

Application of Tableau in Visual Analysis Data of a US Supermarket Sales

Yutao Li

College of Information Science and Technology
Jinan University
Guangzhou
3240805286@qq.com

Abstract—In today's large and complex data background, data needs to be properly interpreted and expressed in order to convey information more clearly. In this paper, a powerful visualization tool, Tableau is used to make visual analysis of online sales data of an American supermarket, the results can better understand the information of sales situation. This can better assist decision-making and provide decision support for the managers of the supermarket.

Keywords—Tableau, visualization, data, supermarket

I. INTRODUCTION

With the development of society, data is playing an increasingly important role in our life. At the same time, the amount of data is getting bigger and bigger, so the traditional manual way of processing data is very expensive. However, visualization has proven effective for not only presenting essential information in vast amounts of data but also driving complex analyses [1]. Massive data can only be perfectly displayed after being properly collected, interpreted and expressed, and visualization is undoubtedly the most effective way to make data user-friendly and easy to understand [2]. Through visualization, librarians can present large amounts of information in a concise manner and analyze large data sets to produce new insights [3]. Also, the emergence of visual software and programming languages have provided us with a variety of choices, making data processing convenient and efficient. For example, with one tool, some information, and a limited quantity of training, librarians may tap into volumes of strategic and operational data to craft beautiful, yet informative data visualizations that examine everything from the utilization of library e-assets or administrations, to the profundity of library assortments on a specific theme [4].

With the popularity of the Internet, the advantages of online shopping are more prominent, progressively become a significant type of shopping. In the future, there will be more enterprises to pick online deals. If they can get timely feedback of sales data and timely adjustment of business strategy, the economic benefits of enterprises and the ability to participate in international competition will be improved. In this context, it is particularly important for merchants to do a good job of online shopping data analysis to obtain business strategies. This paper uses the data visualization tool to carry on the simple research to this, so that merchants could easily observe the operation situation of the supermarket.

II. DATA AND METHOD

The data set I used was from Kaggle, and it is about online orders of a US superstore from 2014-2018. Then I'll introduce the main parameters used. 'Order Date' refers to the date on which a customer places his/her order. The parameters 'State' and 'City' represent states and cities in the United States, respectively. 'Category' is category to which each product belongs to, while 'Sub-Category' is sub-category of each category. In addition, there are three basic variables about

sales. 'Sales' refers to Selling Price of each product, 'Quantity' refers to number of quantity available for a particular product and 'Profit' is profit gained on each product.

The main tool used in this paper is tableau 2021.2. Tableau is a software that can help users explore and understand their data by creating interactive visualizations. The software enjoys the benefits that that it can be used in conjunction with almost any database, and it is easy to use by dragging and dropping to create an interactive visualization expressing the desired format [5]. It can realize data interaction and visualization analysis and dashboard analysis applications, broadly utilized in the field of business intelligence [6].

III. VISUAL ANALYSIS OF A US SUPERMARKET DATA

A. Best-Selling States and their Major Cities

Products from this supermarket are always more popular in some places. Figuring out these areas will help us accurately implement sales strategy, advertising strategy and supply chain management. So, we first analyze the sales.

Filled maps can intuitively reflect the number and corresponding location. The size of the packet bubble chart is also an intuitive indicator of the size of the data.

The filled map(fig.1) shows sales by state, with the darker the color, the higher the sales. We can see that California, Texas and New York are the top three.

