Final Project - Costco

Jinglong Xu

Qi Jiang

Yuan Sheng

ALY6070 Communicate/Visual Data Analysis

June 24, 2020

Shrikant Palkar

Northeastern University - Seattle

Final Project - Costco

In this project, the authors firstly propose six questions based on the sales data for a single department and a retail chain. In the second part, the authors clarify the datasets and variables that we are used for visual. In addition, we pretreate data through data selection and transformation to get the data aggregate for visualization. At last, in the process of data analysis, we create some enlightening visuals in Tableau to explain the proposed questions.

# Questions and Data Source

According to the raw information, we create three questions: 1) What kind of brand of appliances are popular based on sales and quantity? 2) Can we calculate the average refund number and rates of appliance brands based on sales and quantity? 3) Which day of the week has the best sales? Whether sales are subject to holiday marketing?

Besides, there are three questions we will find the answer. 4) Which items should we stock more in which region, during different months of the year, based on the sales? 5) Can we rank warehouses based on the sales of items in terms of quantity as well as Dollars? 6) Is there a correlation of items to account types of the members.

The authors extract “IMDES1” and create a “Label” column, which refers to different product brand names in Excel. We keep other columns as raw data and manipulate the data with R, and then export a clean data report into Tableau for visualization. Besides, the authors create a calendar data for date filtering.

# Variable and Method Used on Analysis

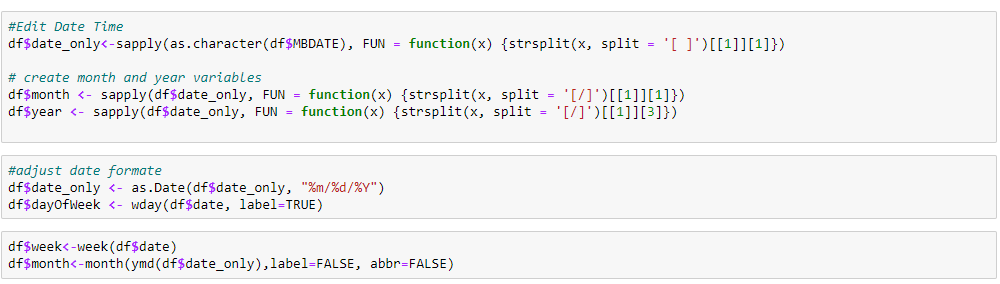
## Variables used on Analysis

MBWHSE: Warehouse Number, MBTRN\_NUM: Transaction Number, MBITEM: Item Number, MBDATE: Date (edit), MBDSUN: Sales Quantity, MBDSEL: Sales, MBRFUD: Refund, IMDES1: Item Description 1, IMDES2: Item Description 2, REGION: Region,  
LCSTA: State, ACCTYP: Account Type, Label : Brand, Day of Week, and Week

## Methods Used on Analysis

We create a LABEL column with all the brands and integrated with MBDSUN, MBDSEL, and date parameters to find out which top three brands have the best performance in Figure 1. Figure 2, we select three variables, including LABEL, MBRFND, MBRDSEL. Then, we build the formula in Tableau to calculate the refund and rate of return. In Figure 3 and Figure 4, the author uses the “lubricate” library in R to change the datatype and to add new columns, which include month, year, day of the week. Next, export a new common-separated values file and interface into Tableau to find the best sales day of the week and holiday season influence. In addition, when the clean data interface into Tableau, the authors change each parameter description for making the audience easy to read and understand. Figures 5, 6, and 7 are three dashboards that analyze the sales by geographic and list Top 5 products by different descriptions. Variables LCSTA and REGION can provide geographic visualization in Tableau. Variables MBDSUN and MBDSEL are quantities of data to measure the total sales amount.

Moreover, Date, IMDES1, and IMDES2 give the category in Tableau. Using MBWHSE, MBDESEL, and MBDSUN, these parameters can obtain the rank of the warehouse based on the sales, as the result shown in Figure 8. MBDESEL, IMDES1, and LABEL variables analyze the relationship between items and the type of customers. The authors use a bar chart to show the percentage of the total.



