Final Report Week 5 Capstone project

Introduction

In this project I'm going to use the location data to explore a geographical location of New-York.I will have the opportunity to use the Foursquare location data to explore, solve and compare neighborhoods of New-York.

During the last decade, New-York have seen an increase number of population. This population come in one of the largest city and one of the most crowded city for different reasons, it can be personally, professionally or as a tourist to visit this "business city". New-York have a very hight number of population per km square. During the the 19 century until today, New-York had welcome a lot of immigrants from all around the world.

It help this city to be one of the largest business city in the world and one of a city where people who work are from all the sides of the world, from Asia, Europe or Pacific. This mixity of culture can be seen in Manhattan as different cluster, as different neighborhoods or boroughs. We can see for example Chinatown, Italian, Indian street or neighborhoods. Even English is the main spoken language many other foreign people speak their own native languages like Chinese, Italian, Indian and other between friends, family and sometime in a company.

The differences of culture can also be seen in the food, I want to said that food bring people together, it's a time that we can share our culture, ideas, thinking with other people. It's great to learn and meet new people from all around the world and learn and see the differences with our own cultures.

In this project, we'll see the different part of New-York, the different cluster, boroughs mainly situated in Manhattan and his areas. We'll see the different neighbourhoods and see what is the main thing that people can do there, like shop, restaurant, coffee or hotel in this area.

For this week, I will submit the following :

- A full report consisting of all of the following components (15 marks):

- Introduction where you discuss the business problem and who would be interested in this project.

- Data where you describe the data that will be used to solve the problem and the source of the data.

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

- Results section where you discuss the results.

- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

- Conclusion section where you conclude the report.

Explore Dataset

- This dataset has a total of 4 boroughs and 306 neighborhoods.

- In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

- This dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

- I downloaded the files and placed it on the server, so you can simply run a wget command and access the data. So let's go ahead and do that

The geograpical coordinate of New York City are 40.7127281, -74.0060152.
This dataframe have 306 neighborhoods and 5 borough. Each neighborhoods have a coordin ates a latitude and longitude.

[118]:

	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

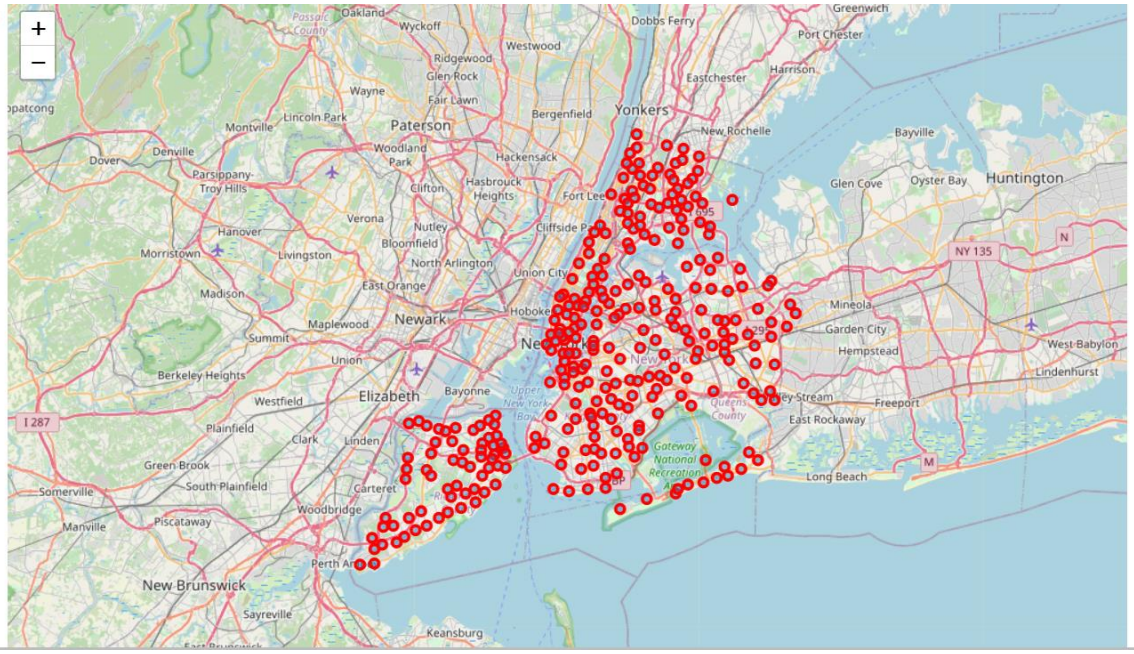
So here we'll focus on New-York city.

Here I create map of New York with neighbourhoods superimposed on the top in red.

Each circle in red is a neighbourhood, we can see in Manhattan there is around 150.

