

---

# **Cousera Capstone Project Report**

**Report by: Hani Raza**

## **Business Problem Statement**

This project will analyse the New York City data. To start with, we will find the most visited commercial shop in New York according to the number of check-ins to that shop, then we will try to find the neighbourhoods that are lacking the selected type of shop which could be potential business opportunity and help the people to open shop in that area since these types of shops are not available.

## **Target Audience**

The target audience of this report is any one that is interested in opening a shop but have no idea what kind of shop should he open and in which neighbourhood.

## **Data Section**

The 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 a file in .txt 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

## Application

We will find the most visited type of shop (commercial) according to the number of check-ins given in the data, then we will try to find neighbourhoods that has none of this type of shop. Examples are for 2000 venues, and the red dot is the centre neighbourhood which has the most number of Bars between selected coordinates. We did find two neighbourhoods that are closest to it having none Bars within 4 kilometres.

## Methodology

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 (Neighborhoods are not included here)

Out[4]:

|      | 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 |
| ...  | ...                      | ...                 | ...           | ...                | ...                |
| 1995 | 49d7dbf5f964a5206c5d1fe3 | BBQ Joint           | 15            | 40.69570820924447  | -73.99105765689987 |
| 1996 | 4edae02c9911a3e78dd03897 | Burger Joint        | 12            | 40.893071          | -73.97342          |
| 1997 | 40d23080f964a5203e011fe3 | American Restaurant | 9             | 40.64147186279297  | -74.01447296142578 |
| 1998 | 4c5eddb07735c9b68ede9272 | Bagel Shop          | 1             | 40.705602          | -74.006405         |
| 1999 | 3fd66200f964a52053eb1ee3 | Hotel               | 14            | 40.642926465040816 | -74.01415535768307 |

2000 rows × 5 columns

```
In [5]: print('Venue count = ',len(venue_data['VenueID'].unique()))
print('Visitors count = ',sum(venue_data['Visitor Count']))
```

```
Venue count = 2000
Visitors count = 12248
```

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

```
rsnop', 16), ('General College & University', 16), ('Hardware Store', 16), ('Japanese Restaurant', 16), ('Latin American Restaurant', 15), ('Athletic & Sport', 15), ('Restaurant', 13), ('Spanish Restaurant', 12), ('Spa / Massage', 12), ('Middle Eastern Restaurant', 12), ('Malaysian Restaurant', 12), ('Record Shop', 12), ('Library', 12), ('Student Center', 11), ('Wings Joint', 11), ('High School', 11), ('Arts & Crafts Store', 10), ('Laundry Service', 10), ('Gas Station / Garage', 10), ('Asian Restaurant', 9), ('Burrito Place', 9), ('Parking', 9), ('Harbor / Marina', 9), ('Thai Restaurant', 9), ('Playground', 8), ('Campground', 8), ('Salad Place', 8), ('Event Space', 8), ('Pool Hall', 8), ('Ramen / Noodle House', 8), ('Automotive Shop', 7), ('Convenience Store', 7), ('Tea Room', 7), ('Post Office', 7), ('Indian Restaurant', 7), ('River', 7), ('Thrift / Vintage Store', 7), ('Paper / Office Supplies Store', 6), ('Cosmetics Shop', 6), ('Dessert Shop', 6), ('Museum', 6), ('Pet Store', 6), ('College & University', 6), ('Brazilian Restaurant', 6), ('Tanning Salon', 5), ('Bike Shop', 5), ('Art Gallery', 5), ('Arts & Entertainment', 5), ('Food', 5), ('Southern / Soul Food Restaurant', 4), ('Smoke Shop', 4), ('College Theater', 4), ('Snack Place', 4), ('Furniture / Home Store', 4), ('Sculpture Garden', 4), ('Mobile Phone Shop', 3), ('Tattoo Parlor', 3), ('Other Nightlife', 3), ('Moving Target', 3), ('Caribbean Restaurant', 3), ('Video Game Store', 3), ('Arcade', 3), ('Steakhouse', 3), ('Greek Restaurant', 3), ('Soup Place', 2), ('Music Store', 2), ('Dumpling Restaurant', 2), ('Performing Arts Venue', 2), ('Camera Store', 2), ('Housing Development', 2), ('Synagogue', 2), ('Mediterranean Restaurant', 2), ('African Restaurant', 2), ('Taxi', 2), ('Professional & Other Places', 2), ('Taco Place', 2), ('Jewelry Store', 1), ('Animal Shelter', 1), ('Factory', 1), ('Cemetery', 1), ('Medical School', 1), ('Pool', 1), ('Garden', 1), ('Hot Dog Joint', 1), ('Outdoors & Recreation', 1), ('Sorority House', 1), ('Casino', 1), ('Temple', 1), ('Historic Site', 1), ('Rest Area', 1), ('History Museum', 1), ('Recycling Facility', 1), ('Bridal Shop', 1), ('Nail Salon', 1), ('Nursery School', 1), ('Antique Shop', 1), ('South American Restaurant', 1), ('Law School', 1), ('Winery', 1)]
```

