Data Analysis Project

Objective: Increase Profit

Richard Reinhart

31 August 2022

This is my personal project to do analysis from a dataset. This analysis has been done to answer the question ‘What are factors that will increase profit?’ from Superstore-Order data. First, I have to understand the data to make the analysis process runs well. The data is about orders from customers, shipping information, detail product purchased, sales amount, discount, and profit for the store from 2014 until 2017.

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State
CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California
US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida
US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida

Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96	2	0	41.9136
42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.94	3	0	219.582
90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14.62	2	0	6.8714
33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	0.45	-383.031
33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.368	2	0.2	2.5164

After that, I prepared and processed the data. The data is in excel file, make sure the data type was correct, no null values, no duplicate data, and no outliers. For cleaning process we can use filters, remove duplicate, trim function, and make simple visualization to see the visual of data. When I made a simple visualization, I found some outliers in sales data. These three values are bigger than $1{,}5 \times IQR$, so I decided to exclude them.

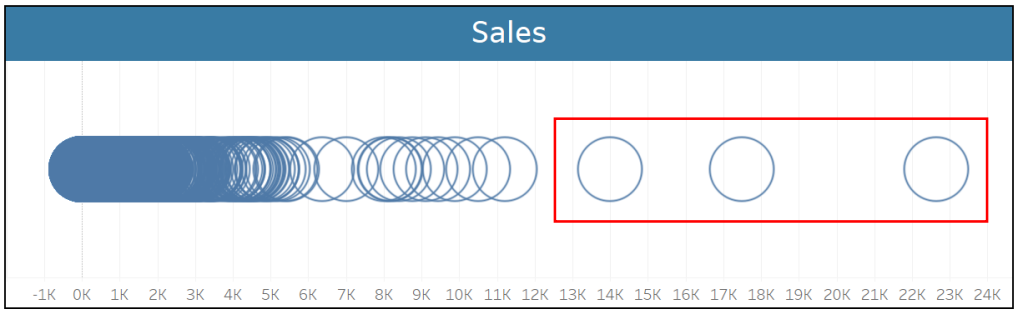


Figure 1: Sales with Outliers

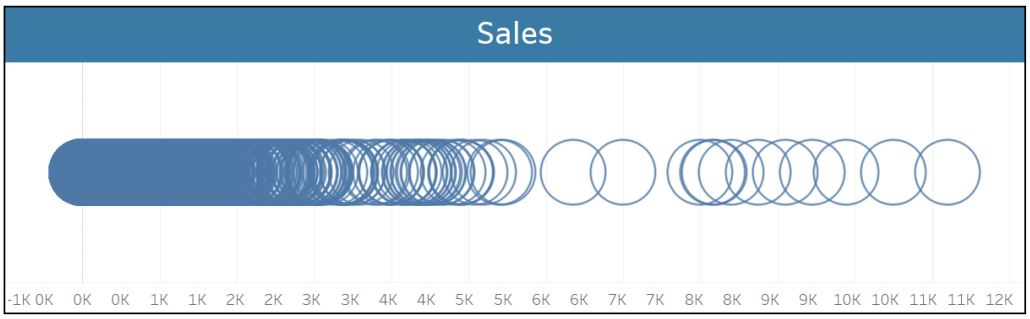


Figure 2: Sales without Outliers

After the data is clean, we can start to analyze it. Focus on the question at the beginning, 'What are factors that will increase profit?' we have to find those factors.

Figure 3: Scatter Plot Sales and Profit

The scatter plot above shows the correlation between Sales and Profit. From that plot, we can see that there is **weak positive correlation between Sales and Profit**, large sales value does not guarantee that it will provide a large profit.

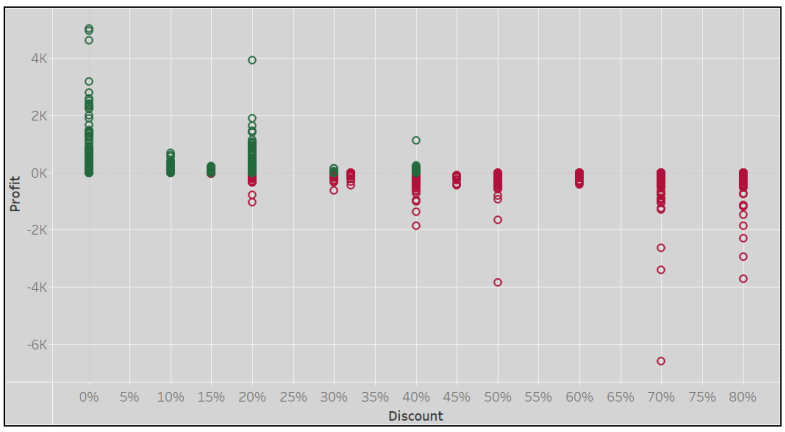
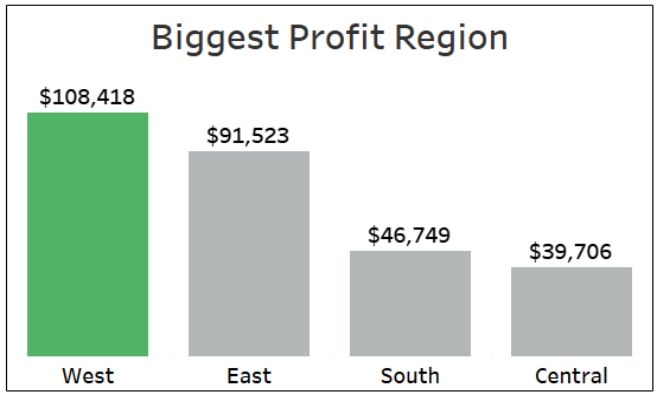
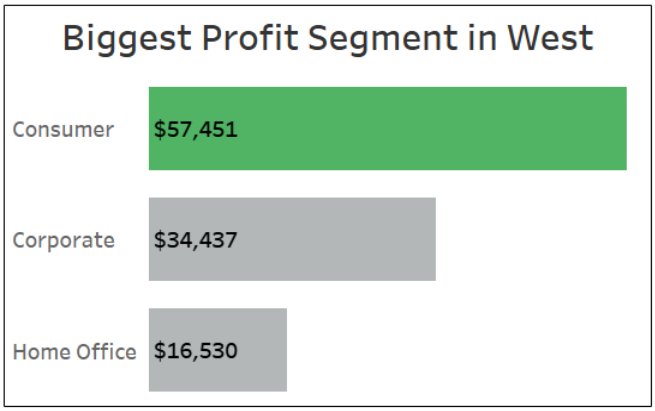


