1. Introduction/Business Understanding

1.1 Description of the problem and the background

As the transportation around the world is becoming more convenient, there are increasing number of people considering immigrate out to other countries. Some are trying to make investment, some are trying to study aboard, however, some are trying to start their own business!

Recently, we could see Trump is helping Hong Kong to voice out the right of freedom. If we would be given chances to live in US, a place full of freedom, what shall we, Hongkongers could do to sustain the daily life and contribute to the country?

So, this Capstone Project aims to discover what are the right business that could be started in the right place. This time, we would wish to research on Brooklyn.

2. Data Requirements

For this project we need following data:

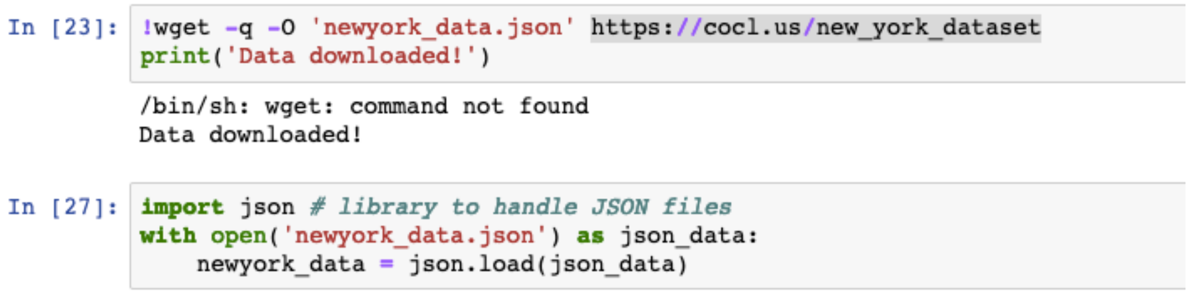
1. New York dataset
   Datasource : https://cocl.us/new_york_dataset

2. Venue nearby Brooklyn
   Datasource : Four Square API

3. Methodology

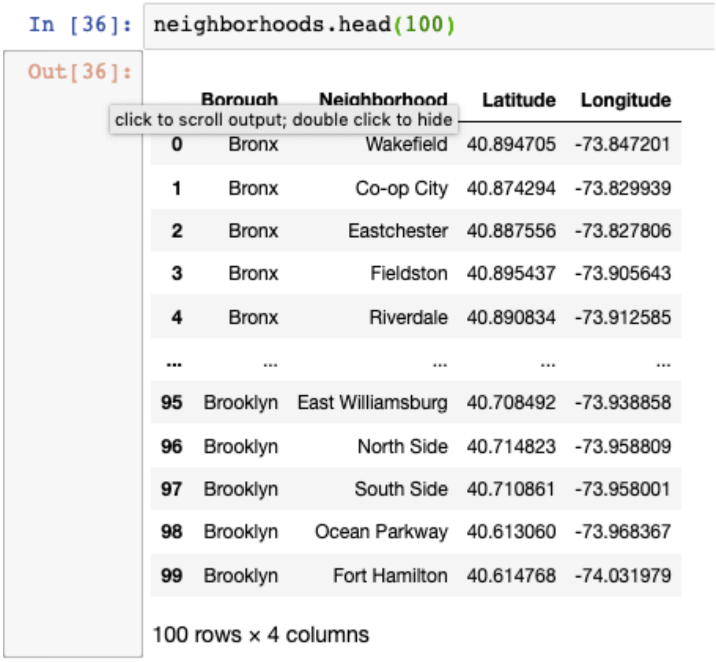
3.1 Data Preparation

3.1.1 Scraping New York Data from COCL



```
In [23]: !wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset
         print('Data downloaded!')

         /bin/sh: wget: command not found
         Data downloaded!

In [27]: import json # library to handle JSON files
         with open('newyork_data.json') as json_data:
             newyork_data = json.load(json_data)
```

After little manipulation, the data-frame is obtained as below:



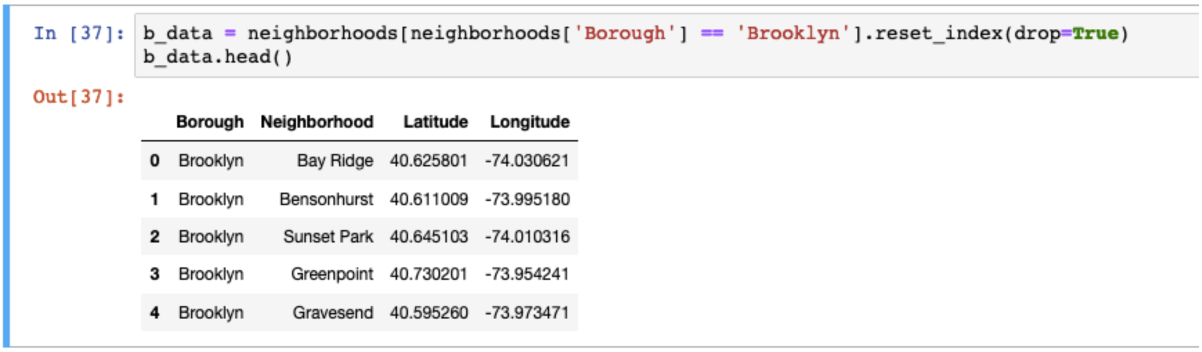
```
In [36]: neighborhoods.head(100)
```