Code 1: Data Type Manipulate



Code 2: Export CSV file

# Findings- Dashboard and Summaries

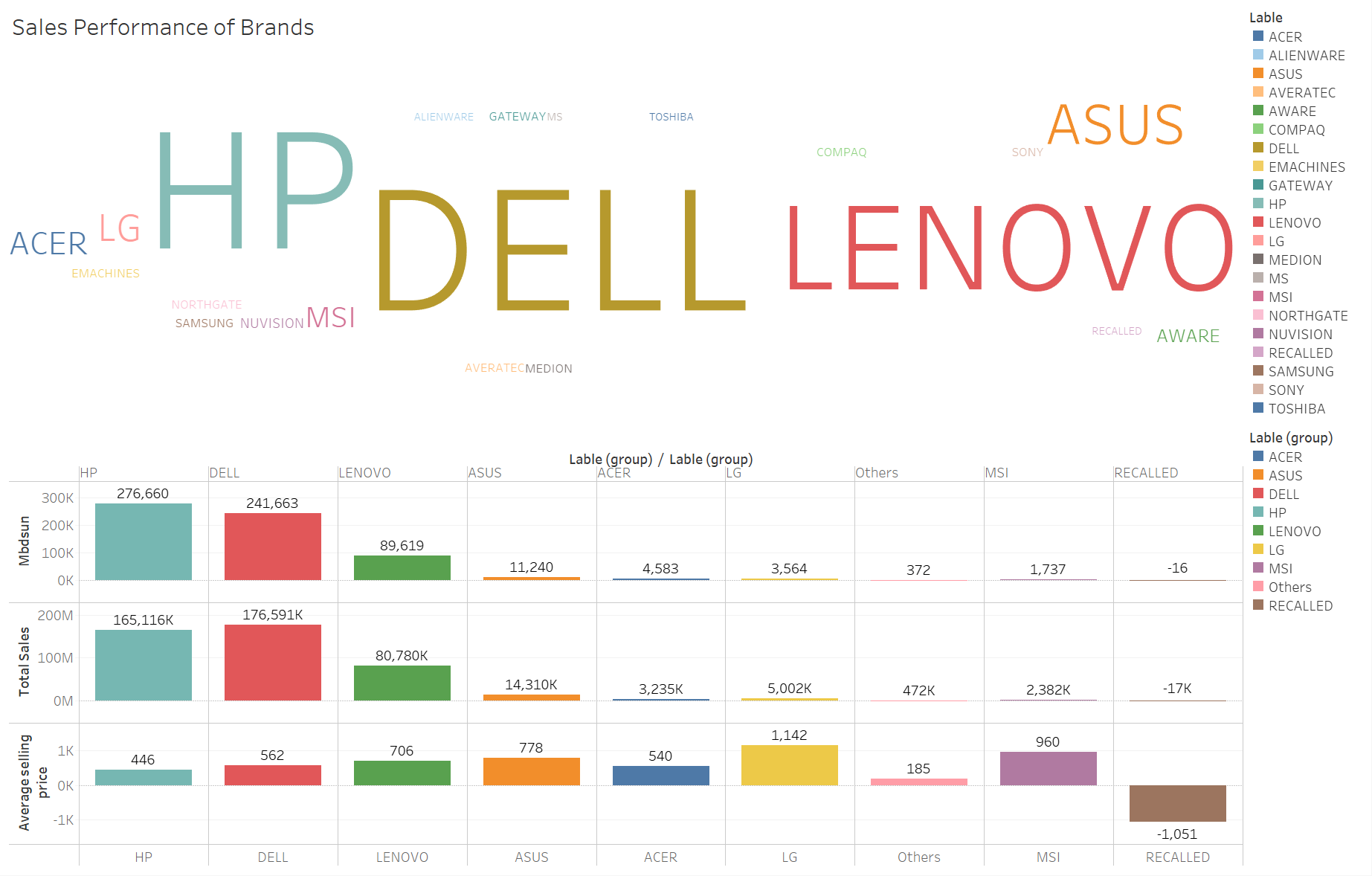


Figure 1: Top three brands

From the above picture, we can see that the top three among twenty-one types of products ranked in the sales data are DELL, HP, LENOVO respectively. The top half of the picture is a word cloud plot, it shows that the relationship between brands and sales. The sales are larger, the size of the brands is bigger. For the bottom part of the dashboard, the sales performance of DELL and HP more than double LENOVO. Averagely, the unit price can be in descending order of LENOVO, DELL, and HP. In conclusion, compared with HP, the average price of DELL plays an important role in boosting sales.