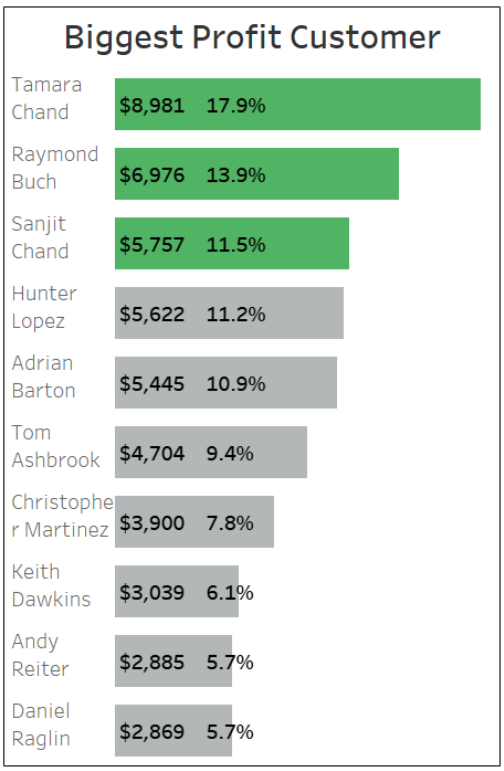
Figure 4: Scatter Plot Discount and Profit

Next, scatter plot Discount and Profit. From the plot above, we can see **negative correlation between Discount and Profit**, the greater discount given, the smaller profit earned.

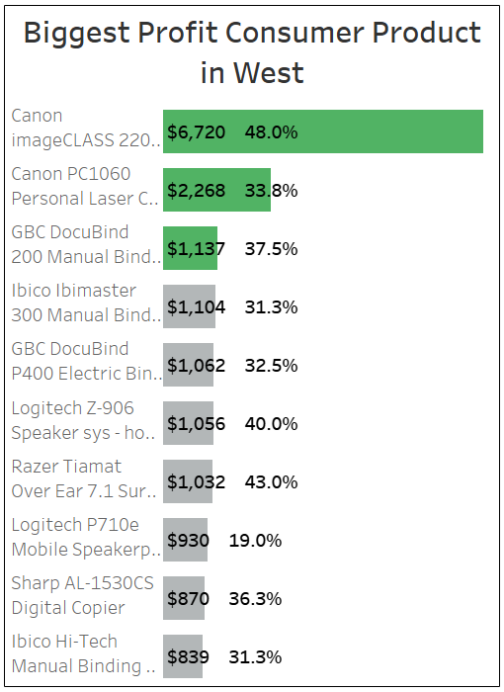




The bar chart above shows that **West** region generated the biggest profit with profit amount **$108,418** or around **37.9%** from total profit in all region. **Consumer** segment contributed **$57,451** or around **53%** from total profit in West. Furthermore, we can look for any Consumer products that generate big profits.



From the bar chart on the left, we know that Canon imageCLASS 2200, Canon PC 1060 Personal Laser, and GBC DocuBind 200 Manual are three consumer products that generated biggest profit amount in West. However, Canon PC 1060 Personal Laser and GBC DocuBind 200 Manual had profit ratio around 30%, which mean the store need to sell more quantity to make big profit. Profit Ratio is a comparison between profit and sales, it can be calculated with total profit divided by total sales.



Top three customer who gave the biggest profit amount is **Tamara Chand** that contributed **17.9%** from total profit, **Raymond Buch** contributed **13.9%**, and **Sanjit Chand** contributed **11.5%**.

Conclusion

- There is **weak positive correlation between Sales and Profit**, large sales amount does not guarantee that it will provide a large profit.
- There is **negative correlation between Discount and Profit**, the greater discount given, the smaller profit earned.
- **West** region generated biggest profit, which is **$108,418** or around **37.9%** from total profit in all region.
- **Consumer** segment contributed **$57,451** or around **53%** from total profit in West.
- **Canon imageCLASS 2200**, **Canon PC 1060 Personal Laser**, and **GBC DocuBind 200 Manual** are three consumer products that generated biggest profit amount in West.
- **Tamara Chand**, **Raymon Buch**, and **Sanjit Chand** are top three customer that generated biggest profit.

Recommendation

- **Decrease percentage of discount** in order to **increase profit**.
- Make a **marketing plan** for **Consumer** segment in **West**.
- Push sales from consumer products that gave big profit, which is **Canon imageCLASS 2200**, **Canon PC 1060 Personal Laser**, and **GBC DocuBind 200 Manual.**
- Establish a **good relationship** with customers that generated biggest profit, which is **Tamara Chand**, **Raymon Buch**, and **Sanjit Chand**. Also give **special offer** of some products that they liked to buy.