

# Analysing the best location for a new restaurant in New-York City

Ahmed Mohy

IBM Data Science Professional Certificate  
Capstone

# **Introduction**

The restaurant industry is difficult to succeed in. Aside from the operational challenges, there are two external factors that are required to survive:

- 1) Regular foot traffic
- 2) An underserved population that they can address.

Foot traffic may come from locals and tourists visiting other nearby venues, from the people that live in that neighbourhood, or in some cases: a busy work centre/industrial area.

A well-organized restaurateur will know their particular cuisine style, price range, and desired target customers. A target customer in this case can be defined as underserved if there's a lack of other options that serve this restaurant's style, price, and theme (i.e., an Italian restaurant in little Italy will have a hard time differentiating itself)

## **Business Explanation**

The restaurant in mind for this study is a new Hong Kong style cafe, modeled after one of my favourite restaurants in Boston: [Double Chin](#)

This restaurant caters to a younger crowd, mostly in their 20's and 30's. Their menu serves fast, cheap, and delicious eats with an emphasis on regularly creating fun dishes.

Their late-night menu begs for them to be the last stop after a long night on the town. The target customer is groups looking for something to eat in between visiting art shows, concerts, bars, and other big social events.

## **Business Problem**

The issue here is that we don't know where the restaurant should be located. The owners had success with their first restaurant in Chinatown, but the crowding of other places serving similar fare made it extremely difficult to stand out.

Aside from Chinatown, the owners don't know where else to look. They need a neighbourhood with a steady stream of customers visiting other nightlife options, with hopefully not too many other Chinese/Asian fusion restaurants in the area.

Can we use publicly available data to source potential new locations for their restaurant?

# Data

To attempt a solution at this problem, we'll be using the [Foursquare API](#) to fetch venue information and analyse different neighbourhoods.

To fulfil the requirements of the business problem stated, we'll specifically be looking for two types of information:

1- Data on the neighbourhoods in general, specifically:

- The categories for each venue in each Manhattan neighbourhood
- Popular days and hours for venues that serve our target customer (nightlife venues)
- A popularity score for each of the target venues, calculated by:
  - The total number of tips
  - The like/dislike volume and ratio

2- Data on venues that might compete with ours, specifically:

- The relative frequency of all late night restaurants
- The price range of late night restaurants in each neighbourhood
- The cuisine type of late night restaurants in each neighbourhood

### Data source:

1- New-York data set:

[https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

2- Foursquare API:

<https://developer.foursquare.com/>

## Data (fetching popular venues in a neighbourhood)

```
: uniques = venues['Category'].value_counts()  
uniques
```

```
: Chinese          9  
Bubble Tea        6  
Cocktail           4  
American           4  
Dim Sum            4  
Vietnamese         4  
Bakery             3  
Noodles            3  
Hotpot             3  
Salon / Barbershop 3  
Optical            2  
Bar                2  
Spa                2  
Coffee Shop        2  
Malay              2  
Asian              2  
Dumplings          2  
Ice Cream          2  
Sandwiches         2  
Boutique           1  
Cosmetics          1
```

## Data (Fetching the popular hours for a venue)

```
: VENUE_ID = venues.loc[0, 'venue.id']
url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
    VENUE_ID,
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
)

details = requests.get(url).json()
popularTimes = details['response']['venue']['popular']
popularTimes
```

```
: {'status': 'Likely open',
  'richStatus': {'entities': [], 'text': 'Likely open'},
  'isOpen': True,
  'isLocalHoliday': False,
  'timeframes': [{'days': 'Today',
    'includesToday': True,
    'open': [{'renderedTime': '8:00 AM-3:00 PM'},
      {'renderedTime': '6:00 PM-8:00 PM'}],
    'segments': []},
    {'days': 'Mon',
      'open': [{'renderedTime': '7:00 AM-8:00 AM'},
        {'renderedTime': '11:00 AM-Noon'},
        {'renderedTime': '3:00 PM-9:00 PM'}],
      'segments': []},
    {'days': 'Tue',
      'open': [{'renderedTime': '7:00 AM-9:00 AM'},
        {'renderedTime': '11:00 AM-Noon'},
        {'renderedTime': '5:00 PM-8:00 PM'}],
      'segments': []},
```



