

## - Data Section

The will data comes from Dingqi Yang 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, which are:

1. User ID (anonymized)
2. Venue ID (Foursquare)
3. Venue category ID (Foursquare)
4. Venue category name (Foursquare)
5. Latitude
6. Longitude
7. Time zone offset in minutes (The offset in minutes between when this check-in occurred and the same time in UTC)
8. UTC time

After extracting and reading the data, I 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 neighbourhood (Neighbourhoods are not included here)

	VenueID	CategoryName	Visitor Count	Latitude	Longitude
0	49bbd6c0f964a520f4531fe3	Arts & Crafts Store	7	40.719810375488535	-74.00258103213994
1	4a43c0aef964a520c6a61fe3	Bridge	37	40.60679958140643	-74.04416981025437
2	4c5cc7b485a1e21e00d35711	Home (private)	1	40.716161684843215	-73.88307005845945
3	4bc7086715a7ef3bef9878da	Medical Center	1	40.7451638	-73.982518775
4	4cf2c5321d18a143951b5cec	Food Truck	4	40.74010382743943	-73.98965835571289

Then I will create a dictionary in order to decide which category is the most popular (commercial type)

```
[('Train Station', 943), ('Park', 778), ('Airport', 769), ('Bar', 756), ('Subway', 587), ('Coffee Shop', 447), ('Gym / Fitness Center', 447), ('Food & Drink Shop', 426), ('Neighborhood', 362), ('Plaza', 342), ('Stadium', 339), ('Bridge', 272), ('Office', 264), ('Department Store', 240), ('Mall', 238), ('Burger Joint', 206), ('American Restaurant', 202), ('Road', 201), ('Bus Stati
```

'Bar' is the most visited commercial category according to given data.

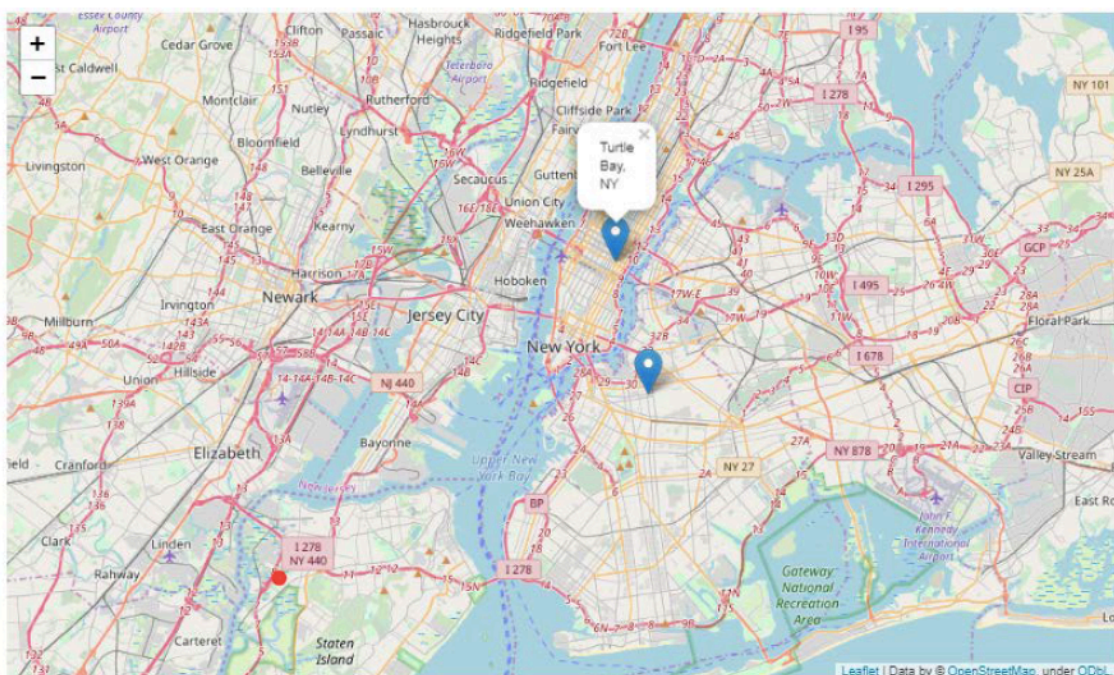
After all this, I will check the coordinates within given n number of kilometres and count how many 'Bar' are there (venues selected as 2000 as a trial).

Coordinates with number of Bar shops within 4 kilometers according to 2000 venues.

```
('40.60613336268842', '-74.17904376983643') : 2  
( '40.719810375488535', '-74.00258103213994') : 0  
( '40.60679958140643', '-74.04416981025437') : 0  
( '40.716161684843215', '-73.88307005845945') : 0
```

Then find the two neighbourhoods that are closest to the coordinate which has the most number of the specific shop type but lacking that within 4 kilometres.

Bedford-Stuyvesant  
Turtle Bay



Red dot is the center

## - Results & Conclusion

In our sample of 2000 venues, I did find more than 10 coordinates that has no Bar (the most visited shop type according to sample) within four-kilometre sphere. And I did manage to get the neighbourhood's names from foursquare database and pin down the two closest neighbourhood's, 'Bedford-Stuyvesant', and 'Turtle Bay', into the map. Of course, it should not be forgotten that the data used above is almost 7-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 above can be of use for future entrepreneurs.