Data Analysis Project

Eleftherios Spirellis.

The scope of this assignment is to analyze data concerning Liquor Stores in USA and reveal the most popular item in terms of bottles sold per zip code and the sales percentages of the stores.

Step 1:

The dataset was added to SQL Workbench by simple running the provided code.

Step 2:

The query used to extract the specified data is the following:

*select \* from finance\_liquor\_sales a*

*where 1=1*

*and a.date >= '2016-01-01'*

*and a.date <= '2019-12-31'*

*order by a.date;*

We exported the data returned by the query into liquor\_sales\_2016\_2019.csv (which is also provided).

Step 3:

We implemented the necessary aggregations into python, using more specifically jupyter notebook, which is also available (Data Analysis - Final Assignment.ipynb). The aggregated data can be found in files sales\_per\_store.csv and popularity\_zip\_code.csv.

For the most popular item per zip code, we grouped the data per zip code and item description, and we summed the number of bottles sold.

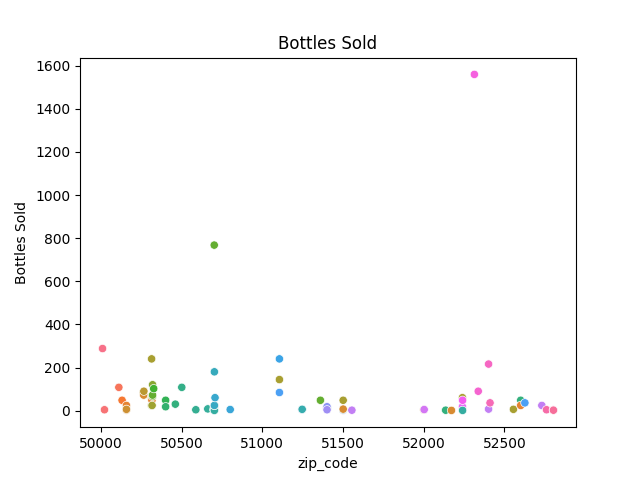
For the sales percentage, even though the description wasn’t very clear, it felt natural to interpret percentage of sales as the sales of each store against the total sales.

Step 4:

We implemented the plots in both python and tableau.

Python with Matplotlib :

1.



2.