# Data (Fetching venue attributes)

```
VENUE_ID = '4db3374590a0843f295fb69b'
url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
    VENUE_ID,
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
)
request2 = requests.get(url).json()
attributes = request2['response']['venue']['attributes']
attributes
```

```
{'groups': [{'type': 'price',
  'name': 'Price',
  'summary': '$',
  'count': 1,
  'items': [{'displayName': 'Price', 'displayValue': '$', 'priceTier': 1}]}],
{'type': 'reservations',
  'name': 'Reservations',
  'count': 3,
  'items': [{'displayName': 'Reservations', 'displayValue': 'No'}]},
{'type': 'payments',
  'name': 'Credit Cards',
  'summary': 'No Credit Cards',
  'count': 7,
  'items': [{'displayName': 'Credit Cards', 'displayValue': 'No'}]},
{'type': 'outdoorSeating',
  'name': 'Outdoor Seating',
  'count': 1,
  'items': [{'displayName': 'Outdoor Seating', 'displayValue': 'No'}]},
{'type': 'serves',
  'name': 'Menus',
  'summary': 'Dinner, Lunch & more',
  'count': 8,
```

## Data (fetching number of tips)

```
|: VENUE_ID = '4db3374590a0843f295fb69b'
url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
    VENUE_ID,
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
)
request2 = requests.get(url).json()
attributes = request2['response']['venue']['tips']
attributes
```

```
|: {'count': 172,
  'groups': [{'type': 'others',
    'name': 'All tips',
    'count': 172,
    'items': [{'id': '4df167a3b0fb807158b979f8',
      'createdAt': 1307666339,
      'text': 'Big tray chicken. Make sure you ask them to add hand-pulled noodles to it.',
      'type': 'user',
      'canonicalUrl': 'https://foursquare.com/item/4df167a3b0fb807158b979f8',
      'lang': 'en',
```

---

# Methodology:

After fetching data for all venues and neighbourhoods in Brooklyn, I filtered out the specific venues I was interested in (Restaurants and Nightlife venues). I then picked venues from this list that had a high number of likes and tips (showing high popularity). This list was used to narrow down the neighbourhood list to just 19 entries

I took the filtered list, and searched for venues again, this time with a little wider search radius and a higher limit on the # of venues

I took the information from this list of “hot” neighbourhoods and pulled together a relative frequency of nightlife venues, restaurants, and then Chinese Restaurants

Out[104]:

	level_0	Borough	Neighborhood	Latitude	Longitude	Night Counts	Restaurant Counts	Total Counts	Night Frequency	Eats Frequency	Chinese Restaurants	Chinese Frequency
0	0	Brooklyn	Carroll Gardens	40.680540	-73.994654	71	27	100	0.710000	0.270000	1.0	0.010000
1	1	Brooklyn	Cobble Hill	40.687920	-73.998561	65	31	100	0.650000	0.310000	3.0	0.030000
2	2	Brooklyn	Crown Heights	40.670829	-73.943291	15	4	19	0.789474	0.210526	0.0	0.000000
3	3	Brooklyn	Gravesend	40.595260	-73.973471	23	7	30	0.766667	0.233333	2.0	0.066667
4	4	Brooklyn	Greenpoint	40.730201	-73.954241	68	32	100	0.680000	0.320000	1.0	0.010000
5	5	Brooklyn	Prospect Heights	40.676822	-73.964859	53	30	84	0.630952	0.357143	1.0	0.011905
6	6	Brooklyn	Red Hook	40.676253	-74.012759	39	11	50	0.780000	0.220000	0.0	0.000000
7	7	Brooklyn	Windsor Terrace	40.656946	-73.980073	22	6	28	0.785714	0.214286	1.0	0.035714

After pulling all the data together in a single data frame, I ran a clustering analysis to group similar neighborhoods together

The different clusters were analysed to see which were most favourable for our new restaurant (a lot of nightlife, very few Chinese restaurants)

Out[116]:

