Battle of the Neighbourhoods

Where to stay in New York City

Samuel Cheng

6th June 2020

1. Introduction

1.1. Background

New York City (NYC) is one the most popular city in the world where talents around the world gathers. In the recent times, NYC seems to be experiencing lost in population based on US Census data with total state population as of mid-2019 at 19.45m, a drop of 0.4% from the previous year. Regardless, NYC remains attractive as it remains the city to start your career, be it as an artist or within the finance sector. NYU is also a renowned university that attracts many overseas students annually. Based on data from New York government, New York is home to 3.1m immigrants.

1.2. Problem

With the influx of foreigners, most would need to locate a place to stay. With so many options, it can be challenging identifying which is the best location to stay within NYC considering the amenities nearby.

1.3. Purpose of report

The purpose of this report is to identify neighbourhoods with higher concentration of amenities as a determining factor where would one consider to stay.

1.4. Target Audience

a) Professionals looking to find accommodation after being relocated to NYC
b) Students looking to find accommodation after enrolling in a university in NYC
c) Property agents looking to better understand how they can segment the neighbourhoods in NYC.
d) Investors looking to purchase a property in NYC.

2. Data

2.1. Data sources

a) NYC location data: https://cocl.us/new_york_dataset

b) Surrounding amenities around a neighbourhood: Foursquare API. By using this API, we will get all the venues (amenities) in NYC neighbourhood. We can then determine neighbourhoods with the highest concentration of amenities.

2.2. Data cleaning

New York City comprises 5 Boroughs and 306 Neighbourhoods. This project will focus on all five boroughs and 306 Neighbourhoods.

The initial step is to download all the data from https://cocl.us/new_york_dataset and convert these data into a dataframe. The data will include the Borough, the Neighbourhood, Latitude and Longitude. The formatted output is as follow:

	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

2.3. Feature Selection

In this project, we will search within 500m radius of each neighbourhood to identify the venues. Within each radius, the maximum venues obtainable is 100 based on the subscription status with four square. We will then calculate the top10 most common venue and tabulate it.

3. Methodology

3.1. Explore Dataset

Using folium library to create with markets to establish a good overview of the dataset that we are examining. This will allow a better appreciation of the area that we are covering for this project.

Map of New York City with each neighbourhood as a market

The neighbourhoods are well spaced out with higher concentration within Manhattan. This is within our expectation.

3.2. Examine Foursquare

Using Foursquare's /venues/explore API to get recommended venues for a specific location. To simplify the results set the limit property is 100 and radius is 500. This request returns a JSON data includes up to 100 venues for a coordinate. A dataframe parameter is then created to obtain the details of each recommended venue. The output is as follow:

```
In [144]: results = requests.get(url).json()
          results
```


```
Out[144]: {'meta': {'code': 200, 'requestId': '5edb243f14a126001bafa804'},
           'response': {'headerLocation': 'Wakefield',
            'headerFullLocation': 'Wakefield, Bronx',
            'headerLocationGranularity': 'neighborhood',
            'totalResults': 10,
            'suggestedBounds': {'ne': {'lat': 40.899205181110005,
              'lng': -73.84125857127495},
             'sw': {'lat': 40.89020517211, 'lng': -73.8531424698231}},
            'groups': [{'type': 'Recommended Places',
              'name': 'recommended',
              'items': [{'reasons': {'count': 0,
                 'items': [{'summary': 'This spot is popular',
                   'type': 'general',
                   'reasonName': 'globalInteractionReason'}]},
                'venue': {'id': '4c537892fd2ea593cb077a28',
                 'name': 'Lollipops Gelato',
                 'location': {'address': '4120 Baychester Ave',
                  'crossStreet': 'Edenwald & Bussing Ave',
                  'lat': 40.894123150205274,
                  'lng': -73.84589162362325,
                  'labeledLatLngs': [{'label': 'display',
                    'lat': 40.894123150205274,
```