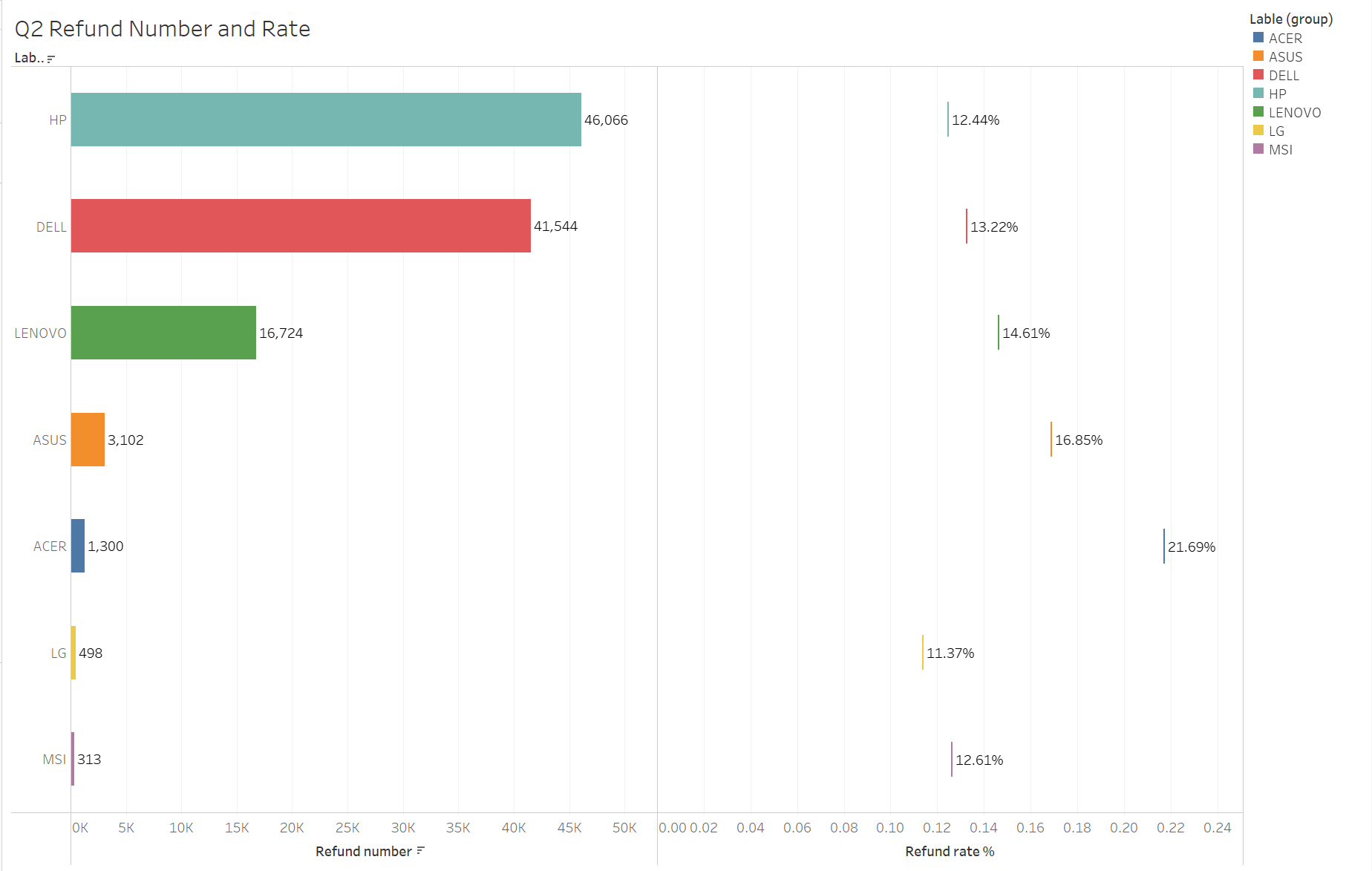


Figure 2: Refund and Rate of Return

From the absolute point of view, HP and DELL are the first two product brands based on refund quantity but based on the first question analysis, we can know that DELL and HP have achieved the best two sales performance, so their returns quantities are seen higher compared with other brands. And in terms of refund rates, these two brands are relatively lower than ACER and ASUS. Therefore, based on the above analysis, it is shown that DELL and HP are popular in America, but the refund rates of ACER and ASUS are higher than other brands probably because of computer configuration and performance.