To display this map, I use the library folio, it's a visualization library, where is possible to zoom, click on the map. Many think that is possible to do on a normal maps.

[123]:



Here we can have an example of a few neighbourhood in Manhattan.

[14]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Here a screen that focus only of Manhattan

In this part I have clean the json data file into a pandas dataframe. For sure before we need to import the pandas library. We can see in the list different areas like restaurants, hotel, bakery and even some sandwiches. Each name have also a categories that explain what types of place it is and each of them have coordinate (latitude and longitude).

[135]:

	name	categories	lat	lng
0	Cheeky Sandwiches	Sandwich Place	40.715821	-73.991830
1	Kiki's	Greek Restaurant	40.714476	-73.992036
2	Hotel 50 Bowery NYC	Hotel	40.715936	-73.996789
3	Renew Day Spa	Spa	40.715559	-73.996747
4	Michaeli Bakery	Bakery	40.714704	-73.991847

In this part we will explore more in details the Manhattan neighbourhoods.

This picture below represents a dataframe, this dataframe have 3077 neighborhoods and 7 attributes that includes the Venue, neiborhoods, venue category, latitude and longitude.

```
Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea
Greenwich Village
East Village
Lower East Side
Tribeca
```

```
(3077, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.910660	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.910660	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.910660	Dunkin'	40.877136	-73.906666	Donut Shop
5	Marble Hill	40.876551	-73.910660	Rite Aid	40.875467	-73.908906	Pharmacy
6	Marble Hill	40.876551	-73.910660	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium
7	Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
8	Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop
9	Marble Hill	40.876551	-73.910660	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
10	Marble Hill	40.876551	-73.910660	Blink Fitness	40.877271	-73.905595	Gym
11	Marble Hill	40.876551	-73.910660	Vitamin Shoppe	40.877160	-73.905632	Supplement Shop
12	Marble Hill	40.876551	-73.910660	Parrilla Latina	40.877473	-73.906073	Steakhouse
13	Marble Hill	40.876551	-73.910660	T.J. Maxx	40.877232	-73.905042	Department Store
14	Marble Hill	40.876551	-73.910660	Boston Market	40.877430	-73.905412	American Restaurant

Here, all this 3077 neighborhoods that we get before are group by ‘Neighborhood’. After that we finally get 327 uniques categories. The picture below show us some example of categories.

```
manhattan_venues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	67	67	67	67	67	67
Carnegie Hill	85	85	85	85	85	85
Central Harlem	46	46	46	46	46	46
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Civic Center	96	96	96	96	96	96
Clinton	100	100	100	100	100	100
East Harlem	42	42	42	42	42	42
East Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100
Flatiron	99	99	99	99	99	99
Gramercy	85	85	85	85	85	85
Greenwich Village	100	100	100	100	100	100
Hamilton Heights	57	57	57	57	57	57
Hudson Yards	55	55	55	55	55	55
Inwood	57	57	57	57	57	57
Lenox Hill	100	100	100	100	100	100
Lincoln Square	94	94	94	94	94	94

This part is the mean in percentage of thing that is possible to found in the different neighborhood. It can be the percentage of restaurant or arcade or bar in the Financial district.

If we take the example of the financial district, we can see that there are 0% of arcade, 5% of American restaurant and 2% of Vietnamese restaurant.

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	A
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
1	Carnegie Hill	0.000000	0.000000	0.00	0.000000	0.011765	0.000000	0.000000	0.000000	0.011765	0.000000	0.011765	0.000000	0.000000	0.
2	Central Harlem	0.000000	0.000000	0.00	0.065217	0.043478	0.000000	0.000000	0.000000	0.000000	0.021739	0.000000	0.000000	0.000000	0.
3	Chelsea	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.110000	0.000000	0.000000	0.000000	0.
4	Chinatown	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.020000	0.
5	Civic Center	0.000000	0.000000	0.00	0.000000	0.052083	0.010417	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.010417	0.
6	Clinton	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.
7	East Harlem	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
8	East Village	0.000000	0.000000	0.00	0.000000	0.010000	0.000000	0.000000	0.010000	0.010000	0.010000	0.000000	0.010000	0.000000	0.
9	Financial District	0.000000	0.000000	0.00	0.000000	0.050000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.
10	Flatiron	0.000000	0.000000	0.00	0.000000	0.020202	0.000000	0.000000	0.000000	0.000000	0.020202	0.000000	0.020202	0.000000	0.
11	Gramercy	0.000000	0.000000	0.00	0.000000	0.035294	0.000000	0.011765	0.000000	0.000000	0.011765	0.000000	0.000000	0.000000	0.
12	Greenwich Village	0.000000	0.000000	0.00	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.
13	Hamilton Heights	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.