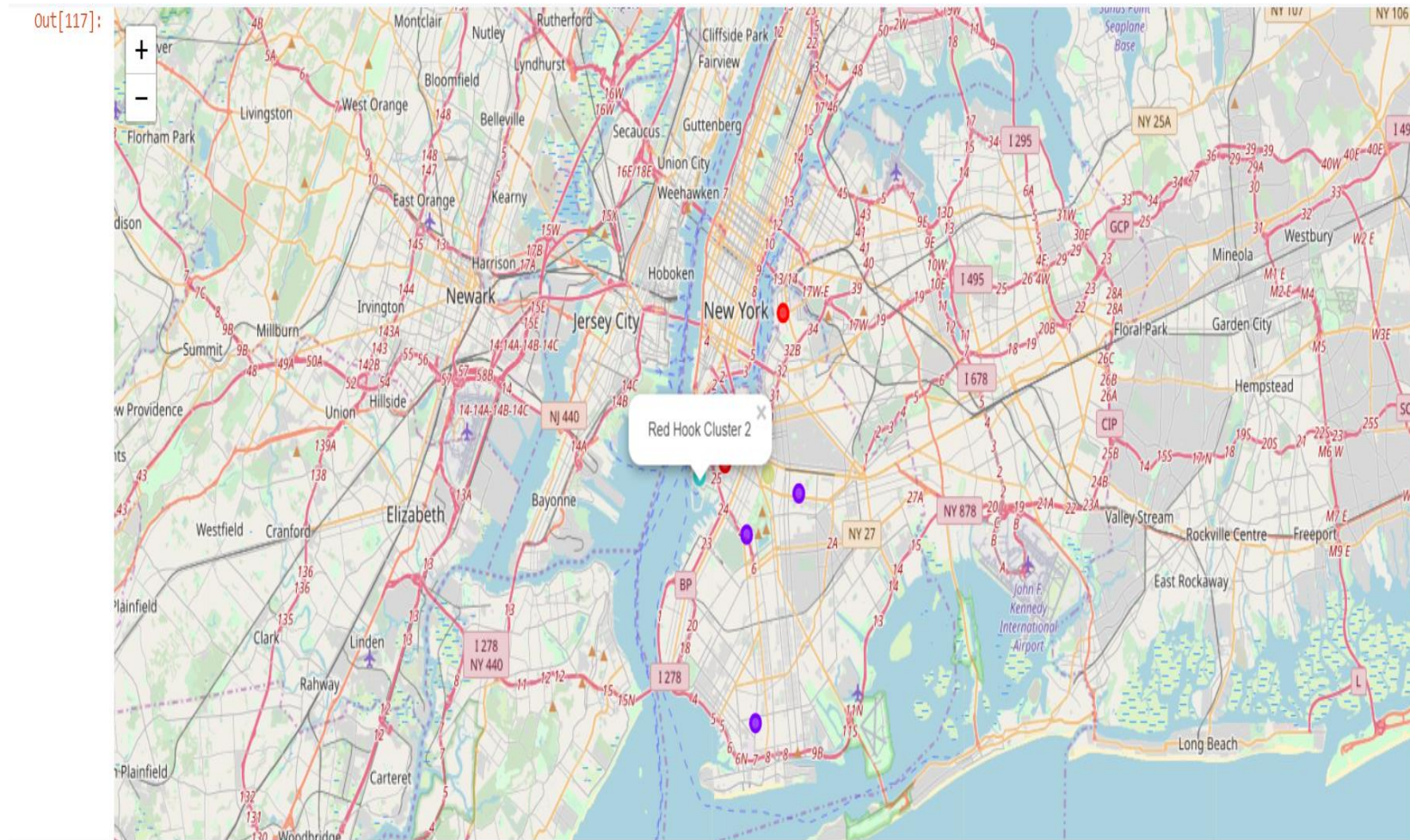
	level_0	Borough	Neighborhood	Latitude	Longitude	Night Counts	Restaurant Counts	Total Counts	Night Frequency	Eats Frequency	Chinese Restaurants	Chinese Frequency	Cluster
0	0	Brooklyn	Carroll Gardens	40.680540	-73.994654	71	27	100	0.710000	0.270000	1.0	0.010000	0
1	1	Brooklyn	Cobble Hill	40.687920	-73.998561	65	31	100	0.650000	0.310000	3.0	0.030000	0
2	2	Brooklyn	Crown Heights	40.670829	-73.943291	15	4	19	0.789474	0.210526	NaN	NaN	1
3	3	Brooklyn	Gravesend	40.595260	-73.973471	23	7	30	0.766667	0.233333	2.0	0.066667	1
4	4	Brooklyn	Greenpoint	40.730201	-73.954241	68	32	100	0.680000	0.320000	1.0	0.010000	0
5	5	Brooklyn	Prospect Heights	40.676822	-73.964859	53	30	84	0.630952	0.357143	1.0	0.011905	3
6	6	Brooklyn	Red Hook	40.676253	-74.012759	39	11	50	0.780000	0.220000	NaN	NaN	2
7	7	Brooklyn	Windsor Terrace	40.656946	-73.980073	22	6	28	0.785714	0.214286	1.0	0.035714	1