Figure 3 tells which day of the week has the best sales in the 2019 fiscal year. In this chart, we can conclude that Goldstar members are the main customers of Costco, which is more than 75% of customers are Goldstar members. Business members are about 22% of all the accounts. In the 2019 fiscal year from September 2018 to March 2019, Friday has the best sales day of the week for business members, whereas Sunday is the best sales day of the week for Goldstar members.

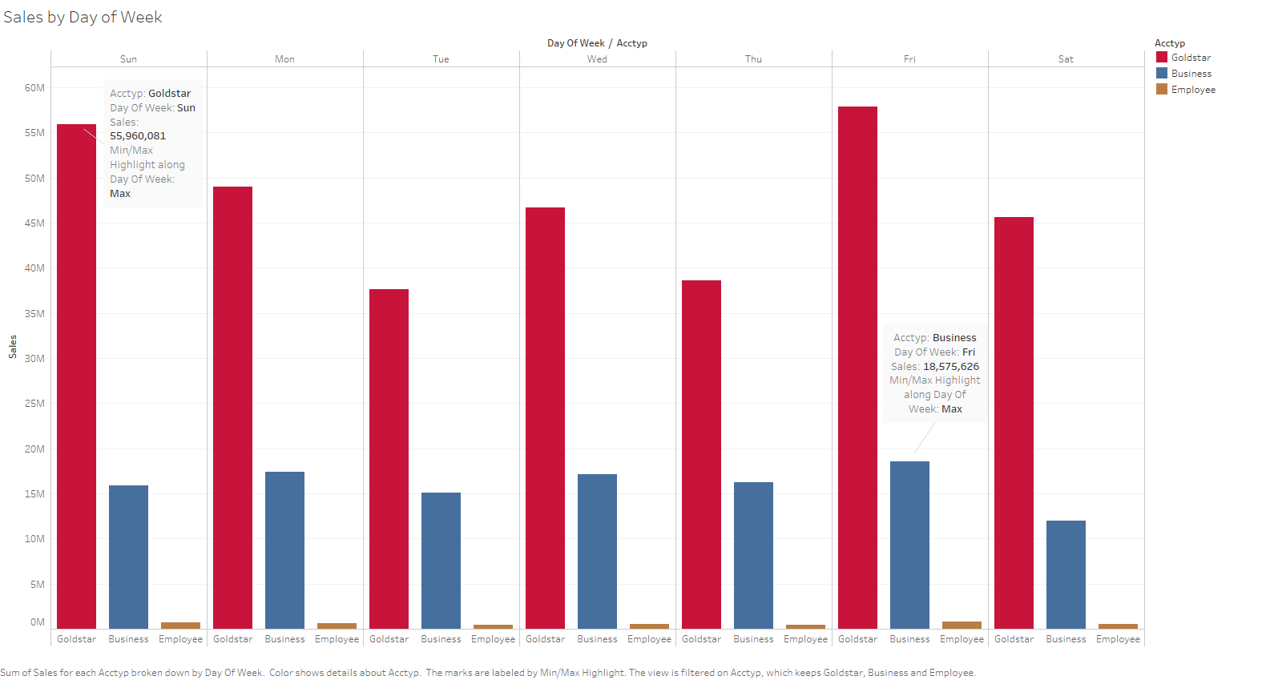


Figure 3: The best sales day of week

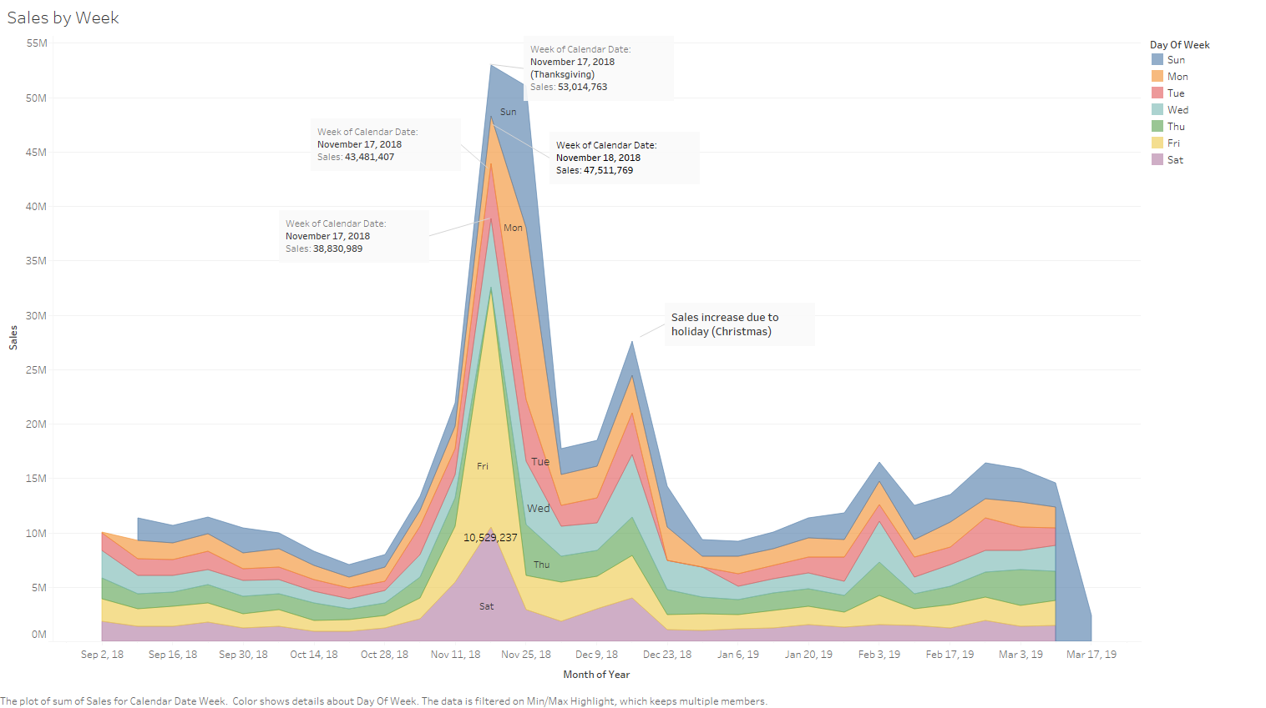


Figure 4: Holiday season impact on sales performance

Figure 4 analyzes whether sales are subject to holiday marketing. The answer is yes. The graph shows the sales increase to the peak during the 2018 Thanksgiving week, and bottom-up before Christmas. Another sales increase shows on the graph is Jan 21, 2019, which is Martin Luther King's holiday.