Out[36]:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
...	...	...	...	...
95	Brooklyn	East Williamsburg	40.708492	-73.938858
96	Brooklyn	North Side	40.714823	-73.958809
97	Brooklyn	South Side	40.710861	-73.958001
98	Brooklyn	Ocean Parkway	40.613060	-73.968367
99	Brooklyn	Fort Hamilton	40.614768	-74.031979

100 rows × 4 columns

3.1.2 Getting the Neighborhood of Brooklyn

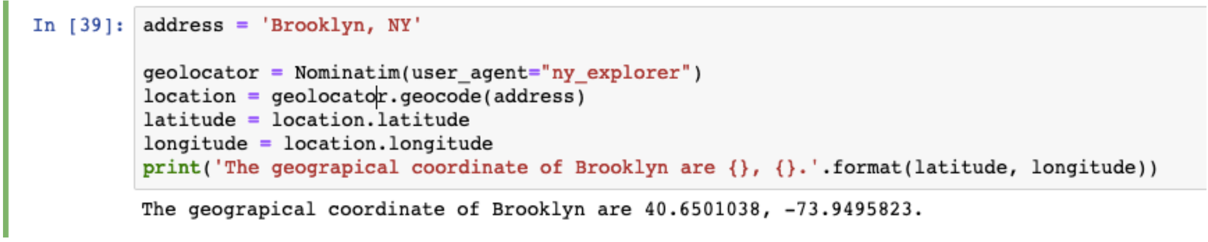


```
In [37]: b_data = neighborhoods[neighborhoods['Borough'] == 'Brooklyn'].reset_index(drop=True)
         b_data.head()
```

Out[37]:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

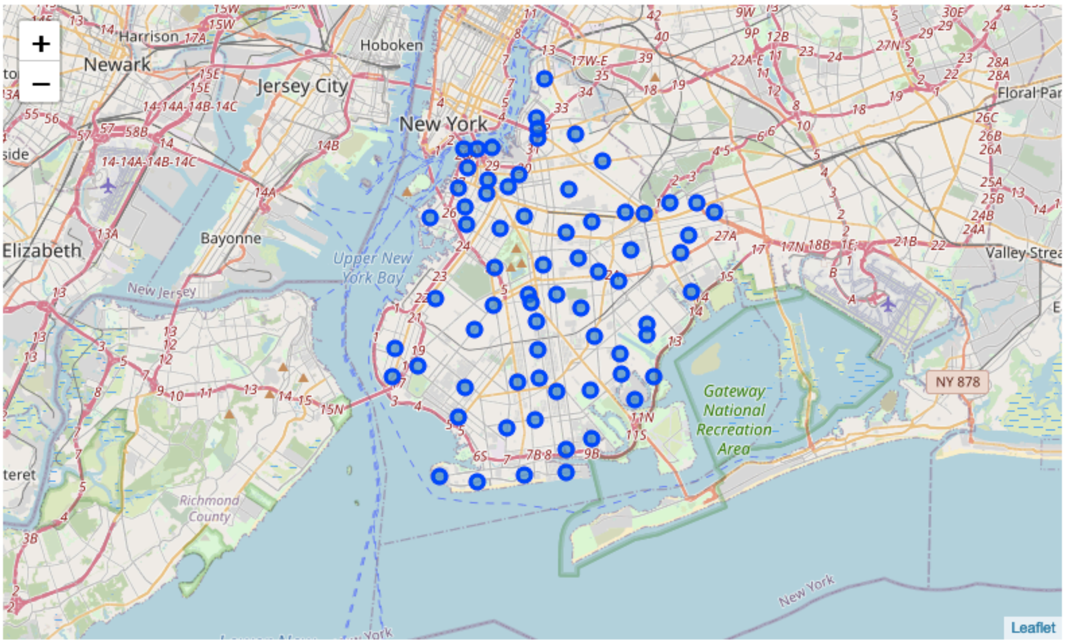
As we are interested in Brooklyn, the geographical coordinate of Brooklyn is needed.



```
In [39]: address = 'Brooklyn, NY'

         geolocator = Nominatim(user_agent="ny_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of Brooklyn are {}, {}.'.format(latitude, longitude))

         The geograpical coordinate of Brooklyn are 40.6501038, -73.9495823.
```