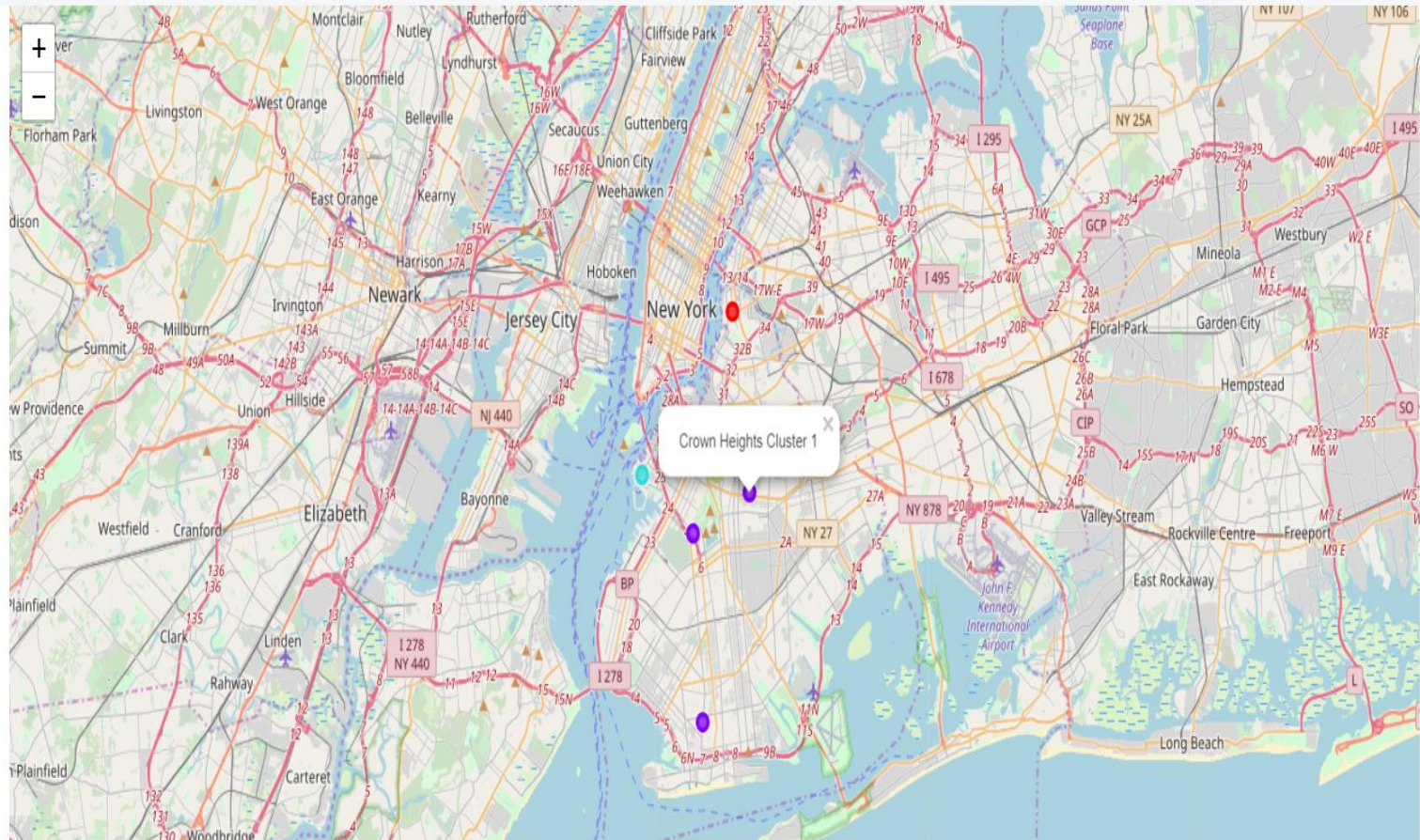


This is based on the high number of nightlife venues and the almost complete absence of other Chinese Restaurants.





Out[117]:



The ideal neighbourhoods for our new location are:

- Red Hook
- Crown Heights
- Prospect Heights
- Windsor Terrace

## **Discussion**

Basing our popularity decision based on number of likes and tips is biased, since it doesn't take into account the length of time a place has been open in. When you use the explore endpoint in the Foursquare API it returns search results based on your own user preferences. Obviously this doesn't give perfectly fair data to inform our model.



## **Conclusion:**

After picking the best neighbourhood, there's a lot more work that needs to be done to actually find the best location. However, this project gives a nice start to the process and narrowed down a very long list (from 70 to 4 choices) which are:

- Red Hook
- Crown Heights
- Prospect Heights
- Windsor Terrace