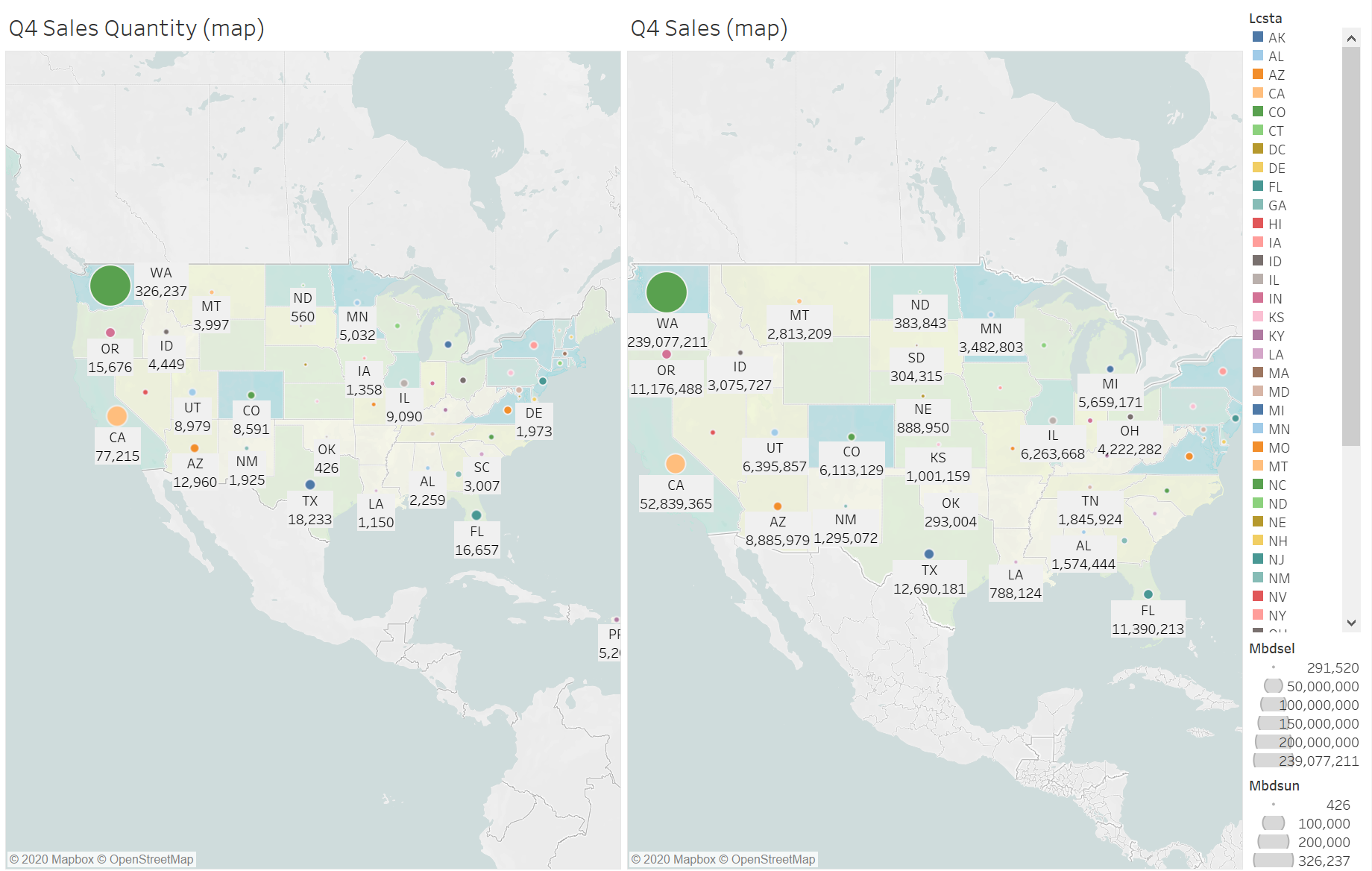


Figure 5: Sales by Geographic

On the first dashboard (Figure 5), we can find that WA is the largest state in the U.S. based on the size of each circle, whether it is sales performance or sales quantity. Followed by that, CA is ranked second on the whole. There are three regions in the state of WA, including EC, NW, BD. Among them, on the second dashboard (Figure 6), undoubtedly, we can find that EC has the best sales and it accounts for nearly 50% of the total sales. More importantly, EC and NW in the state of WA are the top two among all the regions and the percentage of their sales is approximately 60%. On the whole, November is a high peak in promoting sales probably because of the holiday seasons start.

