

Brooklyn_vs_Queens_Smiliar_Real_Estate

Table of Contents

1	<i>Introduction</i>	<i>1</i>
1.1	<i>Target Audience.....</i>	<i>1</i>
2	<i>Data Collection.....</i>	<i>1</i>
2.1	<i>Data Source</i>	<i>1</i>
2.2	<i>Data Format</i>	<i>2</i>
3	<i>Exploring the Data</i>	<i>4</i>
4	<i>Comparing The Data With Venues obtained from four Square API & Conclusion</i>	<i>6</i>
5	<i>Reference.....</i>	<i>7</i>

1 Introduction

New York City (NYC), often called the City of New York or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over about 302.6 square miles (784 km²), New York is also the most densely populated major city in the United States.[10] Located at the southern tip of the U.S. state of New York, the city is the centre of the New York metropolitan area, the largest metropolitan area in the world by urban landmass.[1]

Person X Hailing from Brooklyn Heights, NY, is eyeing to buy a Real Estate property in Queens. Irrespective of prices he wants to get a Property in the neighbourhoods which has same venues and features as of Brooklyn Heights. He wants to understand if he would find neighbourhoods similar to Brooklyn Heights in Queens and if so which Places would be best for him to invest in. He seeks help of a Data Analyst to find out the answers for his questions.

1.1 Target Audience

The report will help in Real Estate brokers from Brooklyn Heights, find appropriate properties for their clients who would like to buy properties with a similar social construct of Brooklyn Heights.

2 Data Collection

2.1 Data Source

The NY Coordinates is available from https://cocl.us/new_york_dataset[2]

2.2 Data Format

The data set contains information in the below format :

```
{'type': 'Feature',  
  'id': 'nyu_2451_34572.1',  
  'geometry': {'type': 'Point',  
    'coordinates': [-73.84720052054902, 40.89470517661]},  
  'geometry_name': 'geom',  
  'properties': {'name': 'Wakefield',  
    'stacked': 1,  
    'annoline1': 'Wakefield',  
    'annoline2': None,  
    'annoline3': None,  
    'annoangle': 0.0,  
    'borough': 'Bronx',  
    'bbox': [-73.84720052054902,  
      40.89470517661,  
      -73.84720052054902,  
      40.89470517661]}}
```

A Properties tag contains the name of the neighbourhood, the Borough it belongs to, and the coordinates of the same. In total there are 5 Boroughs comprising of 306 Neighbourhoods.

```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(  
    len(neighborhoods['Borough'].unique()),  
    neighborhoods.shape[0]  
))
```

The dataframe has 5 boroughs and 306 neighborhoods.

Figure 1. Overview of Data

```
Borough  
Bronx          52  
Brooklyn       70  
Manhattan      40  
Queens         81  
Staten Island  63  
dtype: int64
```

Figure 2. No of Neighbourhoods in each Borough

The Json obtained in the above format is process to obtain the Borough, Neighbourhood, Latitude and Longitude in a data frame as shown below.

```
In [11]: neighborhoods.head()  
Out[11]:
```

	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

Figure 3. Processed NY Data with Boroughs, Neighbourhoods and Coordinates

For our analysis we are interested with only two Boroughs Brooklyn and Queens, with 70 and 81 Neighbourhoods respectively.

```
In [17]: brooklyn_data = neighborhoods[neighborhoods['Borough'] == 'Brooklyn'].reset_index(drop=True)
brooklyn_data
```

Out[17]:

	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
5	Brooklyn	Brighton Beach	40.576825	-73.965094
6	Brooklyn	Sheepshead Bay	40.586890	-73.943186
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
8	Brooklyn	Flatbush	40.636326	-73.958401
9	Brooklyn	Crown Heights	40.670829	-73.943291
10	Brooklyn	East Flatbush	40.641718	-73.936103

Figure 4. Brooklyn Data

```
queens_data = neighborhoods[neighborhoods['Borough'] == 'Queens'].reset_index(drop=True)
queens_data
```

Out[22]:

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138
5	Queens	Corona	40.742382	-73.856825
6	Queens	Forest Hills	40.725264	-73.844475
7	Queens	Kew Gardens	40.705179	-73.829819
8	Queens	Richmond Hill	40.697947	-73.831833
9	Queens	Flushing	40.764454	-73.831773
10	Queens	Long Island City	40.750217	-73.939202

Figure 5. Queens Data

Further, to Compare the Brooklyn Heights Neighbourhood with that of Queens, we isolate and have Brooklyn Heights in a variable.

```
In [18]: brooklyn_my_place = brooklyn_data[brooklyn_data['Neighborhood'] == 'Brooklyn Heights'].reset_index(drop=True)
```

```
In [19]: brooklyn_my_place.shape
```

Out[19]: (1, 4)

3 Exploring the Data

In this section we visualize the Neighbourhoods of Brooklyn and Queens. The maps are created with the respective coordinates obtained from the Json from Section 2.