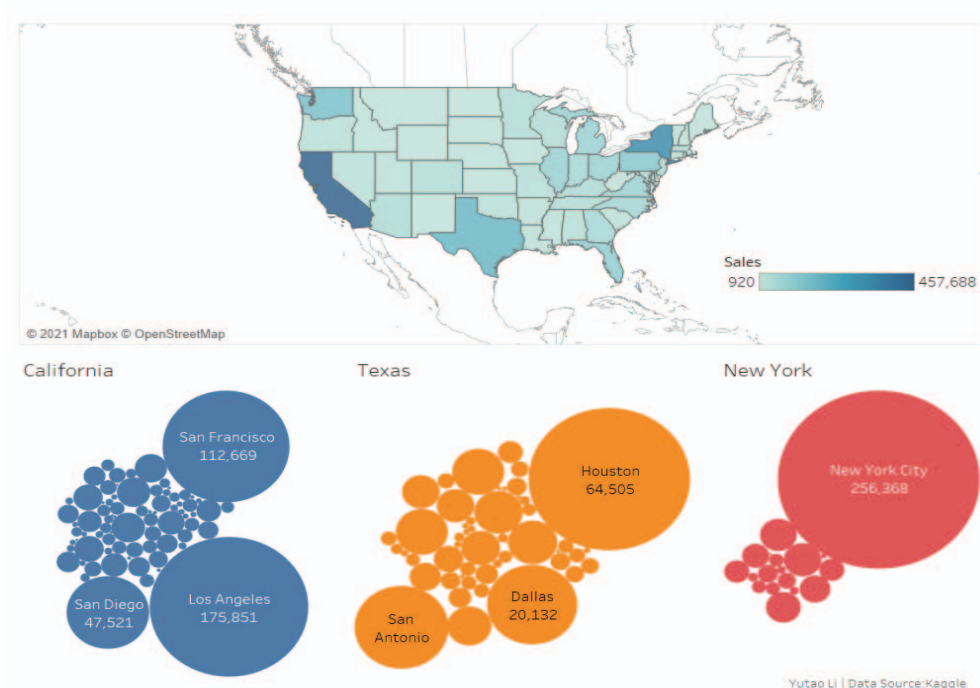
In the packet bubble(fig.1), the size of the bubble represents how much they sell, and the bigger the bubble, the more they sell. And only the cities with high sales are shown here. As shown in the packet bubble(fig.1), in California, there are three cities with high sales: Los Angeles, San Francisco and San Diego. In Texas, Houston, Dallas and San Antonio obviously the top three. However, in New York State, only New York City is relatively high, and it is also the highest in the country.

B. Basic Picture of the States with the Worst Losses

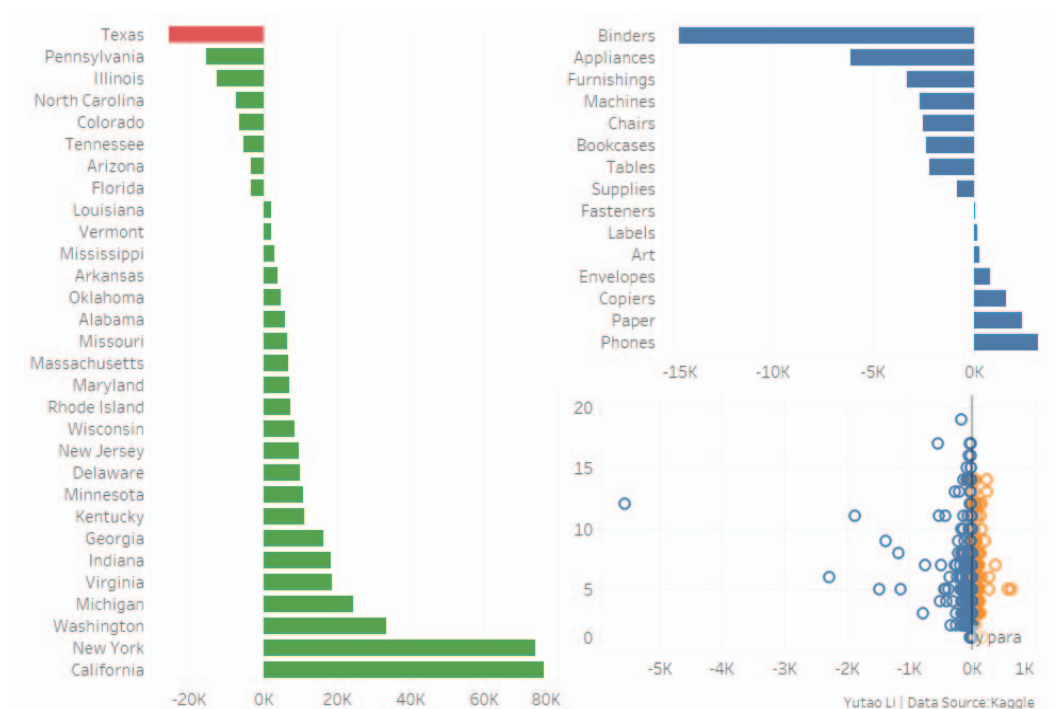
It's important to note that high sales don't necessarily mean high profits. Conversely, some states may have high sales but are losing money, so it is also important to study the profitability of each state. Knowing the state with the biggest losses and its fundamentals will help us propose targeted solutions.

The bar chart can be used to quickly compare the value of each information. When the data is divided into several categories, the bar chart effect can be used to easily find the comparison between the data of each item[7]. Scatter diagram can show the specific distribution of each case. So, it's a good way to use them to analyze this problem.