3.3. Pre-processing data

We first create a dataframe to store the data of the recommended venues:

(9934, 7)

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

As this give us over 9934 results, we would like to group these neighbourhood to have a better understanding on the concentration of amenities around each neighbourhood.

```
nyc_venues_count = nyc_venues.groupby('Neighbourhood', as_index=False).count()
nyc_venues_count
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allerton	29	29	29	29	29	29
1	Annadale	11	11	11	11	11	11
2	Arden Heights	3	3	3	3	3	3
3	Arlington	5	5	5	5	5	5

Then is was ranked in descending order in terms of number of venues in each neighbourhood.

```
nyc_venues_sorted = nyc_venues_count.sort_values(by='Venue', ascending=False)
nyc_venues_sorted
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
185	Murray Hill	124	124	124	124	124	124
49	Chelsea	105	105	105	105	105	105
150	Lenox Hill	100	100	100	100	100	100
154	Little Italy	100	100	100	100	100	100
50	Chinatown	100	100	100	100	100	100
53	Civic Center	100	100	100	100	100	100

The top 50 neighbourhood with most concentrated venues were selected. We will focus on these 50 neighbourhoods and will..

Select the top 50 Neighbourhoods by concentration

```
nyc_venues_top = nyc_venues_sorted.iloc[0:50,0:5]
nyc_venues_top
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude
185	Murray Hill	124	124	124	124
49	Chelsea	105	105	105	105
150	Lenox Hill	100	100	100	100
154	Little Italy	100	100	100	100
50	Chinatown	100	100	100	100

```
m = list(nyc_venues_top['Neighbourhood'])
nyc_venues1 = nyc_venues[nyc_venues['Neighbourhood'].isin(m)]
nyc_venues1
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Ve Ca
60	Kingsbridge	40.881687	-73.902818	Garden Gourmet Market	40.881350	-73.903389	Gc Sh
61	Kingsbridge	40.881687	-73.902818	Leche y Miel	40.883742	-73.901857	La Ar Re
62	Kingsbridge	40.881687	-73.902818	Kingsbridge Social Club	40.884545	-73.901964	Pi:

A one-hot encoding to the venue categories is then created and a mean of each category determined:

```
nyc_grouped = nyc_onehot.groupby('Neighbourhood').mean().reset_index()
nyc_grouped.head()
```

	Neighbourhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arca
0	Astoria	0.0	0.0	0.0	0.0	0.010000	0.0	0.0	0.0
1	Battery Park City	0.0	0.0	0.0	0.0	0.015385	0.0	0.0	0.0
2	Bay Ridge	0.0	0.0	0.0	0.0	0.033708	0.0	0.0	0.0
3	Bayside	0.0	0.0	0.0	0.0	0.039474	0.0	0.0	0.0
4	Belmont	0.0	0.0	0.0	0.0	0.010204	0.0	0.0	0.0

Then we calculate the frequency of categories for each neighbourhood and get the top10 most common venues for each of the top 50 neighbourhoods we have identified.

	Neighbourhood	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
0	Astoria	Middle Eastern Restaurant	Bar	Mediterranean Restaurant	Seafood Restaurant	Greek Restaurant	Hookah Bar	Pizza Place
1	Battery Park City	Park	Hotel	Memorial Site	Gym	Coffee Shop	Playground	Beer Garden
2	Bay Ridge	Spa	Italian Restaurant	Bar	Pizza Place	Greek Restaurant	Thai Restaurant	American Restaurant
3	Bayside	Bar	Pub	Indian Restaurant	Mexican Restaurant	Sushi Restaurant	Greek Restaurant	Pizza Place
4	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Dessert Shop	Bank	Donut Shop