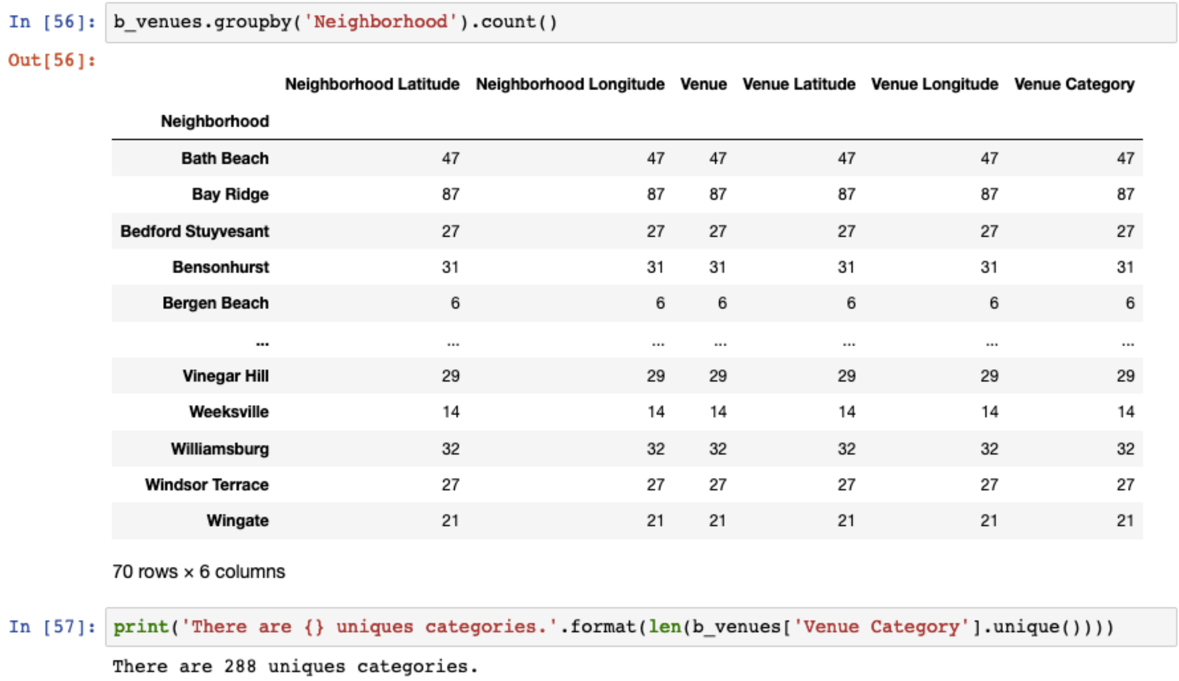
I used python **folium** library to visualize geographic details of the neighbourhood. I used latitude and longitude values to get the visual as below:

Out[41]:



3.2. Exploratory Data Analysis:

3.2.1 Using Foursquare Location Data



```
In [56]: b_venues.groupby('Neighborhood').count()
```

Out[56]:

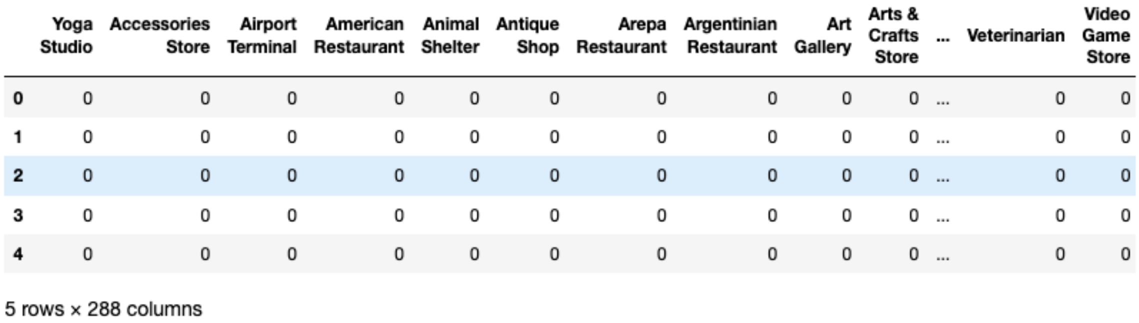
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bath Beach	47	47	47	47	47	47
Bay Ridge	87	87	87	87	87	87
Bedford Stuyvesant	27	27	27	27	27	27
Bensonhurst	31	31	31	31	31	31
Bergen Beach	6	6	6	6	6	6
...	...	...	...	...	...	...
Vinegar Hill	29	29	29	29	29	29
Weeksville	14	14	14	14	14	14
Williamsburg	32	32	32	32	32	32
Windsor Terrace	27	27	27	27	27	27
Wingate	21	21	21	21	21	21

70 rows × 6 columns

```
In [57]: print('There are {} uniques categories.'.format(len(b_venues['Venue Category'].unique())))

         There are 288 uniques categories.
```

There are 288 unique categories.