On the X-axis of the bar chart(fig.2)is profit. The X-axis of the scatter plot(fig.2)is also profit, and the Y-axis is quantity.



The best-selling states and their major cities



Basic picture of the states with the worst losses

For the scatter chart(fig.2), trades with profits greater than zero are marked in yellow, and the rest in blue.

As the bar chart on the left(fig.2)shows, Texas has suffered the worst losses. Its product, Binders, lost the most money, more than twice as much as Appliances, the next largest. But not all of them lost money. Paper and Phones are still profitable.

Another point of note can be visually seen in the scatter diagram(fig.2)in the lower right corner. There are far more losses than gains, and the maximum losses are far greater than gains.

C. Sales and Profits of the Various Products

After the analysis of the different states, we turned to the margins for different kinds of products. Knowing the sales of

different types of products allows us to determine whether these products should continue to be sold. Further, we can figure out how to allocate the quantity of goods.

Same as the last question, scatter diagram is very appropriate.

The product types are divided into four areas. The ordinate is based on the average of sales, and the abscissa is based on the average of profits.

The products at the top right of the chart(fig.3), with high sales and high profits, are the core products. For the products marked orange, the supermarket could consider increasing the number of orders to get a higher profit.

At the bottom right of the chart(fig.3), the products are not selling enough, but they are profitable, so we may consider expanding the market appropriately.

The product at the top left of the chart(fig.3)sells well but loses money. You can try to cut costs and discounts.

The product categories in the lower left of the chart(Fig.3), such as Fasteners and Labels, have very low profit and sales and consider discontinuing sales of these products.

D. Cost Profile of the Various Products

In addition to sales volume and profit, another essential factor is cost. Clarifying average and total costs is an important aspect of cost control.

The chart on the left is a good way to visualize the difference in costs, while the bar chart on the right is suitable to rank the average cost.

In the left graph(fig.4), the red circle represents sales, the blue circle represents profit on sales, and the line segment between the two points represents the difference between sales and profit on sales, namely total cost. Chairs and telephones have higher values.

What's more important is the average cost on the right-hand side(fig.4), if the average cost is too high, one possibility

is that even if the sales volume is high, the profit is not high. So, the cost control of Machines and Copiers needs a lot of attention. Perhaps this explains why they sell more but make less money. Comparatively, the total profit of paper is high, even if the sales volume is not very high, it is better because of the low cost.

E. Sales in Different Months and Working Days

Order quantities tend to be periodic. Identifying the peak time will help us better deploy our manpower. Relatively few employees can be deployed when overall sales are low at a certain time, and vice versa. So, we focus on sales in different months and working days.

It's a good idea to have a dashboard that combines monthly, weekly and daily data.

This picture(fig.5)clearly shows the comparative analysis of sales in different months and working days. The squares in the picture represent daily sales, with darker squares representing more sales and lighter squares representing less sales. The same goes for months and weeks.