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

After all this, we will check the coordinates within given n number of kilometers 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.

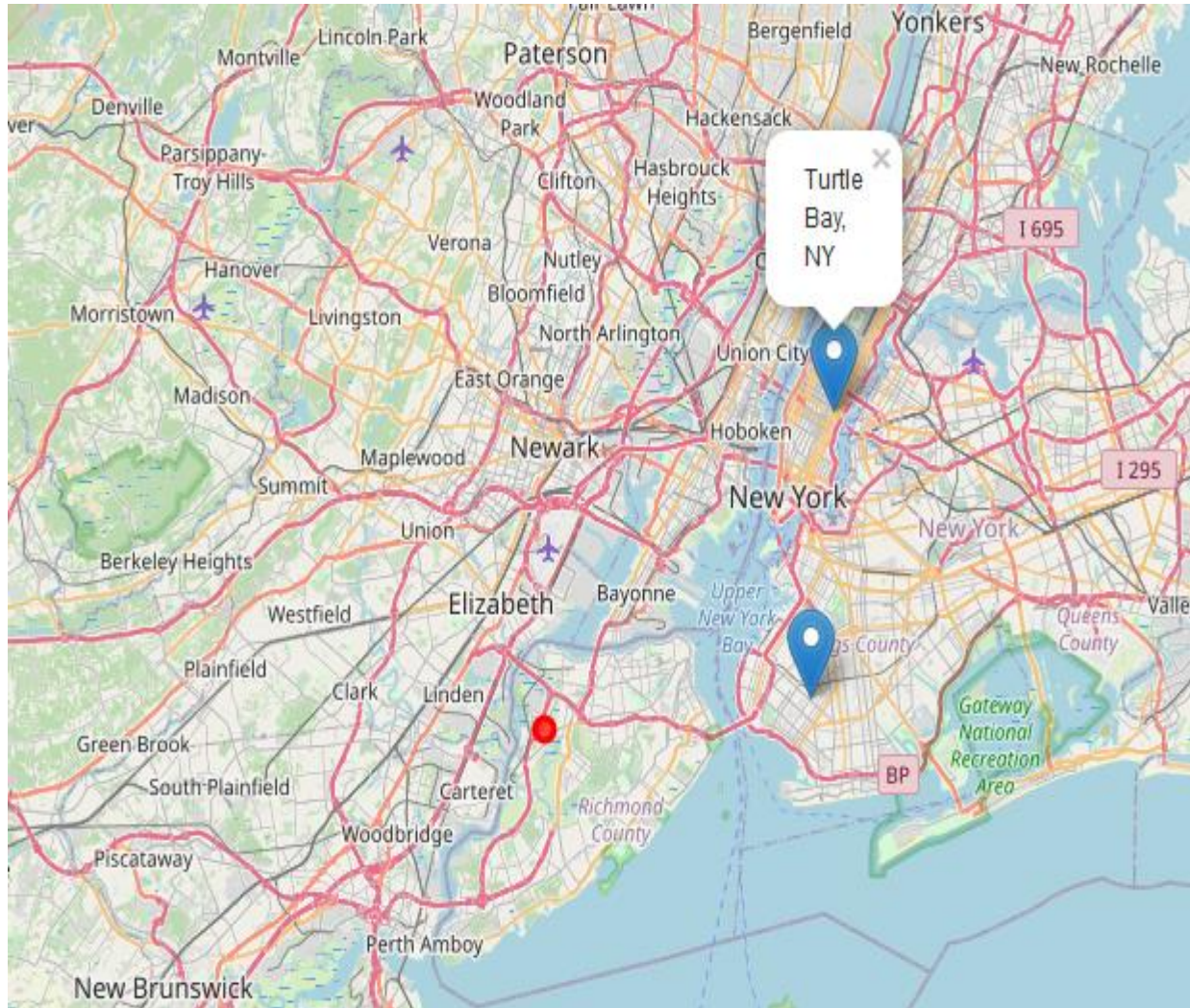
```
('40.60613336268842', '-74.17904376983643') : 2
('40.719810375488535', '-74.00258103213994') : 0
('40.60679958140643', '-74.04416981025437') : 0
('40.716161684843215', '-73.88307005845945') : 0
('40.69042711809854', '-73.95468677509598') : 0
('40.751591431346306', '-73.9741214009634') : 0
('40.61900594093755', '-73.99037472596906') : 0
('40.71976226666666', '-74.250014') : 0
('40.86198150306815', '-74.04790453737951') : 0
('40.82678953781387', '-73.94950923509141') : 0
('40.906627', '-73.777774') : 0
('40.73067679262482', '-74.06567180055883') : 0
('40.64531729239498', '-73.77383708953857') : 0
('40.79059946897114', '-73.98023377661316') : 0
('40.655535144394925', '-74.00862937888984') : 0
('40.86283581665962', '-74.19723987579346') : 0
('40.901057866884024', '-74.1507625579834') : 0
('40.963240757086346', '-74.09463109843134') : 0
('40.77615805031661', '-73.82361593360073') : 0
('40.92431190858341', '-73.99688829591487') : 0
('40.786713', '-74.175476') : 0
('40.74710920467287', '-74.15280867121942') : 0
('40.771046', '-74.065758') : 0
('40.725891', '-73.79143081000001') : 0
('40.83162208784265', '-74.13679361343384') : 0
('40.828602195433966', '-73.87925863265991') : 0
('40.60014371940161', '-73.94659322025349') : 0
('40.8706300946324', '-74.09792627389562') : 0
('40.677558813432675', '-73.74452479795166') : 0
('40.82358025225002', '-74.22280201227520') : 0
```