Udon Restaurant	Used Bookstore	Vegetarian / Vegan Restaurant	Veterinarian	Video Game Store	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.014925	0.044776	0.000000	0.014925	0.000000
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.023529	0.000000	0.00	0.000000	0.011765	0.035294	0.000000	0.000000	0.035294
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.00	0.000000	0.010000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.010000	0.000000	0.000000	0.010000	0.000000
0.00	0.000000	0.010000	0.000000	0.000000	0.00	0.020000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000
0.00	0.000000	0.010417	0.000000	0.000000	0.00	0.010417	0.000000	0.00	0.010417	0.020833	0.020833	0.000000	0.000000	0.031250
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.020000	0.030000	0.000000	0.000000	0.000000
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.00	0.000000	0.020000	0.000000	0.000000	0.00	0.020000	0.000000	0.00	0.000000	0.030000	0.010000	0.000000	0.000000	0.000000
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.020000	0.000000	0.010000	0.010000
0.00	0.000000	0.020202	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.030303	0.000000	0.000000	0.020202
0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.023529	0.000000	0.000000	0.011765
0.01	0.000000	0.000000	0.000000	0.000000	0.00	0.020000	0.000000	0.00	0.000000	0.010000	0.000000	0.000000	0.000000	0.010000

It could be also interesting to see the top five of most popular “attraction” activities or work that is possible to do in each neighborhood. Here is some example, as said before we can see that in the financial district there are a lot of coffee shop and Pizza place to eat, on the third rank it’s the American restaurant.

```
----Battery Park City----
           venue  freq
0           Park  0.12
1    Coffee Shop  0.07
2          Hotel  0.06
3      Wine Shop  0.04
4  Memorial Site  0.04


----Carnegie Hill----
                venue  freq
0         Coffee Shop  0.09
1                Café  0.05
2         Pizza Place  0.05
3         Yoga Studio  0.04
4  Japanese Restaurant  0.04


----Central Harlem----
                venue  freq
0   African Restaurant  0.07
1   Chinese Restaurant  0.07
2    French Restaurant  0.04
3  American Restaurant  0.04
4                 Bar  0.04
```

