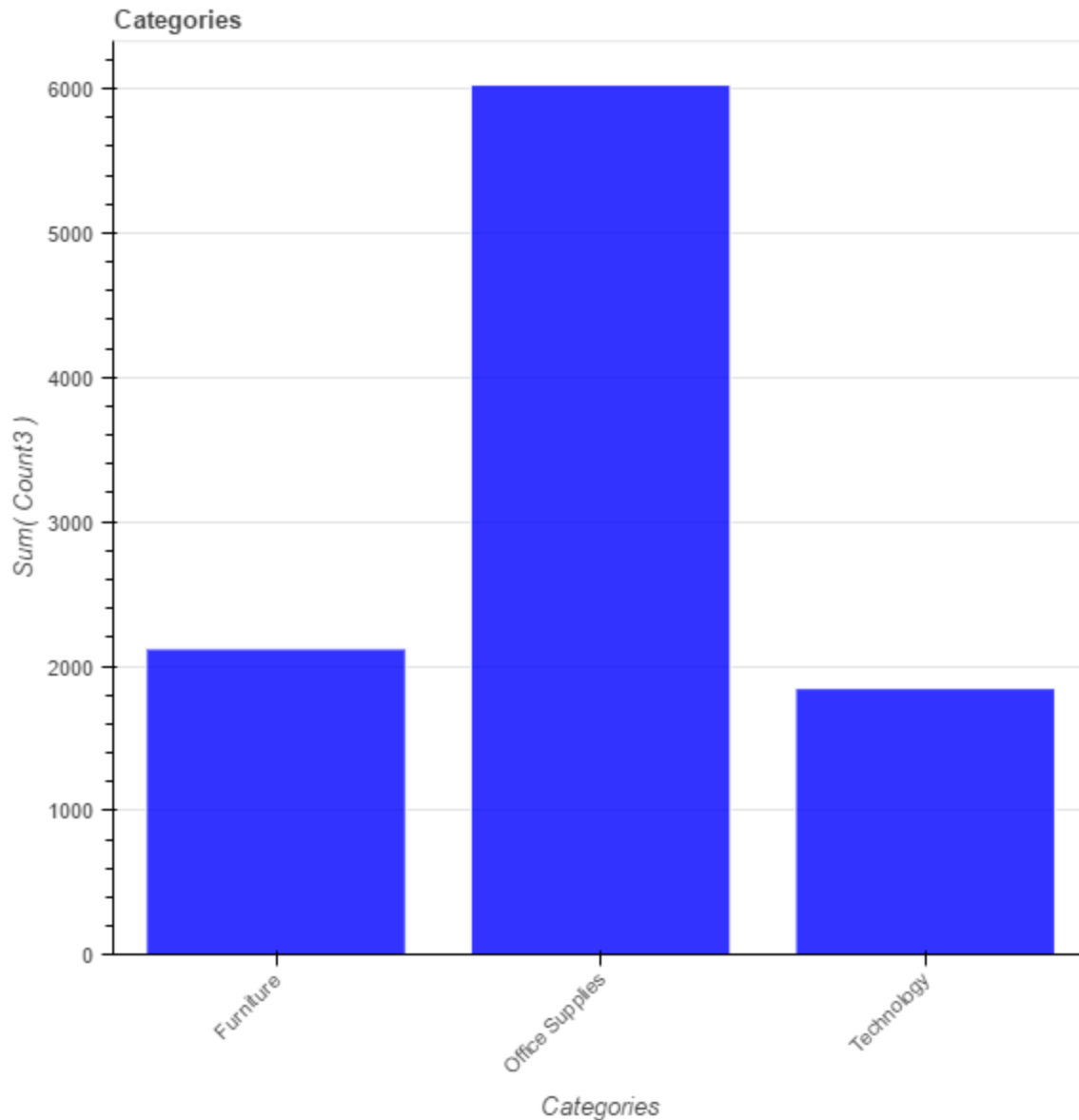
[Figure- Top 15 cities]

The above figure represents the Top 15 cities which plays a major role in increasing the sales of the company. If we compare the figure 1 and figure 2 we can see the difference that thought the West region places more order but if we differentiate sales according to City we can find that New York City which is located in the east contribute more to sales followed by Los Angeles situated in the West.