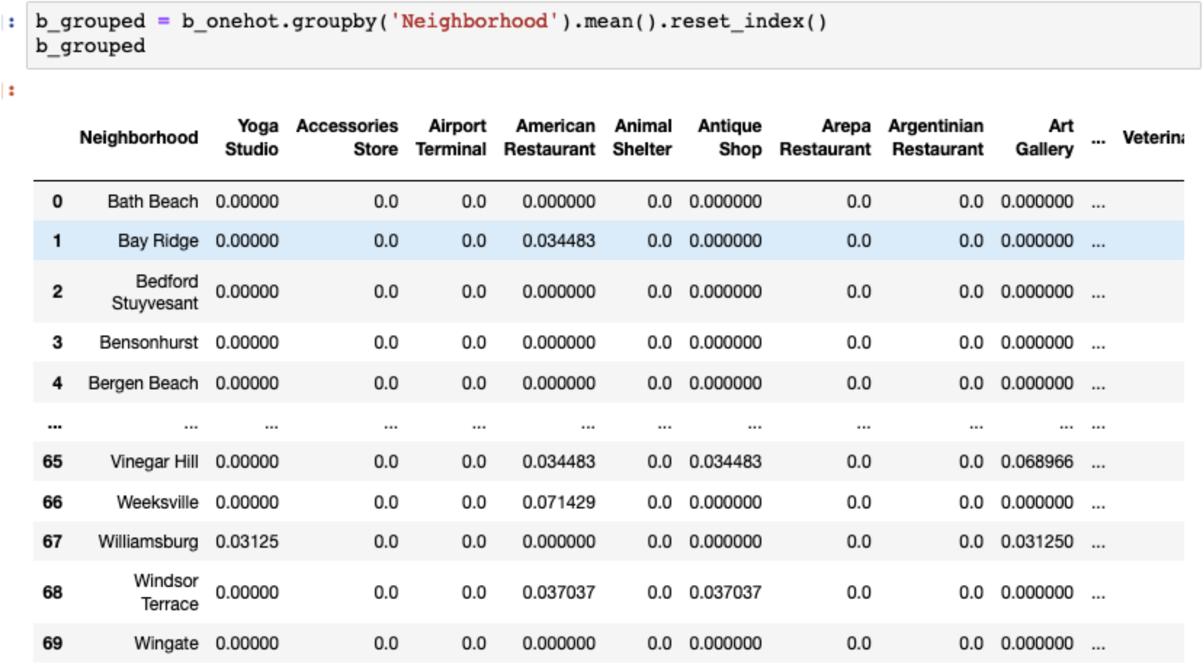
So, create a data-frame with pandas one hot encoding for the venue categories.



	Yoga Studio	Accessories Store	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	...	Veterinarian	Video Game Store
0	0	0	0	0	0	0	0	0	0	0	...	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0

5 rows × 288 columns

Use pandas groupby on neighborhood column and calculate the mean of the frequency of occurrence of each venue category.



```
b_grouped = b_onehot.groupby('Neighborhood').mean().reset_index()
b_grouped
```

	Neighborhood	Yoga Studio	Accessories Store	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Veterin
0	Bath Beach	0.00000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	...	
1	Bay Ridge	0.00000	0.0	0.0	0.034483	0.0	0.000000	0.0	0.0	0.000000	...	
2	Bedford Stuyvesant	0.00000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	...	
3	Bensonhurst	0.00000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	...	
4	Bergen Beach	0.00000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	...	
...	...	...	...	...	...	...	...	...	...	...	...	
65	Vinegar Hill	0.00000	0.0	0.0	0.034483	0.0	0.034483	0.0	0.0	0.068966	...	
66	Weeksville	0.00000	0.0	0.0	0.071429	0.0	0.000000	0.0	0.0	0.000000	...	
67	Williamsburg	0.03125	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.031250	...	
68	Windsor Terrace	0.00000	0.0	0.0	0.037037	0.0	0.037037	0.0	0.0	0.000000	...	
69	Wingate	0.00000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	...	