Find the two neighborhoods that are closest to the coordinate which has the most number of the specific shop type but lacking that within 4 kilometers

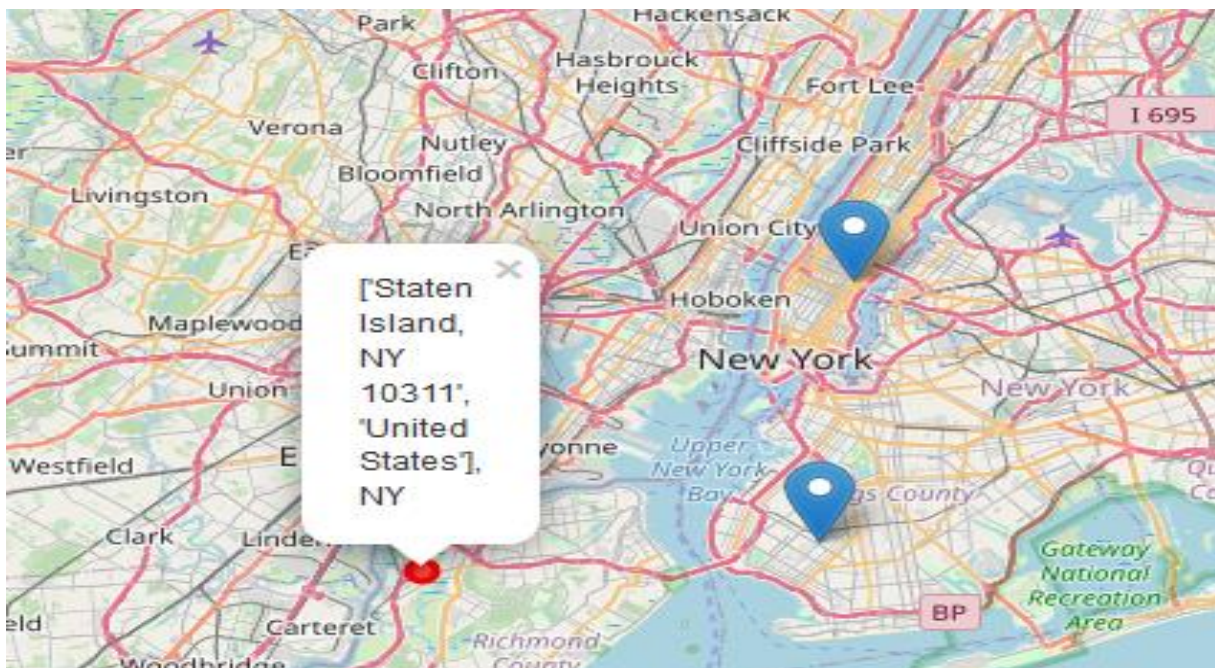
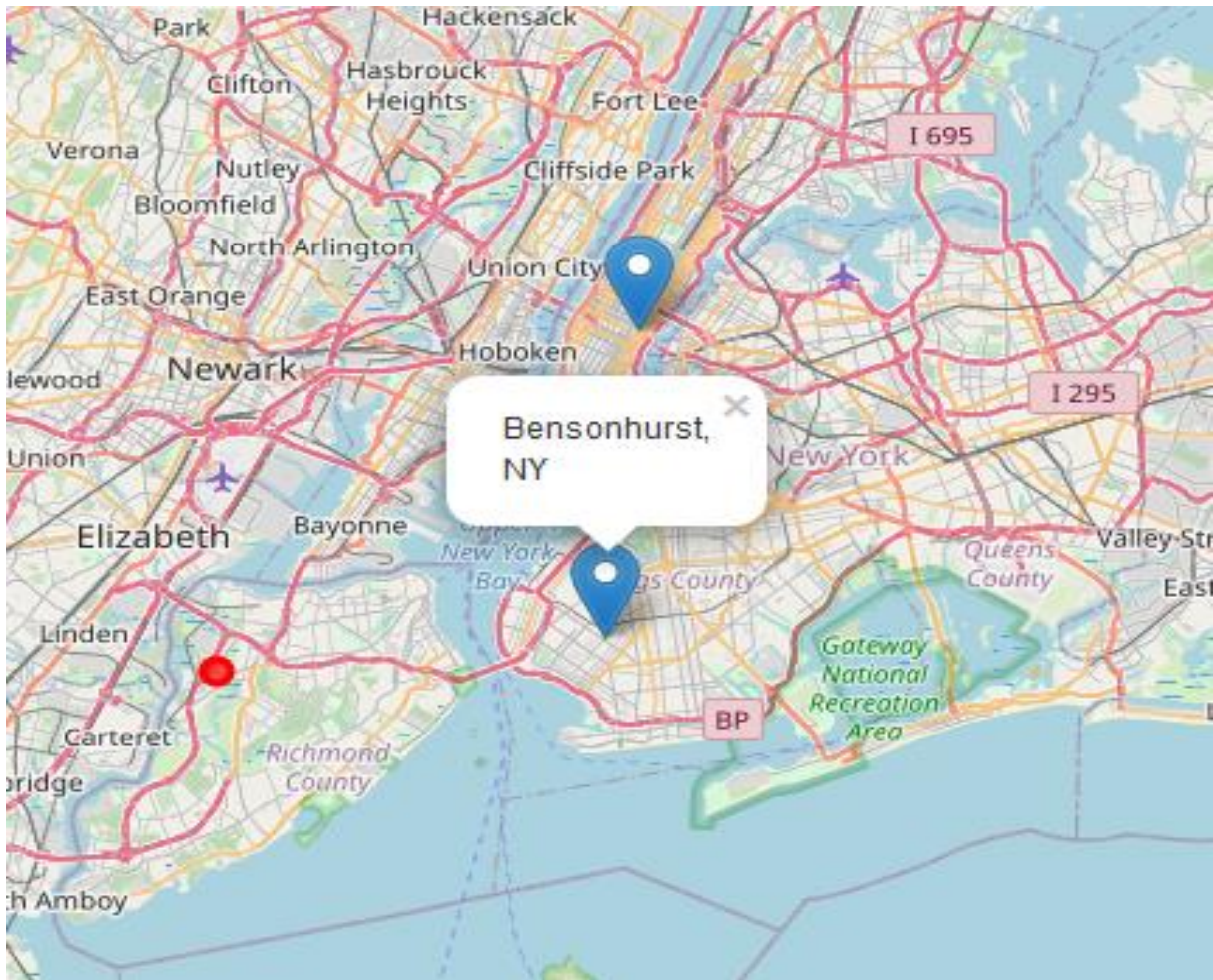
```
for ne in neighborhoods:  
    print(ne)
```

Bensonhurst  
Turtle Bay

### **Visualization:**







Red dot is the center

---

## **Results & 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 above can be of use for future entrepreneurs.