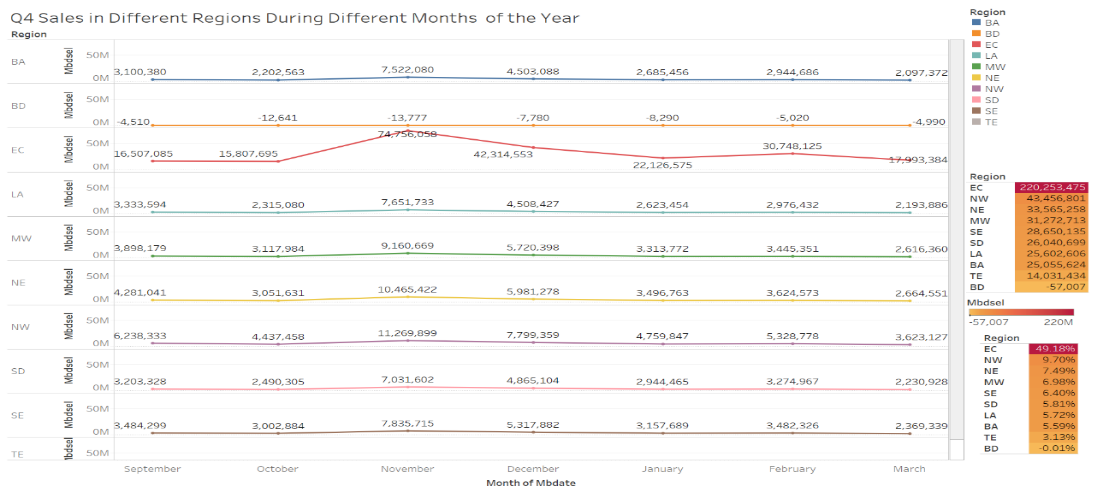


Figure 6: Sales by different regions and month of year

On the third dashboard (Figure 7), after filtering the condition and selecting a specific region, we can know which items we stock during different months, based on the sales. For example, the sales in December are higher than in other months. Top five IMDES1 items in December in the EC region are DELL 5K15 I5-8250U/12/1/T, HP 14-DF0023CL 14" LT, DELL I5570 15.6" LT, DELL I5570 15.6" TCH LT and HP 15-CS0053CL 15.6" LT, and Top five IMDES2 items are I3-8130U/4GB/128SSD, I3-8130U/12GB/1TB, I7-8550U/12GB/1TB, I7-8550U/16GB/512SSD and I5-8250U/12GB/1TB.