Output each neighborhood along with the top 5 most common venues:

```
----Bath Beach----
                  venue  freq
0           Pizza Place  0.09
1    Chinese Restaurant  0.09
2              Pharmacy  0.06
3         Bubble Tea Shop  0.04
4   Fast Food Restaurant  0.04


----Bay Ridge----
                 venue  freq
0                  Spa  0.07
1   Italian Restaurant  0.07
2          Pizza Place  0.05
3     Greek Restaurant  0.05
4  American Restaurant  0.03


----Bedford Stuyvesant----
           venue  freq
0  Deli / Bodega  0.07
```

```
1    Pizza Place  0.07
2    Coffee Shop  0.07
3           Café  0.07
4            Bar  0.07


----Bensonhurst----
                venue  freq
0  Chinese Restaurant  0.10
1    Sushi Restaurant  0.06
2          Donut Shop  0.06
3      Ice Cream Shop  0.06
4  Italian Restaurant  0.06


----Bergen Beach----
                venue  freq
0     Harbor / Marina  0.33
1  Athletics & Sports  0.17
2          Playground  0.17
3      Baseball Field  0.17
4          Donut Shop  0.17


----Boerum Hill----
            venue  freq
0     Coffee Shop  0.05
1    Dance Studio  0.05
2             Bar  0.04
3  Sandwich Place  0.03
4          Bakery  0.03


----Borough Park----
                  venue  freq
0                  Bank  0.21
1           Pizza Place  0.16
2  Fast Food Restaurant  0.11
3              Pharmacy  0.11
4                 Hotel  0.05


----Brighton Beach----
                        venue  freq
0          Russian Restaurant  0.07
1                       Beach  0.07
2                  Restaurant  0.07
3  Eastern European Restaurant  0.07
4                        Bank  0.05


----Broadway Junction----
                 venue  freq
0  Fried Chicken Joint  0.11
1           Donut Shop  0.11
2                Diner  0.11
3            Nightclub  0.06
4          Gas Station  0.06


----Brooklyn Heights----
```

```
                venue  freq
0       Deli / Bodega  0.05
1        Yoga Studio  0.04
2               Park  0.04
3        Pizza Place  0.04
4  Mexican Restaurant  0.03


----Brownsville----
                venue  freq
0          Restaurant  0.16
1  Chinese Restaurant  0.11
2                Park  0.11
3   Convenience Store  0.05
4            Pharmacy  0.05


----Bushwick----
                    venue  freq
0                     Bar  0.09
1      Mexican Restaurant  0.07
2             Coffee Shop  0.07
3           Deli / Bodega  0.07
4  Thrift / Vintage Store  0.04


----Canarsie----
                  venue  freq
0                   Gym   0.2
1      Asian Restaurant   0.2
2  Caribbean Restaurant   0.2
3                  Food   0.2
4              Bus Line   0.2


----Carroll Gardens----
                venue  freq
0  Italian Restaurant  0.11
1         Coffee Shop  0.07
2         Pizza Place  0.05
3        Cocktail Bar  0.04
4              Bakery  0.04


----City Line----
                 venue  freq
0           Donut Shop  0.08
1  Fried Chicken Joint  0.06
2        Grocery Store  0.06
3    Mobile Phone Shop  0.06
4         Liquor Store  0.06


----Clinton Hill----
                venue  freq
0         Pizza Place  0.06
1  Italian Restaurant  0.05
2           Wine Shop  0.04
3  Mexican Restaurant  0.04
4     Thai Restaurant  0.04
```

```
----Cobble Hill----
         venue  freq
0   Playground  0.04
1  Pizza Place  0.04
2  Coffee Shop  0.04
3          Bar  0.04
4  Yoga Studio  0.03


----Coney Island----
                          venue  freq
0          Caribbean Restaurant  0.15
1              Baseball Stadium  0.15
2                       Brewery  0.08
3  Theme Park Ride / Attraction  0.08
4                  Dessert Shop  0.08


----Crown Heights----
         venue  freq
0  Pizza Place  0.14
1         Café  0.09
2       Museum  0.09
3     Pharmacy  0.05
4  Coffee Shop  0.05


----Cypress Hills----
                       venue  freq
0        Fried Chicken Joint  0.13
1  Latin American Restaurant  0.10
2                 Donut Shop  0.07
3             Ice Cream Shop  0.07
4       Fast Food Restaurant  0.07


----Ditmas Park----
                  venue  freq
0  Caribbean Restaurant  0.07
1          Burger Joint  0.05
2            Donut Shop  0.05
3    Chinese Restaurant  0.05
4        Clothing Store  0.05


----Downtown----
                   venue  freq
0           Burger Joint  0.05
1            Pizza Place  0.05
2            Coffee Shop  0.05
3         Sandwich Place  0.04
4  Performing Arts Venue  0.02


----Dumbo----
            venue  freq
0            Park  0.08
1  Scenic Lookout  0.07
2     Coffee Shop  0.07
3     Yoga Studio  0.03
```

```
4             Gym  0.03


----Dyker Heights----
            venue  freq
0    Burger Joint   0.2
1      Bagel Shop   0.2
2  Cosmetics Shop   0.2
3     Golf Course   0.2
4   Grocery Store   0.2


----East Flatbush----
               venue  freq