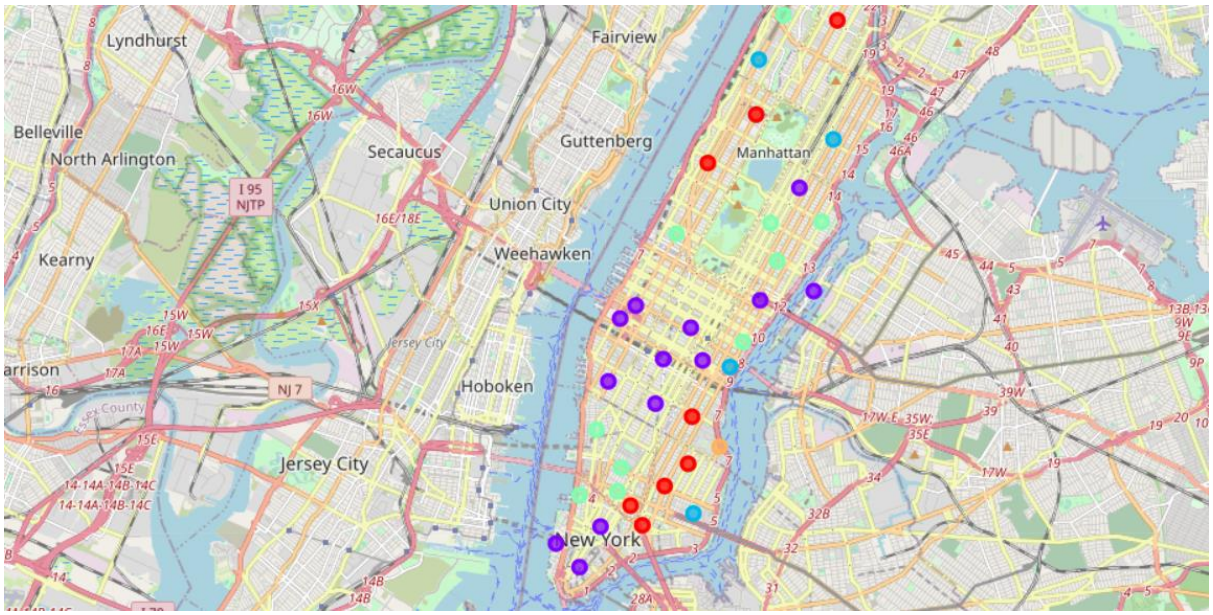
```
----East Harlem----
                      venue  freq
0         Mexican Restaurant  0.12
1                    Bakery  0.10
2            Thai Restaurant  0.07
3  Latin American Restaurant  0.07
4                        Spa  0.05


----East Village----
              venue  freq
0       Cocktail Bar  0.05
1  Mexican Restaurant  0.05
2                 Bar  0.05
3        Pizza Place  0.04
4        Coffee Shop  0.04


----Financial District----
                venue  freq
0         Coffee Shop  0.07
1         Pizza Place  0.05
2  American Restaurant  0.05
3              Hotel  0.05
4   Italian Restaurant  0.03
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Gym	Wine Shop	Shopping Mall	Gourmet Shop	Plaza	Boat or Ferry
1	Carnegie Hill	Coffee Shop	Pizza Place	Café	Yoga Studio	Gym	Wine Shop	Bar	Bookstore	Japanese Restaurant	Grocery Store
2	Central Harlem	African Restaurant	Chinese Restaurant	French Restaurant	American Restaurant	Bar	Cosmetics Shop	Fried Chicken Joint	Seafood Restaurant	Food Truck	Market
3	Chelsea	Art Gallery	Coffee Shop	Café	Bakery	Ice Cream Shop	American Restaurant	Italian Restaurant	Theater	Seafood Restaurant	Hotel
4	Chinatown	Chinese Restaurant	Optical Shop	Bakery	Cocktail Bar	Bubble Tea Shop	Salon / Barbershop	Spa	Ice Cream Shop	American Restaurant	Coffee Shop

Same thing with this, we can see the most famous venue

	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Sandwich Place	Gym	Coffee Shop	Tennis Stadium	Supplement Shop	Steakhouse	Shopping Mall	Seafood Restaurant	Yoga Studio
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Optical Shop	Bakery	Cocktail Bar	Bubble Tea Shop	Salon / Barbershop	Spa	Ice Cream Shop	American Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	2	Café	Bakery	Mobile Phone Shop	Chinese Restaurant	Grocery Store	Donut Shop	Deli / Bodega	New American Restaurant	Mexican Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	2	Mexican Restaurant	Café	Lounge	Pizza Place	Restaurant	Deli / Bodega	American Restaurant	Bakery	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Deli / Bodega	Chinese Restaurant	Sushi Restaurant	Cocktail Bar	Yoga Studio

We can see in this part the different borough of New-York with a specific neighbourhood.



This map show us the different cluster

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Chinatown	Chinese Restaurant	Optical Shop	Bakery	Cocktail Bar	Bubble Tea Shop	Salon / Barbershop	Spa	Ice Cream Shop	American Restaurant	Coffee Shop
6	Central Harlem	African Restaurant	Chinese Restaurant	French Restaurant	American Restaurant	Bar	Cosmetics Shop	Fried Chicken Joint	Seafood Restaurant	Food Truck	Market
12	Upper West Side	Italian Restaurant	Coffee Shop	Wine Bar	Bakery	Bar	Dessert Shop	Indian Restaurant	Mediterranean Restaurant	Mexican Restaurant	Middle Eastern Restaurant
19	East Village	Bar	Mexican Restaurant	Cocktail Bar	Coffee Shop	Pizza Place	Japanese Restaurant	Juice Bar	Ice Cream Shop	Ramen Restaurant	Wine Bar
22	Little Italy	Bubble Tea Shop	Chinese Restaurant	Pizza Place	Mediterranean Restaurant	Spa	Bakery	Hotel	Café	Thai Restaurant	Ice Cream Shop
25	Manhattan Valley	Coffee Shop	Pizza Place	Bar	Mexican Restaurant	Yoga Studio	Grocery Store	Playground	Park	Latin American Restaurant	Korean Restaurant
27	Gramercy	Bagel Shop	Italian Restaurant	Coffee Shop	Pizza Place	Bar	Playground	Mexican Restaurant	Ice Cream Shop	Thai Restaurant	Grocery Store
31	Noho	Coffee Shop	Pizza Place	Italian Restaurant	Japanese Restaurant	Sandwich Place	Grocery Store	Wine Shop	Mexican Restaurant	Sushi Restaurant	Seafood Restaurant

Finally here have an example of a cluster in Manhattan

Discussion

- We see that each neighbourhood is different. We can find various thing to do on each neighbourhood.
- Borough is a bigger than neighbourhood so it give only a global vision of the different cluster and categories.
- We see in Manhattan in Inwood that the Mexican restaurant is the 1 most common venue.

Conclusion

This project is very interesting it help people to have a better understanding of a cluster. It help to understand the neighbourhood with the different venues. This analyse can help a person to find the best place, the most strategic place to have more benefit.

We need also to remember that more data we'll have more the result will be closer to the reality. We see in New-York some neighbourhood don't have for example a playgroud, it could be interesting made it one here compares to other borough where there are already a lot. This project makes us a better understanding of a city and how to find the best place to make a business.