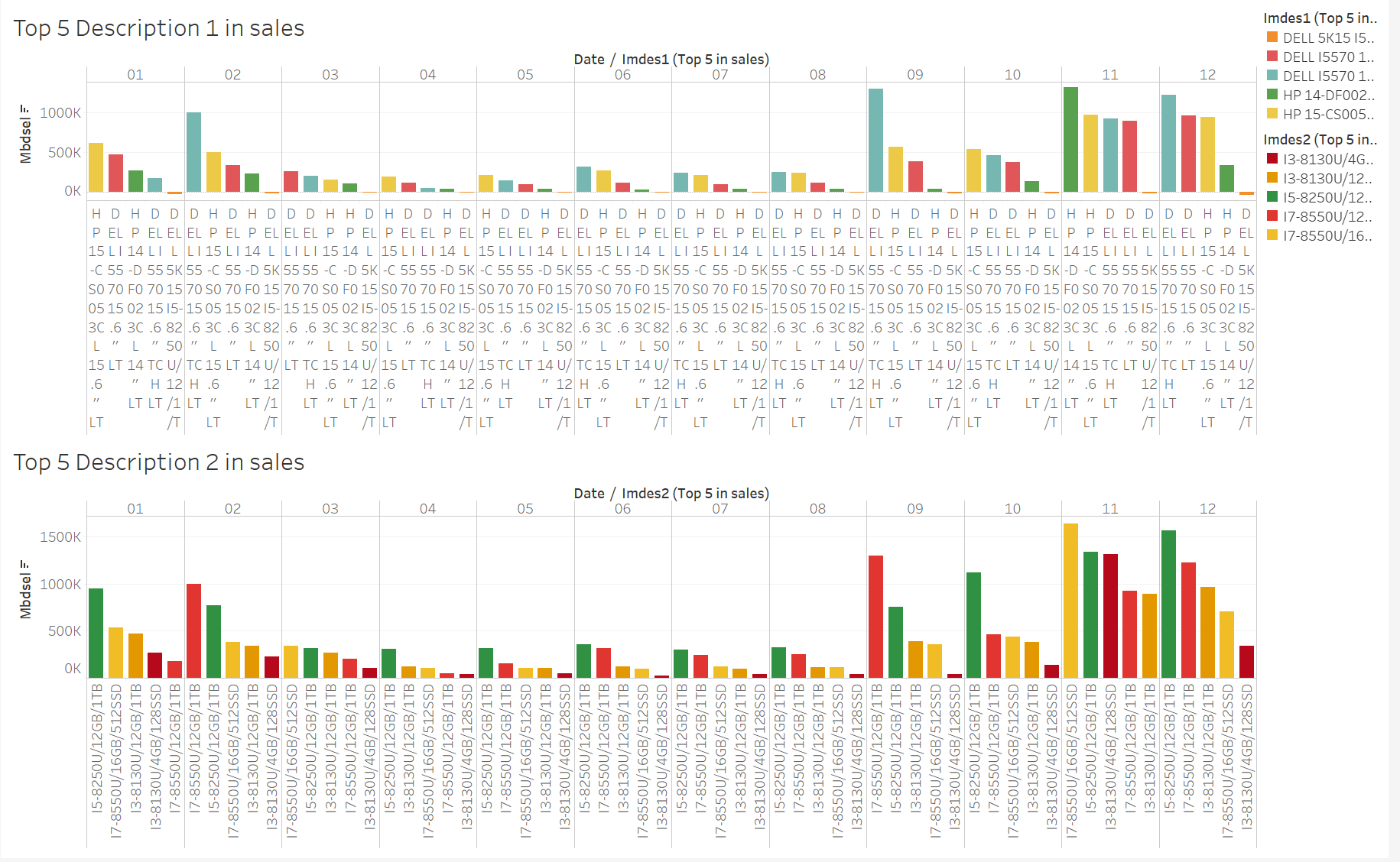


Figure 7: Top 5 Products with Two Descriptions

A screenshot of a social media post

Description automatically generated

Figure 8: Warehouse Sales Rank

The author uses the number of warehouses, sales quantity, and sales of these three variables. Figure 8 obviously shows that No.847 (US E-Commerce) is far ahead of other warehouses both sales and sales quantity and we can know that the sales in most regions are dependent on this warehouse. Additionally, more than 1,000k of sales are warehouse No.2 (Portland), No.246 (Tigard), No.111 (Christiana), and No.110 (Issaquah).