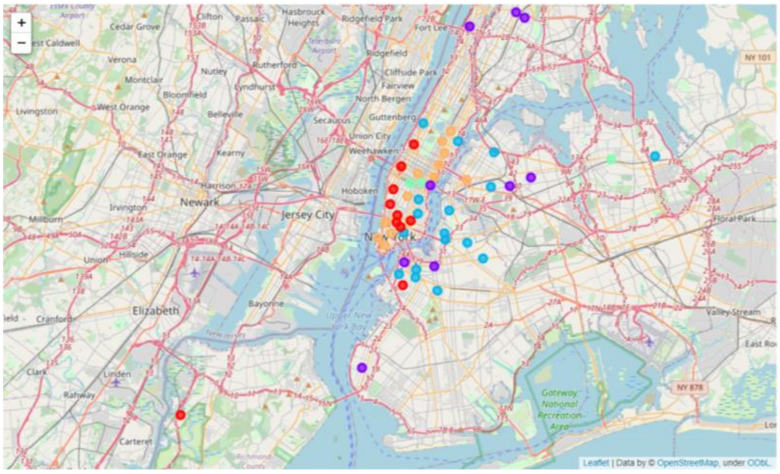
3.4. Clustering

For this project, k-means is an appropriate clustering algorithm. Because we have an unlabelled dataset, so this is an unsupervised learning project. K-means clustering will partition n, the number of observation, into k, the number of clusters, in which each observation belongs to the cluster with the nearest mean. By clustering the neighbourhood, we would be able to identify the characteristics of each neighbourhood.

4. Results

4.1. Visualise clusters

Folium library is used to create a map with clustered markets:



The whole of NYC top 50 neighbourhoods have been split into 5 different clusters.

4.2. Print Clusters

Cluster 1

```
nyc_cluster1
```

	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 Mos Commc Venue
66	Carroll Gardens	Italian Restaurant	Coffee Shop	Pizza Place	Cocktail Bar	Bakery	Bar	Spa
112	Lincoln Square	Italian Restaurant	Plaza	Café	Gym / Fitness Center	Theater	Concert Hall	Perform Arts Ver
113	Clinton	Theater	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Wine Shop	Gym	Pizza Place
116	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	Bakery	American Restaurant	Café	Park
117	Greenwich Village	Italian Restaurant	Sushi Restaurant	Café	Chinese Restaurant	Gym	Indie Movie Theater	Gourme Shop
121	Little Italy	Mediterranean Restaurant	Italian Restaurant	Bubble Tea Shop	Pizza Place	Ice Cream Shop	Bakery	Spa
122	Soho	Italian Restaurant	Sandwich Place	Mediterranean Restaurant	Ice Cream Shop	Bakery	Coffee Shop	Seafooc Restaur
123	West Village	Italian Restaurant	Wine Bar	New American Restaurant	Bakery	Cocktail Bar	Park	America Restaur
244	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	Bakery	American Restaurant	Café	Park
248	Noho	Italian Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Grocery Store	Wine Shop	Hotel

Cluster 2

	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
5	Kingsbridge	Pizza Place	Sandwich Place	Deli / Bodega	Bar	Supermarket	Mexican Restaurant	Latin American Restaurant
16	Fordham	Mobile Phone Shop	Shoe Store	Pizza Place	Bank	Donut Shop	Pharmacy	Gym / Fitness Center
34	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Dessert Shop	Bank	Donut Shop
46	Bay Ridge	Spa	Italian Restaurant	Bar	Pizza Place	Greek Restaurant	Thai Restaurant	American Restaurant
64	Brooklyn Heights	Deli / Bodega	Yoga Studio	Pizza Place	Park	Gym	Bakery	Italian Restaurant
84	Clinton Hill	Italian Restaurant	Pizza Place	Mexican Restaurant	Wine Shop	Thai Restaurant	Yoga Studio	Japanese Restaurant
101	Washington Heights	Café	Bakery	Mobile Phone Shop	Latin American Restaurant	Chinese Restaurant	Donut Shop	Pizza Place
130	Woodside	Grocery Store	Filipino Restaurant	Thai Restaurant	Bakery	Latin American Restaurant	Bar	Donut Shop
131	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mexican Restaurant	Thai Restaurant	Mobile Phone Shop
274	Tudor City	Park	Café	Mexican Restaurant	Pizza Place	Deli / Bodega	Sushi Restaurant	Vietnamese Restaurant

