1)First I imported the ‘finance\_liquor\_sales’ Dataset to my SQL Workbench.

2)To get the results we wanted between the years 2016 and 2019,I wrote the below query:

select \*

from finance\_liquor\_sales

where date>='2016-01-01' and date<='2019-12-31';

3) Then I exported the results to a csv file called ‘finance\_liquor\_sales2016\_2019’.

4) I imported Pandas and Matplotlib with the new csv.

import pandas as pd  
import matplotlib.pyplot as plt  
df=pd.read\_csv('finance\_liquor\_sales2016\_2019.csv')

5)I grouped all data by “zip\_file” and “item\_description” and used the sum method so I found the total battles sold,after that I found the items had been sold the most.

groupedf=df.groupby(['zip\_code','item\_description'])['bottles\_sold'].sum()  
mpi=groupedf.groupby('zip\_code').idxmax() #mpi=most popular items

6)I wanted to mind the percentage of sales per store so I divided sales\_by\_store with total\_sales and multiply it by 100.

sbs=df.groupby('store\_name')['sale\_dollars'].sum()#sbs=sales by store  
total\_sales=sbs.sum()  
sales\_percentage=(sbs/total\_sales)\*100

7)Then all I had to do is to visualize this Data.

result=df.groupby('zip\_code')['bottles\_sold'].sum()  
plt.scatter(result.index,result.values)  
plt.xlabel('Zip Code')  
plt.ylabel('Bottles sold')  
plt.show()