The Battle of Neighborhood

Introduction

- In New York City data. First, we will find the most visited commercial shop according to the number of check-ins, then we will try to find the neighbourhoods that are lacking the selected type of shop which could be potential business opportunity.

Data

- The data comes from the following link
- https://sites.google.com/site/yangdingqi/home/foursquare-dataset. It contains 227,428 check-ins in New York city. The data contains two files in tsv format. Each file contains 8 columns.
- After extracting and reading the data, we will translate the above data into a Pandas data frame for processing which would look like this. These are the data elements that are needed when we call Foursquare web service call in order to get the venues available in that neighborhood

Methodology

- Basic skills from week 3 lab
- Majorly relied on Foursquare API to retrieve all venues of each neighborhoods, then group by each neighborhoods and to count how many venues before filter top 10 most common venue types of each neighborhoods

Results and Conclusion

- In our sample of 2000 venues, we did find more than 10 coordinates that has no Bar (the most visited shop type according to sample) within four-kilometer sphere. And we did manage to get the neighborhoods' names from foursquare database and pin down the two closest
- neighborhoods, 'Bedford-Stuyvesant', and 'Turtle Bay', into the map. Of course, it should not be forgotten that the data used above is almost 6-year old so further research might be needed.
- Anyways, the results according to the data in hand can be checked from the map and analysis can be of use for future entrepreneurs.