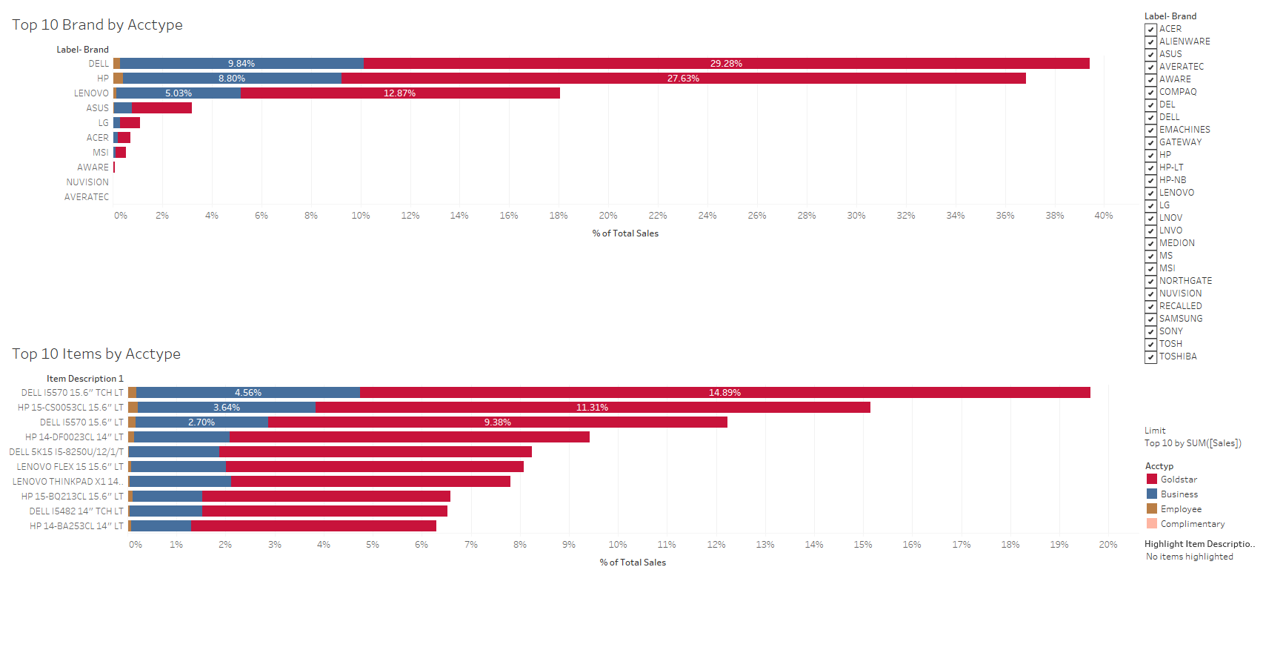


Figure 9: Correlation between items and account type

Figure 9 displays the correlations between product items and account types. On top of the dashboard, it shows all the customers prefer to buy Dell, HP, and Lenovo these good quality laptops or PCs. On the bottom of the dashboard, we can see both Goldstar customers and business customers tend to purchase 15.6” monitor screen size laptops.

# Summary

Goldstar individual customers are the main group of customers. The data set shows electronic products related to laptops or personal computers. Viewing the results, Dell, HP, and Lenovo are the best seller of all the brands. In other words, both business customers and Goldstar customers prefer buying their products. The best sales day of the week is different from business customers and individual customers. Business customers prefer making orders on Friday, whereas individual customs like purchasing on Sunday. According to our observation, the holiday season will trigger the sales increase. EC (US E-commerce) region operates all the online shopping of all-over North America. Therefore, the company needs to store most of the inventories near the E-commerce warehouse. Furthermore, we find customers like 15.6” screen size laptops or monitors. Thus, the company needs to consider keeping these products' SKU (Stock Keeping Unit).