Cluster 3

	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 Con Ven
49	Greenpoint	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Yoga Studio	Café	French Restaurant	Sushi Restaurant	New Ame Res
59	Prospect Heights	Bar	Mexican Restaurant	Thai Restaurant	Wine Shop	Gourmet Shop	Cocktail Bar	Café	Yoga Studio	Ice Sho
62	Bushwick	Bar	Coffee Shop	Mexican Restaurant	Pizza Place	Deli / Bodega	Bakery	Thrift / Vintage Store	Discount Store	Caf
65	Cobble Hill	Coffee Shop	Playground	Pizza Place	Bar	Bakery	Italian Restaurant	Deli / Bodega	Cocktail Bar	Yog Stu
86	Downtown	Pizza Place	Burger Joint	Coffee Shop	Sandwich Place	Chinese Restaurant	Dance Studio	Performing Arts Venue	Bar	Deli Bod
87	Boerum Hill	Coffee Shop	Dance Studio	Bar	Sandwich Place	Bakery	French Restaurant	Furniture / Home Store	Arts & Crafts Store	Yog Stu
95	East Williamsburg	Bar	Deli / Bodega	Bakery	Coffee Shop	Cocktail Bar	Concert Hall	Mexican Restaurant	Music Venue	Veg / Ve Res
96	North Side	Coffee Shop	Pizza Place	American Restaurant	Bar	Yoga Studio	Wine Bar	Vegetarian / Vegan Restaurant	Jewelry Store	Bak
97	South Side	Coffee Shop	Bar	American Restaurant	Pizza Place	Yoga Studio	Wine Bar	Breakfast Spot	Art Gallery	Mex Res
100	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Spa	American Restaurant	Ice Cream Shop	Bar	Vietnamese Restaurant	Bub Tea
108	Yorkville	Coffee Shop	Italian Restaurant	Gym	Bar	Deli / Bodega	Sushi Restaurant	Japanese Restaurant	Diner	Mex Res
111	Upper West Side	Bar	Italian Restaurant	Dessert Shop	Indian Restaurant	Wine Bar	Sports Bar	Coffee Shop	Ice Cream Shop	Tha Res
118	East Village	Mexican Restaurant	Bar	Cocktail Bar	Coffee Shop	Japanese Restaurant	Wine Bar	Speakeasy	Juice Bar	Pizz Plac
126	Gramercy	Bagel Shop	Coffee Shop	Bar	Pizza Place	Mexican Restaurant	Cocktail Bar	Grocery Store	American Restaurant	Itali Res
129	Astoria	Middle Eastern Restaurant	Bar	Mediterranean Restaurant	Seafood Restaurant	Greek Restaurant	Hookah Bar	Pizza Place	Café	Deli Bod
151	Bayside	Bar	Pub	Indian Restaurant	Mexican Restaurant	Sushi Restaurant	Greek Restaurant	Pizza Place	American Restaurant	Don Sho
277	Sunnyside Gardens	Bar	Grocery Store	Pizza Place	Turkish Restaurant	Coffee Shop	American Restaurant	Pharmacy	Korean Restaurant	Ban

Cluster 4

	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	10t Mo Co Ve
115	Murray Hill	Korean Restaurant	Coffee Shop	Hotel	Sandwich Place	Pizza Place	Japanese Restaurant	Bar	Gym / Fitness Center	Burger Joint	Bal
180	Murray Hill	Korean Restaurant	Coffee Shop	Hotel	Sandwich Place	Pizza Place	Japanese Restaurant	Bar	Gym / Fitness Center	Burger Joint	Bal
250	Midtown South	Korean Restaurant	Hotel	Japanese Restaurant	American Restaurant	Dessert Shop	Café	Coffee Shop	Burger Joint	Hotel Bar	Gy Fit Ce