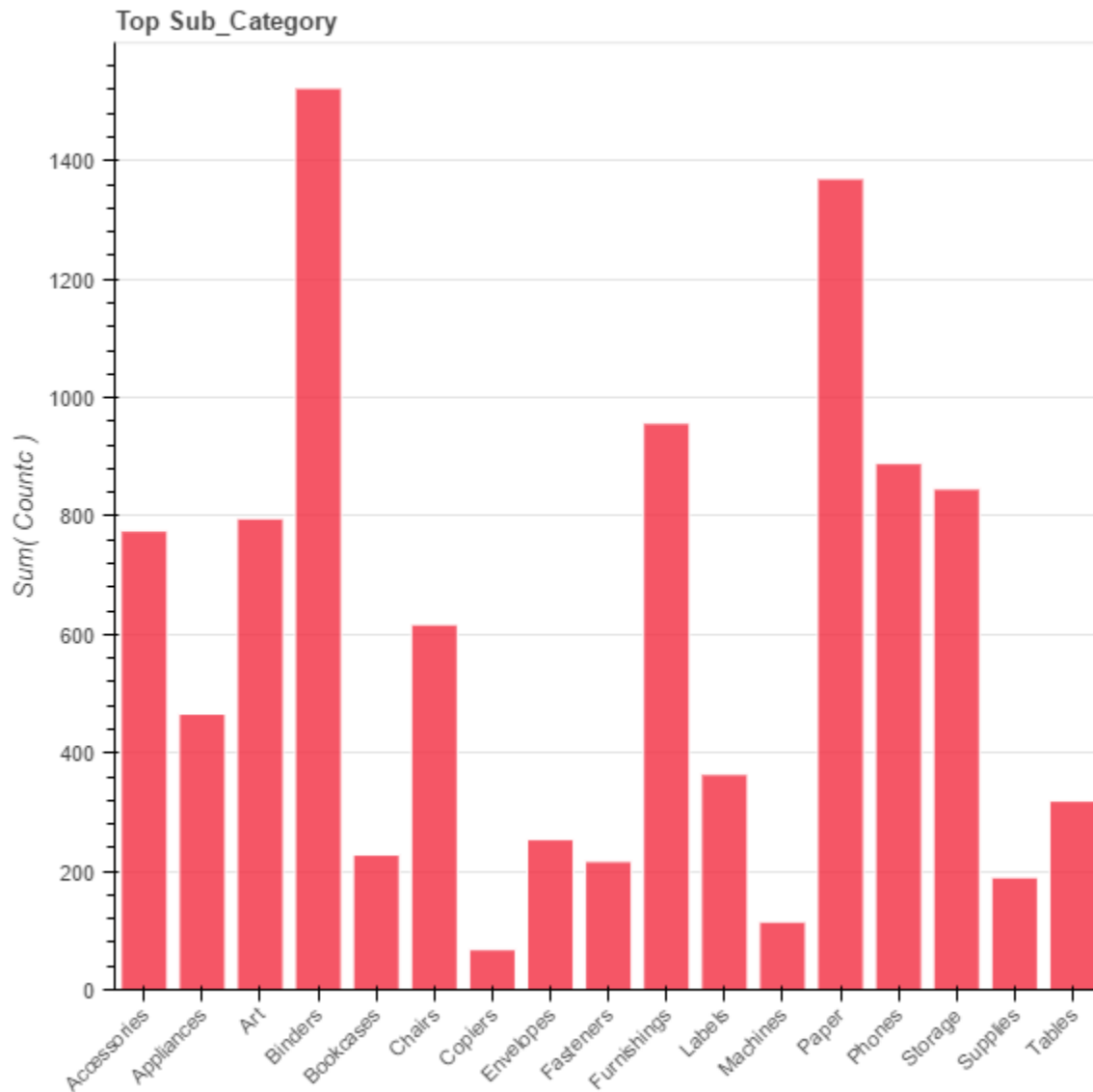
[Figure-3 Segment]

After I had the rough idea from where the customers were, I didn't know what they do? So for that I created a Bar chart of Segment, from this I can tell which segment would be my target for future as well as I can plan some strategies to keep them busy shopping.