0         Print Shop  0.08
1           Pharmacy  0.08
2   Department Store  0.08
3      Moving Target  0.08
4  Food & Drink Shop  0.08


----East New York----
                 venue  freq
0        Deli / Bodega  0.14
1          Bus Station  0.14
2                  Gym  0.07
3             Pharmacy  0.07
4  Fried Chicken Joint  0.07


----East Williamsburg----
           venue  freq
0            Bar  0.10
1  Deli / Bodega  0.07
2   Cocktail Bar  0.07
3         Bakery  0.05
4    Coffee Shop  0.05


----Erasmus----
                  venue  freq
0  Caribbean Restaurant  0.18
1     Convenience Store  0.05
2     Mobile Phone Shop  0.05
3    Chinese Restaurant  0.05
4        Sandwich Place  0.05


----Flatbush----
                  venue  freq
0    Mexican Restaurant  0.10
1              Pharmacy  0.10
2           Coffee Shop  0.10
3  Caribbean Restaurant  0.10
4            Donut Shop  0.05


----Flatlands----
                  venue  freq
0              Pharmacy  0.16
1   Fried Chicken Joint  0.11
```

```
2  Caribbean Restaurant  0.11
3  Fast Food Restaurant  0.11
4     Electronics Store  0.05


----Fort Greene----
                venue  freq
0         Flower Shop  0.05
1         Pizza Place  0.05
2           Wine Shop  0.05
3  Italian Restaurant  0.05
4          Playground  0.03


----Fort Hamilton----
                venue  freq
0         Pizza Place  0.05
1  Italian Restaurant  0.05
2       Deli / Bodega  0.05
3      Sandwich Place  0.05
4  Chinese Restaurant  0.05


----Fulton Ferry----
                 venue  freq
0                 Park  0.14
1  American Restaurant  0.05
2       Scenic Lookout  0.05
3               Bridge  0.04
4          Pizza Place  0.04


----Georgetown----
                venue  freq
0                Bank  0.10
1  Italian Restaurant  0.07
2            Pharmacy  0.07
3          Donut Shop  0.07
4      Shipping Store  0.03


----Gerritsen Beach----
             venue  freq
0      Pizza Place  0.10
1              Bar  0.10
2   Ice Cream Shop  0.10
3      Event Space  0.05
4  Harbor / Marina  0.05


----Gowanus----
                    venue  freq
0      Italian Restaurant  0.06
1                     Bar  0.06
2  Furniture / Home Store  0.06
3              Food Truck  0.04
4             Pizza Place  0.04


----Gravesend----
                venue  freq
```

```
0  Italian Restaurant  0.12
1         Pizza Place  0.12
2  Chinese Restaurant  0.08
3              Bakery  0.08
4              Lounge  0.08


----Greenpoint----
          venue  freq
0           Bar  0.09
1   Pizza Place  0.08
2  Cocktail Bar  0.06
3   Coffee Shop  0.06
4   Yoga Studio  0.03


----Highland Park----
                  venue  freq
0         Grocery Store  0.08
1                Garden  0.08
2  Gym / Fitness Center  0.08
3                  Park  0.08
4         Big Box Store  0.08


----Homecrest----
                venue  freq
0                Bank  0.10
1          Donut Shop  0.08
2         Pizza Place  0.05
3  Chinese Restaurant  0.05
4    Sushi Restaurant  0.05


----Kensington----
             venue  freq
0  Thai Restaurant  0.09
1    Grocery Store  0.09
2   Ice Cream Shop  0.06
3   Sandwich Place  0.06
4      Pizza Place  0.06


----Madison----
                venue  freq
0          Bagel Shop   0.2
1       Deli / Bodega   0.1
2                 Spa   0.1
3        Dessert Shop   0.1
4  Italian Restaurant   0.1


----Manhattan Beach----
            venue  freq
0        Bus Stop  0.18
1            Café  0.18
2  Ice Cream Shop  0.09
3            Food  0.09
4           Beach  0.09
```

```
----Manhattan Terrace----
            venue  freq
0     Pizza Place  0.15
1  Ice Cream Shop  0.11
2      Donut Shop  0.11
3        Pharmacy  0.07
4  Cosmetics Shop  0.04


----Marine Park----
              venue  freq
0     Deli / Bodega   0.1
1       Pizza Place   0.1
2      Soccer Field   0.1
3              Park   0.1
4  Basketball Court   0.1


----Midwood----
               venue  freq
0        Pizza Place   0.4
1     Ice Cream Shop   0.1
2             Bakery   0.1
3        Candy Store   0.1
4  Convenience Store   0.1


----Mill Basin----
                 venue  freq
0   Chinese Restaurant  0.11
1  Japanese Restaurant  0.08
2          Pizza Place  0.08
3             Pharmacy  0.06
4                 Bank  0.06


----Mill Island----
                   venue  freq
0                   Pool   1.0
1            Yoga Studio   0.0
2            Opera House   0.0
3           Outlet Store   0.0
4  Outdoors & Recreation   0.0


----New Lots----
                 venue  freq
0             Pharmacy  0.10
1          Pizza Place  0.10
2  Fried Chicken Joint  0.10
3          Bus Station  0.05
4       Breakfast Spot  0.05


----North Side----
         venue  freq
0  Coffee Shop  0.09
1  Pizza Place  0.06
2          Bar  0.05
3  Yoga Studio  0.04
4     Wine Bar  0.04
```