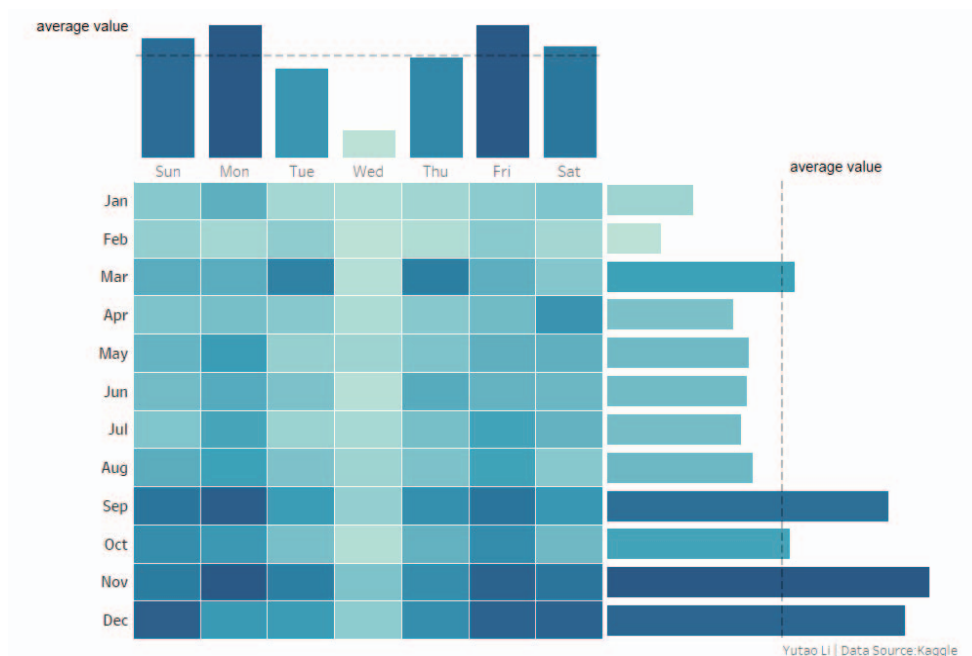
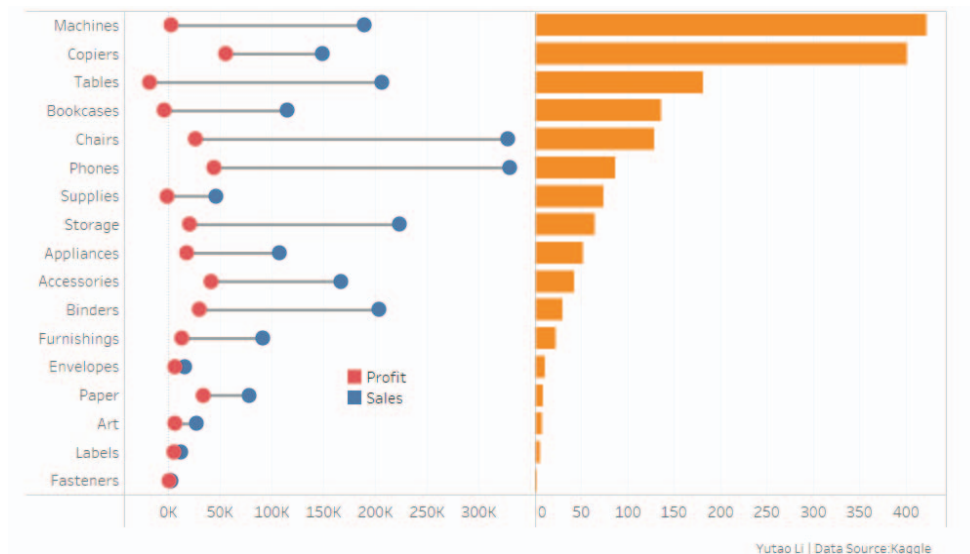
It's hard to tell from daily sales alone, but we can clearly see the monthly variation and also variation in working days. Sales on Mondays, Fridays, Saturdays and Sundays are above average while the remaining three days were below average. Overall, sales were close to average every day, with the exception of Wednesday, when sales were particularly low. Meanwhile, March, September, November and December are above the monthly average, with November and December sales particularly high. February's total sales were undoubtedly the lowest.

CONCLUSION

In summary, this paper mainly studies the sales data visualization system in the following aspects:(1) the best-selling states and their major cities, (2) basic picture of the states with the worst losses, (3)the sales and profits of the various products, (4)the cost profile of the various products, (5)sales in different months and working days; These analyses



The sales and profits of the various products



visualized the sales volume, profit, cost, working hours and other aspects of the supermarket operators, so that they could easily reflect the operation situation of the supermarket. This facilitates the decision-making of supermarket managers.

This visualization system is a system with infinite possibilities. Different people analyze different things from different perspectives, so the analysis results are different. There is still a shortcoming to be improved. To be specific, only basic analysis and visualization of the data are carried out, and complex mathematical model exploration is lacking.

REFERENCES

[1] D. Keim, H. Qu, K. and L. Ma, Big-data visualization, IEEE Computer Graphics and Applications, 2013, 33(4): 20-21.

[2] B. Ling, Application of Tableau in visualization analysis of medical and health data, China digital medicine, 2018, 13(10): 72-74+77.

[3] A. Deardorff, Tableau (version. 9. 1), Journal of the Medical Library Association, 2016, 104(2): 182-183.

[4] A. S. Murphy, How data visualization supports academic library assessment: Three examples from The Ohio State University Libraries using Tableau, College & Research Libraries News, 2015, 76(9): 482-486.

[5] I. Ko, H. Chang, Interactive visualization of healthcare data using tableau[J]. Healthcare informatics research, 2017, 23(4): 349-354.

[6] Y. Yue, Application of Tableau in the Analysis of shipping enterprise line revenue Data, Containerization, 2018, 29(08): 8-9.

[7] H. Bing, H. Liangan, G. Junjie, Data Visualization Application and Practice, Beijing: Enterprise Management Press, 2015: 23-68.