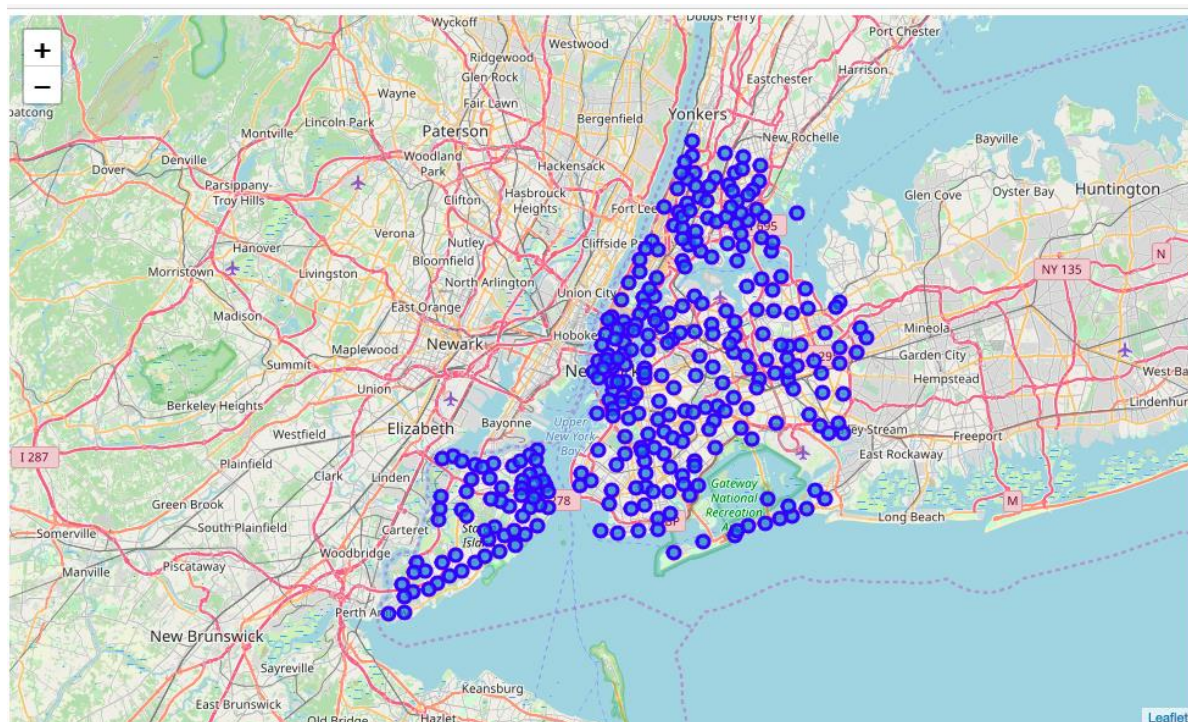


Figure 6. New York City Map plot with 5 Boroughs and 306 Neighbourhoods

A Visual Representation of the Brooklyn with Brooklyn Heights Highlighted

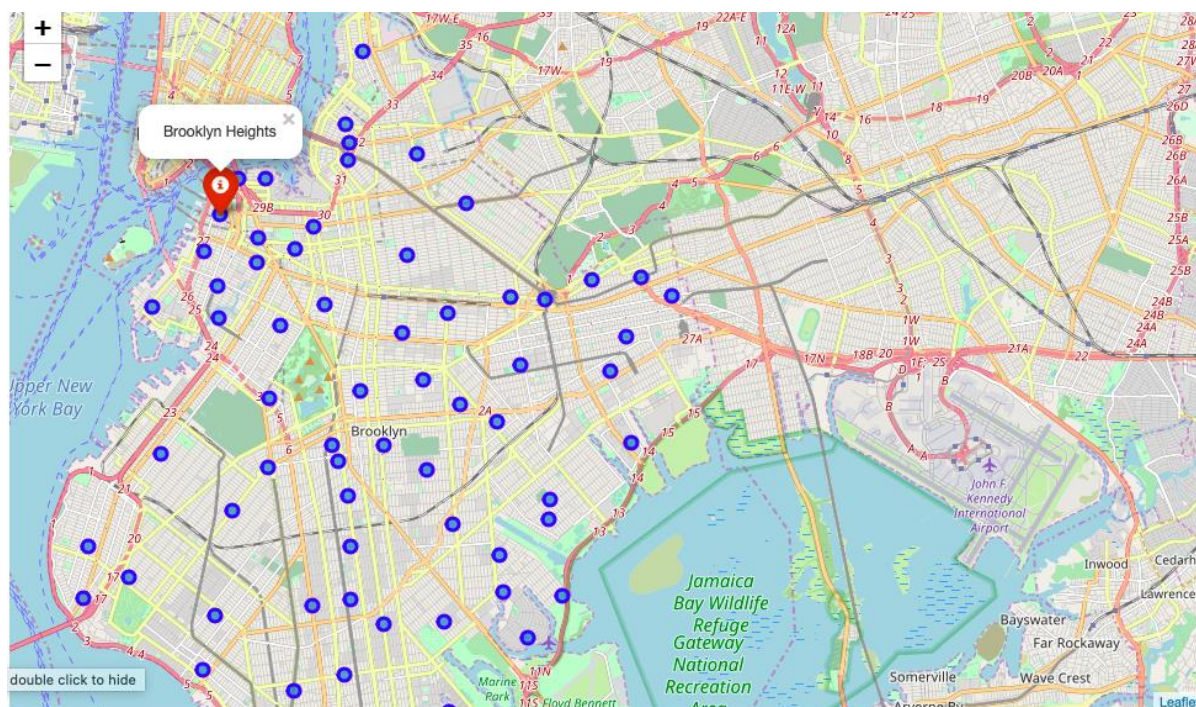