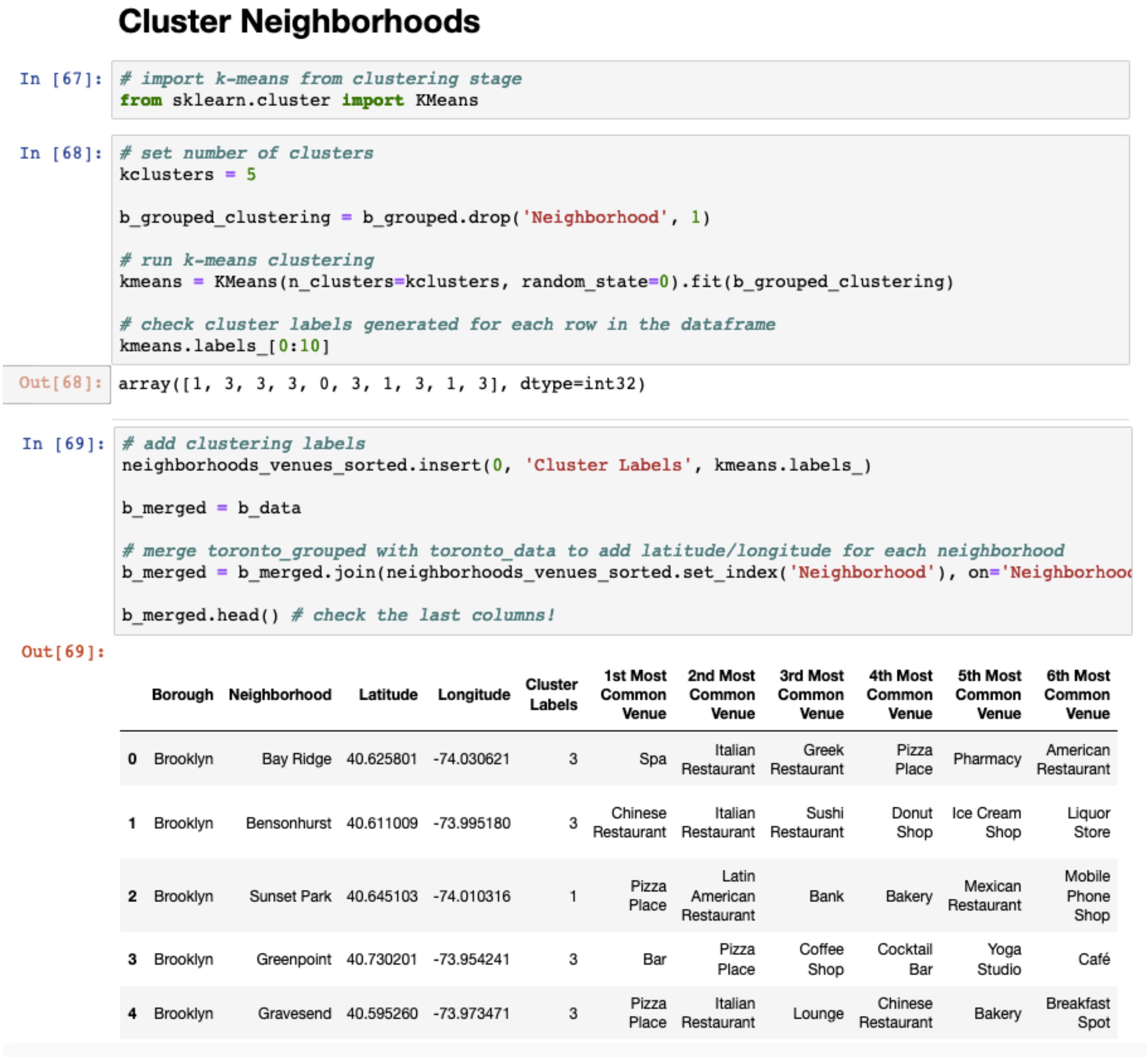
```
----Ocean Hill----
                             venue  freq
0                    Deli / Bodega  0.14
1                Convenience Store  0.07
2                    Grocery Store  0.07
3  Southern / Soul Food Restaurant  0.07
4                      Supermarket  0.07


----Ocean Parkway----
          venue  freq
0   Pizza Place  0.09
1  Liquor Store  0.09
2    Restaurant  0.09
3  Dessert Shop  0.05
4    Bagel Shop  0.05


----Paerdegat Basin----
                venue  freq
0  Child Care Service  0.14
1         Auto Garage  0.14
2       Moving Target  0.14
3                Food  0.14
4            Bus Line  0.14


----Park Slope----
                 venue  freq
0          Coffee Shop  0.08
1         Burger Joint  0.07
2  American Restaurant  0.05
3               Bakery  0.03
4            Bookstore  0.03


----Prospect Heights----
                venue  freq
0                 Bar  0.09
1  Mexican Restaurant  0.07
2     Thai Restaurant  0.04
3         Cocktail Bar  0.04
4        Gourmet Shop  0.04


----Prospect Lefferts Gardens----
                  venue  freq
0                  Café  0.08
1           Pizza Place  0.08
2                Bakery  0.08
3  Caribbean Restaurant  0.06
4         Deli / Bodega  0.04


----Prospect Park South----
                  venue  freq
0  Caribbean Restaurant  0.10
1  Fast Food Restaurant  0.06
2           Pizza Place  0.06
```

```
3         Grocery Store  0.06
4     Mobile Phone Shop  0.06


----Red Hook----
                venue  freq
0  Seafood Restaurant  0.08
1         Art Gallery  0.08
2                 Bar  0.06
3                Park  0.06
4                Farm  0.04


----Remsen Village----
                  venue  freq
0  Caribbean Restaurant  0.17
1  Fast Food Restaurant  0.11
2                   Gym  0.06
3           Supermarket  0.06
4                   Spa  0.06


----Rugby----
                  venue  freq
0         Grocery Store  0.12
1  Caribbean Restaurant  0.12
2                  Bank  0.12
3                 Diner  0.06
4           Pizza Place  0.06


----Sea Gate----
         venue  freq
0  Sports Club  0.25
1          Spa  0.25
2        Beach  0.25
3  Bus Station  0.25
4  Yoga Studio  0.00


----Sheepshead Bay----
                venue  freq
0  Turkish Restaurant  0.12
1        Dessert Shop  0.12
2      Sandwich Place  0.08
3         Yoga Studio  0.04
4  Italian Restaurant  0.04


----South Side----
                 venue  freq
0          Coffee Shop  0.07
1                  Bar  0.07
2          Pizza Place  0.05
3  American Restaurant  0.05
4          Yoga Studio  0.03


----Starrett City----
         venue  freq
0     Pharmacy   0.2
```

```
1   Donut Shop   0.1
2  Pizza Place   0.1
3     Gym Pool   0.1
4     Bus Stop   0.1


----Sunset Park----
                       venue  freq
0                     Bakery  0.09
1         Mexican Restaurant  0.09
2                Pizza Place  0.09
3  Latin American Restaurant  0.09
4                       Bank  0.09


----Vinegar Hill----
         venue  freq
0   Food Truck  0.17
1  Coffee Shop  0.10
2         Café  0.07
3  Art Gallery  0.07
4         Park  0.03


----Weeksville----
                venue  freq
0      Discount Store  0.14
1          Donut Shop  0.07
2                Café  0.07
3  Chinese Restaurant  0.07
4       Grocery Store  0.07


----Williamsburg----
         venue  freq
0  Coffee Shop  0.09
1          Bar  0.06
2   Bagel Shop  0.06
3  Yoga Studio  0.03
4  Event Space  0.03


----Windsor Terrace----
           venue  freq
0  Grocery Store  0.07
1           Park  0.07
2          Diner  0.07
3           Café  0.07
4          Plaza  0.07


----Wingate----
                 venue  freq
0  Fried Chicken Joint  0.10
1        Deli / Bodega  0.05
2               Bakery  0.05
3            Juice Bar  0.05
4         Liquor Store  0.05
```

