DATA ANALYSIS OF

SALES OF A SUPER STORE

SURYA MURALI

M12462296

**INTRODUCTION:**

This report presents the quantitative findings of the sales of a super store based out of Canada. The exact location from where this dataset was taken is: <https://community.tableau.com/docs/DOC-1236>

The dataset provides information of the orders of customers from different locations related to office supplies, furniture and stationaries. The superstore is assumed to be headquartered in Columbus, Ohio, United States and is named as “Cenation”. The names, geographic locations and few other data have been altered for statistical purposes. The original dataset had 8400 rows, out of which a sample of 4862 rows have been used in this report.

This report analyzes the effects of discounts, customer satisfaction, order priorities, shipment processes, type of products sold, location of customers, etc. on the sales of the store. It also provides a location wise and a year wise analysis.

**STATISTICAL ANALYSIS:**

Let us start by analyzing the total sales of its products across cities. Cenation operates in 10 cities across the United States. The graph reveals that bulk of its revenues comes from Cincinnati, followed by San Francisco.

Now let us analyze the Profit margin for each city. Profit margin is calculated as net profit divided by sales. In the Excel, it is calculated by dividing the value of the corresponding Profit cell by the value in the corresponding Sales cell. In terms of average profit margin per city, Atlanta tops the charts, which is closely followed by Miami. However, there is not much difference among the other cities except for Boston, which clearly ranks the least among the them.

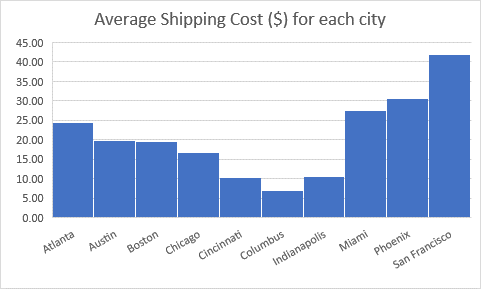
These findings are in sync with the average customer satisfaction ratings for each city. Boston ranks lowest even here and hence corresponds to low profits. Customer satisfaction rating is the rating provided by the customer in a scale of 1 to 5 in regard to their purchases with the store (5 being the best rating).

One of the key aspects of customer satisfaction is how soon the customer gets the order delivered after he purchases. Plotting the graphs for average shipping duration for each city, we observe the following:

It is puzzling to note that, despite being the company’s headquarters, Columbus has the highest shipping duration (delays). But on further analysis, we find out that Columbus has around 33% of its orders marked as low priority. San Francisco has the lowest shipping duration because major chunk of its orders has a critical, high or medium priority. The split up is shown below:

Each product has different shipping costs and unit prices. The average shipping cost and average unit price for each product seem to follow suit.

But Shipping cost majorly depends on where the product is being shipped. When we plot a graph for Average shipping cost for the ten cities, we find that, San Francisco has the highest average shipping cost, followed by Phoenix and Miami. Since these cities are the farthest from Columbus, they naturally tend to have high shipment costs. Cincinnati and Indianapolis have low shipping costs as they neighbor Columbus.

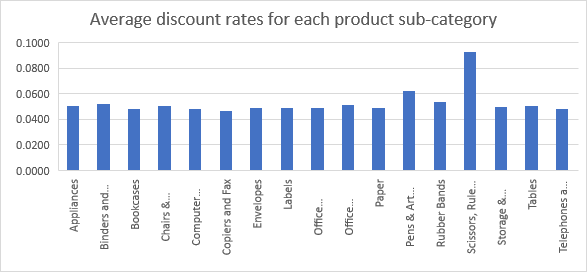


The shipment details for each city also indicates that neighbouring cities of Columbus use a lot of delivery trucks, while faroff places use Expressand regular aeroplanes for delivery.

There has been a steady rise and fall of customer satisfaction rating in the last 4 years from 2009 to 2012. This is in correspondence with the average profit margin achievedd by the company in the last 4 years. This indicates that company’s growth / profits depend heavily on customer satisfaction. Note that the ‘Year’ values in the graph are retreived using the excel formula function: ‘=YEAR(cell\_reference)’

The average customer satisfactcion rating for each product sold by t he company reveals that “Scissors, Rulers and Trimmers” have got the best ratings followed by “Pens & Art Supplies”.

These two products respectively have been offered the highest discount rates by the company. Thus costing less could potentially be one of the reasons for these two products to be popular among customers.



But when you plot average discount rates against the quantity of orders, there is not much of skewness involved. The graph is relatively flat. The Kurtosis value is -0.67 and its skewness is 0.91. This shows that the discount rates don’t have any effect on the quantity of orders.

The average sales for different discount rates also do not provide much information that we can infer. However, there is an outlier at a discount rate of 0.25.

The average profits of the company start reducing when the discounts are being increased. The company would have been in a loss during such scenarios.

The average customer satisfaction rating has a direct relationship with the order priority. Critical orders have the best customer satisfaction rating and the low priority orders have the least rating. This might partially be due to the quick repsonse of the company during critical orders and delayed responses towards low priority orders.

|  |  |
| --- | --- |
| **Order Priority** | **Average of Customer Satisfaction Rating** |
| Critical | 4.81 |
| High | 4.59 |
| Medium | 3.91 |
| Low | 1.99 |

Here, the skewness factor is -1.5. Since the absolute value of skewness is greater than 1,the graph is highly skewed. Here, it is positively skewed and hence Median(4.25) > Mean(3.83).

The Bubble Chart above shows the relationship between customer satisfaction rating, profit margin and shipping duration. The X-axis refers to the customer satisfaction rating, Y-axis to the Profit margin and the size of the bubble refers to the number of days between the day the customer ordered and the day the customer receivedd the product.

Customer satisfaction rating is in a scale of 1 to 5. The Profit margin above the X axis is positive, and if it is below, it indicates that it is negative.

The first big bubble indicates that it has high shipping duration, negative profit and hence the least vallue in the customer rating. While the smallest bubble in the right shows that it has the least shipping period, high profits and thus has the highest rating from the customers.

The customer satisfaction against the average profit margin represented in Data bars would look like this:



Red indicates loss, while green indicates profit. Clearly, the customer satisfaction rating of 4 and above profits the company, while the rest don’t.

**CONCLUSION:**

From these data, it can be inferred that the company has to boost its sales in regions like Boston and Chicago. It should also concentrate more on its low priority orders and act with minimal delays and thus gain wider audience.

**Note**: There were few instances where countif() and countifs() function in Excel was used.

Syntax: countif(range, criteria)

Countifs(criteria\_range1, criteria1, criteria\_range2, criteria2)

The results of the function is shown below:

|  |  |
| --- | --- |
| **Customer Satisfaction Rating** | **Number of Instances** |
| 1 | 698 |
| 2 | 269 |
| 3 | 554 |
| 4 | 1223 |
| 5 | 2117 |

|  |  |
| --- | --- |
| **Discount** | **Number of Instances** |
| Greater than or equal to 0 and less than 0.50 | 4850 |
| Greater than or equal to 0.50 and less than 1.00 | 11 |