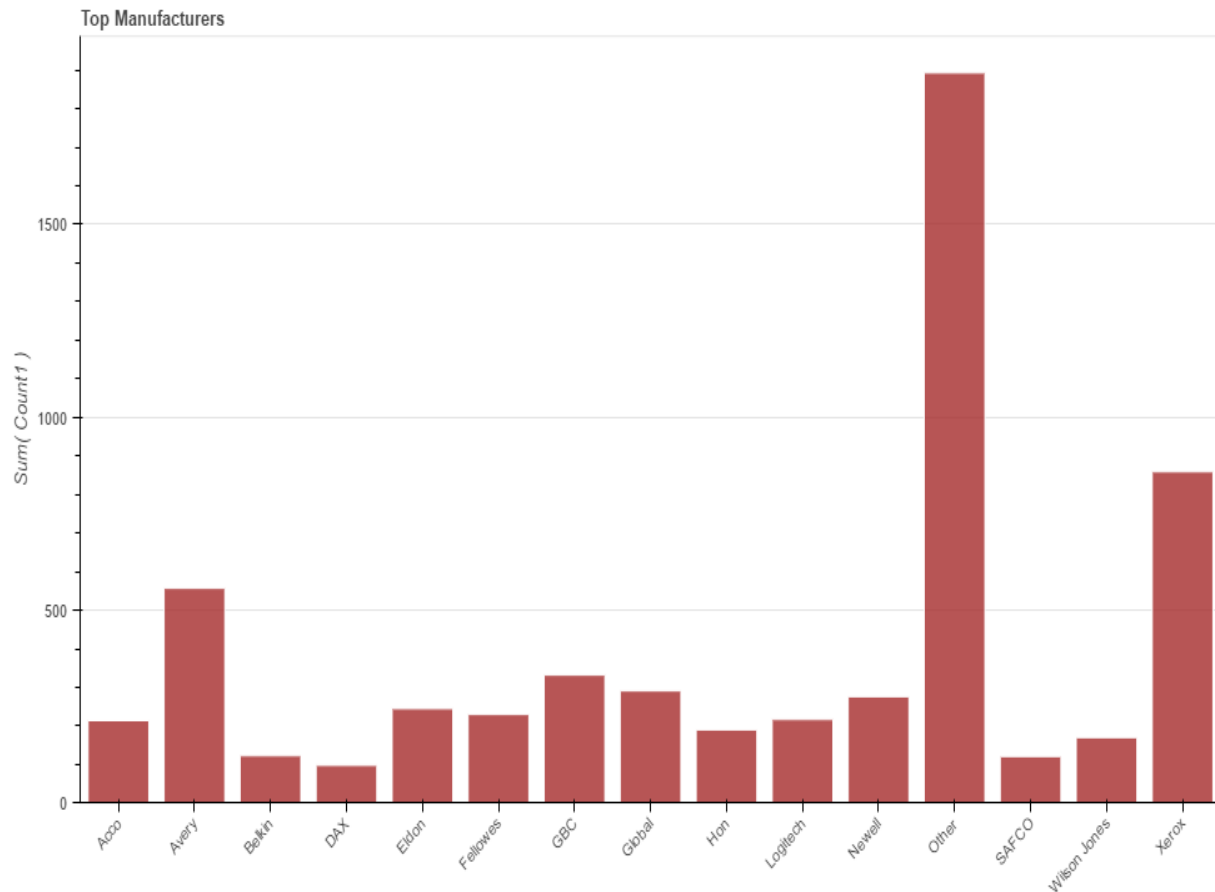
[Figure 4 Categories]

After I have known which segments boosts my sales, I still don't know what they ordered most frequently? So for that, I created a Bar graph according to the categories which showed me most of orders which were placed. Also, it shows that all product belongs to the Offices Supplies category followed by Furniture and at last Technological products.



[Figure 5 Sub Categories]

From the graph I can tell that Papers and Binders are the two sub categories boosting the sale of the company. Also, we can see that figure 5 supports the output of figure 4 stating that Office Supplies have the highest sales. Moreover, New York City and Los Angeles are the two most important technological as well as Financial hub of USA. So this figure 5 also supports figure 2.



[Figure-6 Manufacturer]

From the graph of Manufacturer, we came to know that “other” category of manufacturer has highest sales because there are small things people buy if they get it cheap. But, if you look more carefully you can infer that customers would prefer to buy products of “Xerox” which is associated with Xerox machines and other product which usually come under Office supplies.

We also counted the Average time duration required to deliver the parcel to the customer at any location which turned to 4 days which is good because you don’t want to keep your customers waiting.

Conclusion

From the above graphs it can be inferred that office supplies plays a major role in increasing the revenue of the company. Also, we got which city placed highest order. Moreover, we have the graphs for segment and Categories, so if the company focus more on office Supplies and within daily intervals give customers Discount related to their categories, then customer would be more likely to place order and would stay for a longer time with the company which would eventually increase their sales and market.