4. Analysis

I will use *clustering* (KMeans).



Cluster Neighborhoods

In [67]:
```
# import k-means from clustering stage
from sklearn.cluster import KMeans
```

In [68]:
```
# set number of clusters
kclusters = 5

b_grouped_clustering = b_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(b_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[68]:
```
array([1, 3, 3, 3, 0, 3, 1, 3, 1, 3], dtype=int32)
```

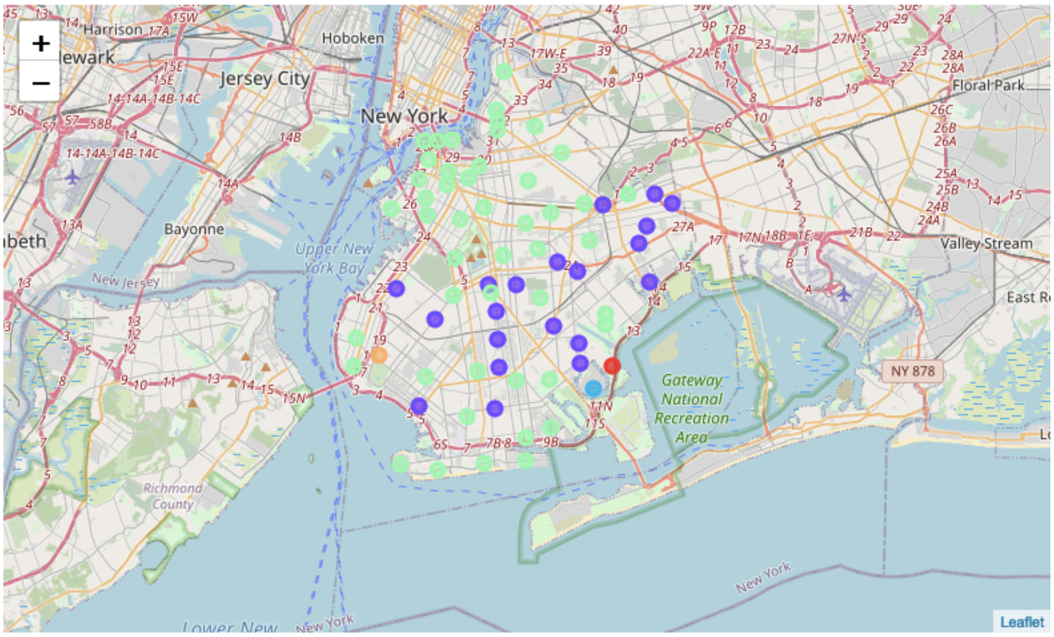
In [69]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

b_merged = b_data

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
b_merged = b_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhoo

b_merged.head() # check the last columns!
```

Out[69]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	3	Spa	Italian Restaurant	Greek Restaurant	Pizza Place	Pharmacy	American Restaurant
1	Brooklyn	Bensonhurst	40.611009	-73.995180	3	Chinese Restaurant	Italian Restaurant	Sushi Restaurant	Donut Shop	Ice Cream Shop	Liquor Store
2	Brooklyn	Sunset Park	40.645103	-74.010316	1	Pizza Place	Latin American Restaurant	Bank	Bakery	Mexican Restaurant	Mobile Phone Shop
3	Brooklyn	Greenpoint	40.730201	-73.954241	3	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Yoga Studio	Café
4	Brooklyn	Gravesend	40.595260	-73.973471	3	Pizza Place	Italian Restaurant	Lounge	Chinese Restaurant	Bakery	Breakfast Spot

We can represent these 5 clusters in a leaflet map using Folium library as below:

```
Out[72]:
```



5. Results & discussion

We could see the first, third and fifth cluster only contain one neighbourhood. Both of them are neither catering nor services.

For the first cluster, the neighbourhood is Bergen Beach. The common venue are Harbor, Playground etc. Therefor if you are moving in to this area, you may consider to start business on catering or continence store.

For the third cluster, the neighbourhood is Mill Island. The most common venue are Pool and Women’s Store. Therefore if you are moving in to this area, you may consider to start business on selling swimming products and bikini.

For the fifth cluster, the neighbourhood is Dyker Heights. The common venue are Golf Course and Bagel Shop. Therefore if you are moving in to this area, you may consider to start business on selling sunscreen products and hats.

6. Conclusion

Freedom is not free. Be brave to take risk if you are moving to new places and restarting your new life.

Although I had some suggestion above, they might not always be true. You shall still search for new ideas and go ahead for it. Data is everywhere around us, and data could help in solving many problems.

Also, one dataset could help in solving multiple problems, depends on how you use it, understand it and analysis it.