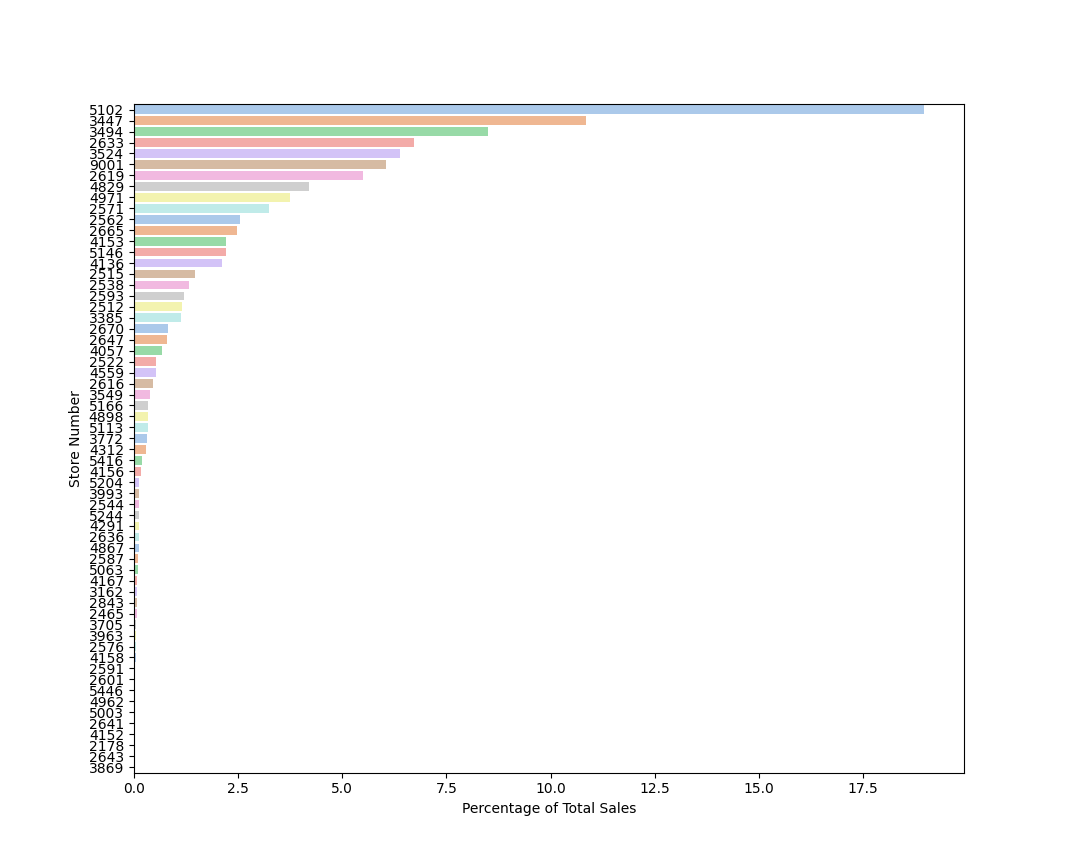
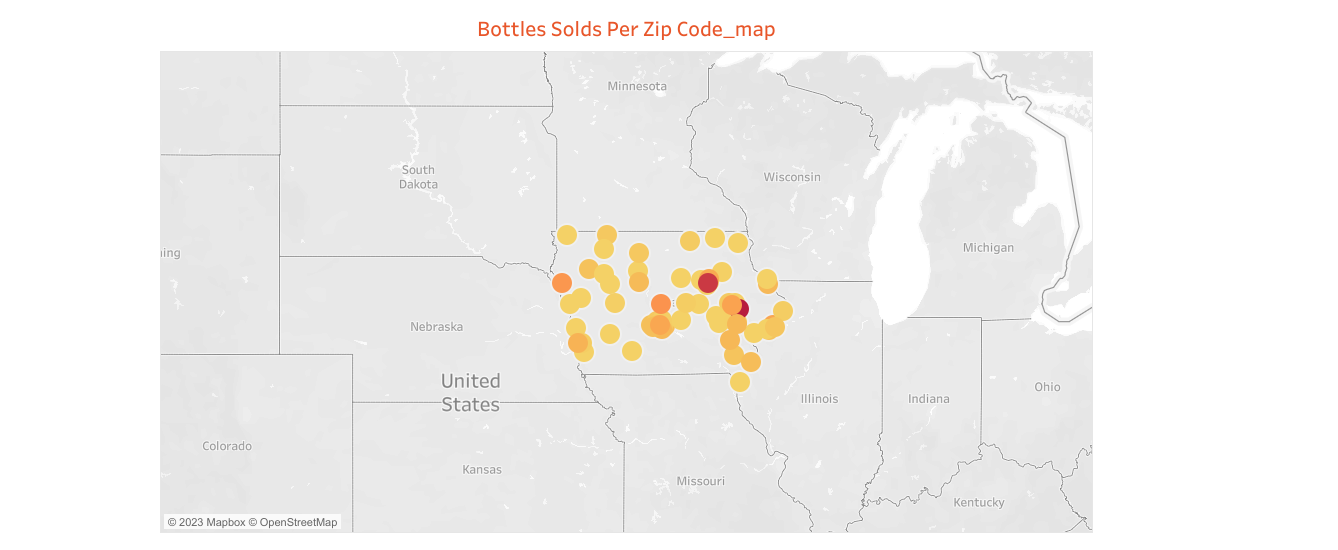
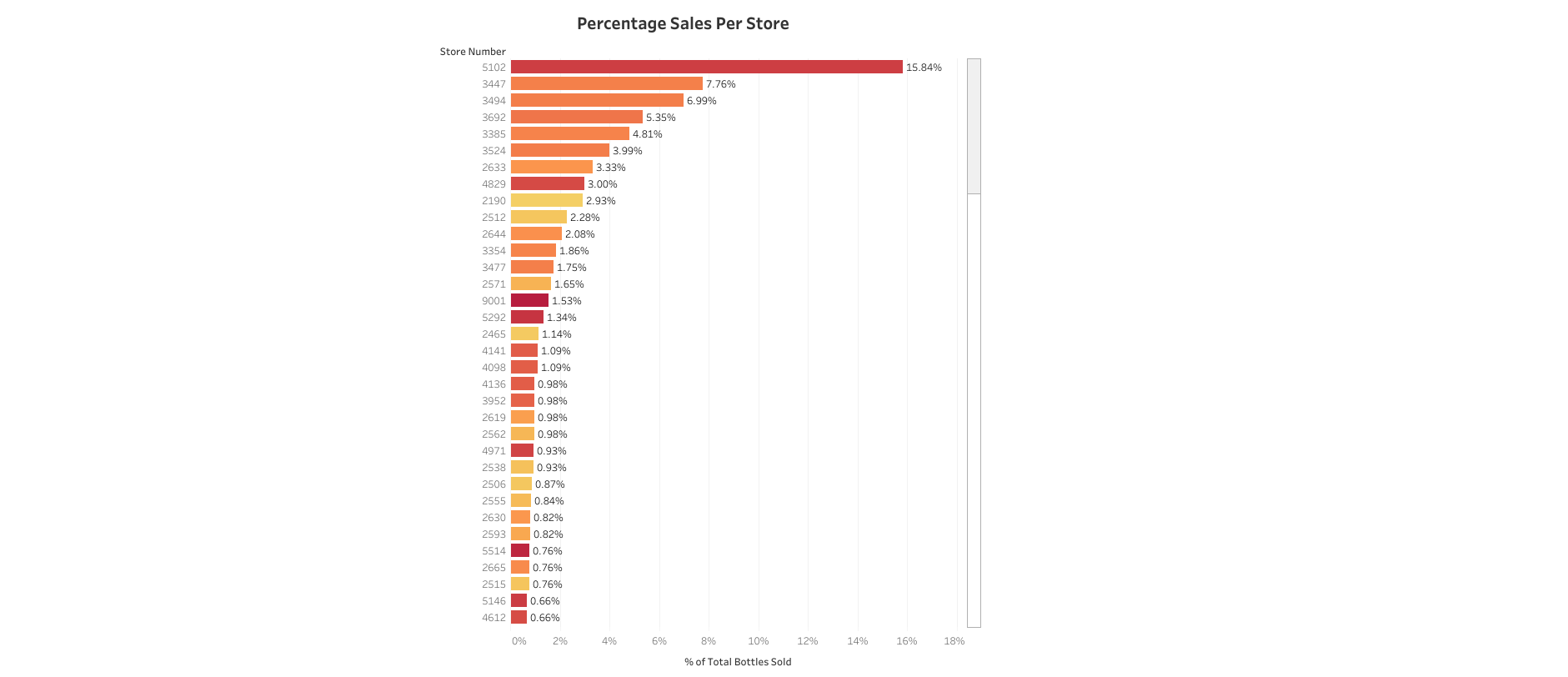
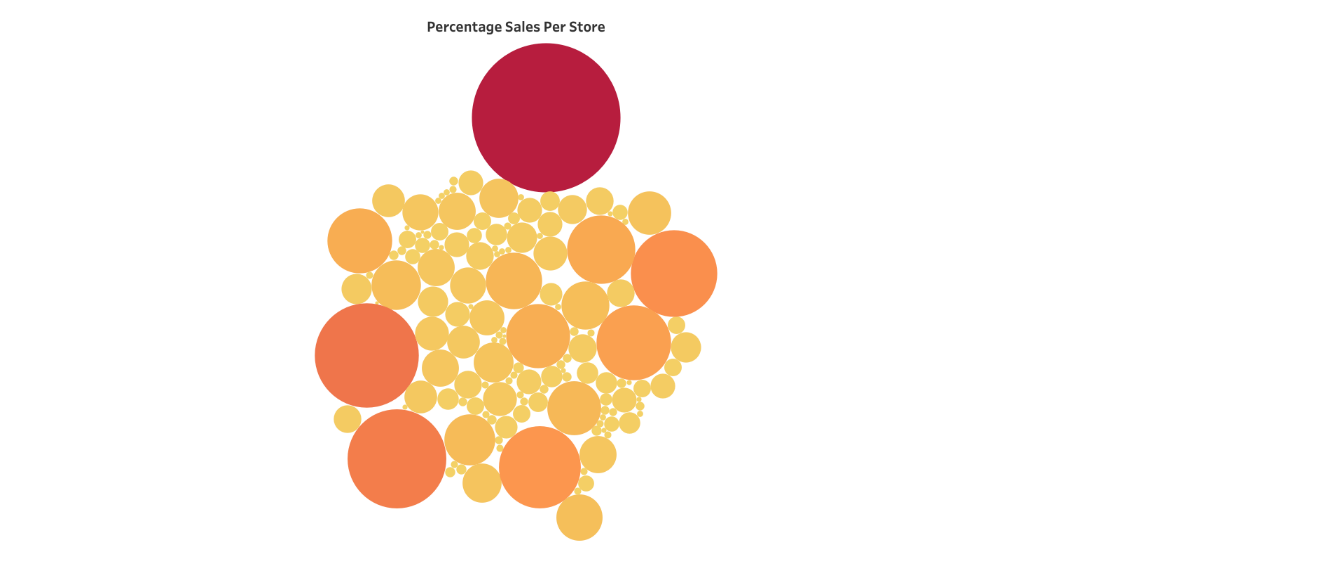


Tableau :









We note here that the plots seem to differ from the ones provided in the example. That suspect that it is because of the different time line (our data refer to 2016-2019).