Cluster 5

	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 Mo Comm Venue
107	Upper East Side	Italian Restaurant	Coffee Shop	Bakery	Gym / Fitness Center	Spa	Exhibit	Juice Bar	Yoga Studio	French Restau
109	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Cocktail Bar	Sushi Restaurant	Café	Burger Joint	Gym	Gym / Center
114	Midtown	Coffee Shop	Hotel	Bakery	Theater	Clothing Store	Sushi Restaurant	Japanese Restaurant	Pizza Place	Cuban Restau
120	Tribeca	Park	Italian Restaurant	American Restaurant	Café	Wine Bar	Spa	Coffee Shop	Men's Store	Greek Restau
127	Battery Park City	Park	Hotel	Memorial Site	Gym	Coffee Shop	Playground	Beer Garden	Shopping Mall	Plaza
128	Financial District	Coffee Shop	Pizza Place	American Restaurant	Cocktail Bar	Gym	Mexican Restaurant	Café	Falafel Restaurant	Event
139	Long Island City	Coffee Shop	Hotel	Bar	Pizza Place	Mexican Restaurant	Café	Supermarket	Gym / Fitness Center	Deli / E
247	Carnegie Hill	Coffee Shop	Pizza Place	Café	Bookstore	Gym / Fitness Center	Gym	Japanese Restaurant	Yoga Studio	Wine S
249	Civic Center	Coffee Shop	Hotel	Cocktail Bar	American Restaurant	French Restaurant	Spa	Park	Gym / Fitness Center	Yoga S
271	Sutton Place	Gym / Fitness Center	Coffee Shop	Park	Italian Restaurant	Bagel Shop	Gym	Hotel	Furniture / Home Store	Health Beauty Servic
273	Turtle Bay	Italian Restaurant	Café	Coffee Shop	Sushi Restaurant	Park	Deli / Bodega	Wine Bar	Seafood Restaurant	French Restau
276	Flatiron	Gym / Fitness Center	Café	Italian Restaurant	Yoga Studio	Park	Japanese Restaurant	Spa	Coffee Shop	Medite Restau

4.3. Answer the question

By examining the cluster, the below are how I would describe the clusters:

Cluster	Characteristics	Potential profile of individuals who would like to live in this cluster
1	Higher concentration of Artistic related venues and cafes.	• University students • Artists
2	Well distributed concentration of food,	• Families

	essential services and lifestyle shops.	• Middle aged - older folks
3	High concentration of bars and restaurants	• University students • Young adults who enjoy drinking
4	High concentration of Asian restaurants and near Korean town	• Koreans (this is a little of an anomaly)
5	High centration of Hotels, restaurants, and gyms	• Professionals who prioritise convenience by staying near office

5. Discussion

For this project, k-means work as anticipated. We have used amenities as the key independent variable on whether a particular location would appeal to the individual.

Although this does give us insight to each location, to finally decide on which neighbourhood to pick, rental price of the neighbourhood does play an important role. Also, the crime rate in that neighbourhood would also play a critical role.

6. Conclusion

In this project, we have used Foursquare to solve the problem "where to stay in NYC".

The neighbourhood data has been collected from government data, formatted, and processed. Details of venues surround the neighbourhood was extracted from Foursquare's API. We further narrowed our search by picking the top 50 neighbourhood with the highest venue concentration. One-hot encode the venues' categories and calculated the frequencies to obtain the Top10 common venues for each neighbourhood.

K-means algorithm was used to cluster the dataset into 5 clusters. Folium was used to visualise the result and through examining the result of each cluster, we have managed to describe the characteristic of each clusters.

Data on rental rate and crime rate can help to improve the model to obtain a more convincing proposal to potential tenants.

7. References

https://developer.foursquare.com/docs
https://cocl.us/about-polr