Figure 6. Visual representation of Brooklyn Heights.

A Visual Representation of the Queens Highlighted

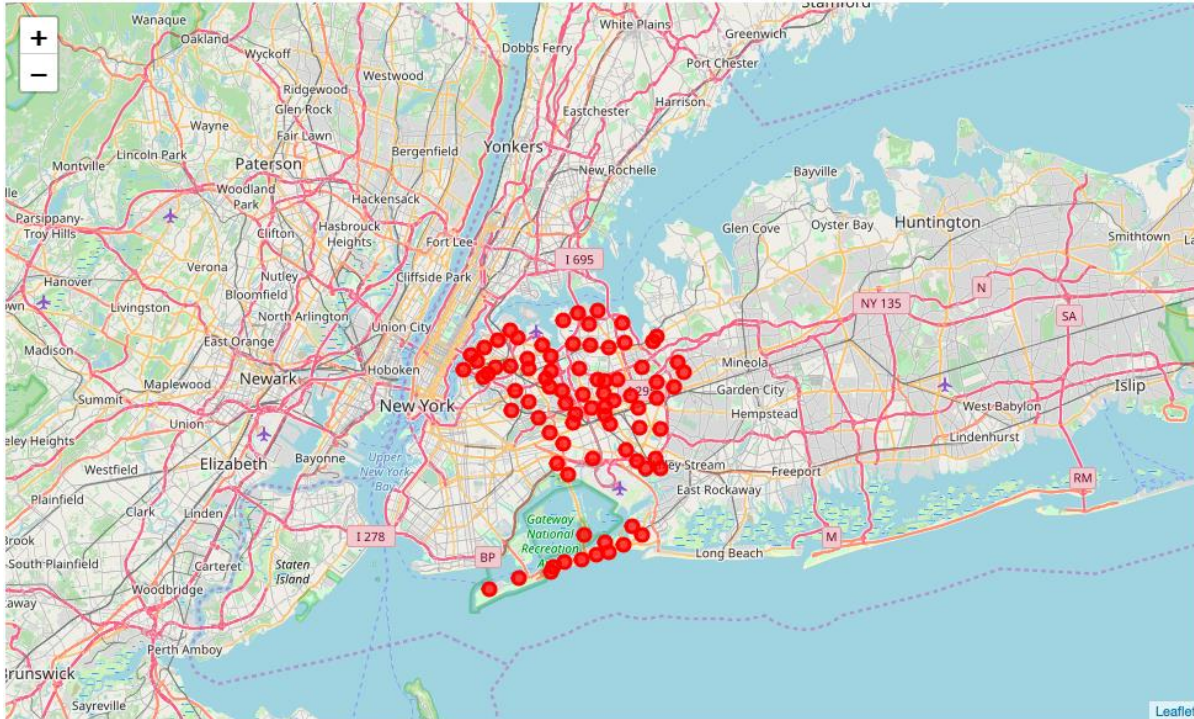


Figure 7. Visual representation of Queens

A Visual Representation of the Brooklyn and Queens Highlighted

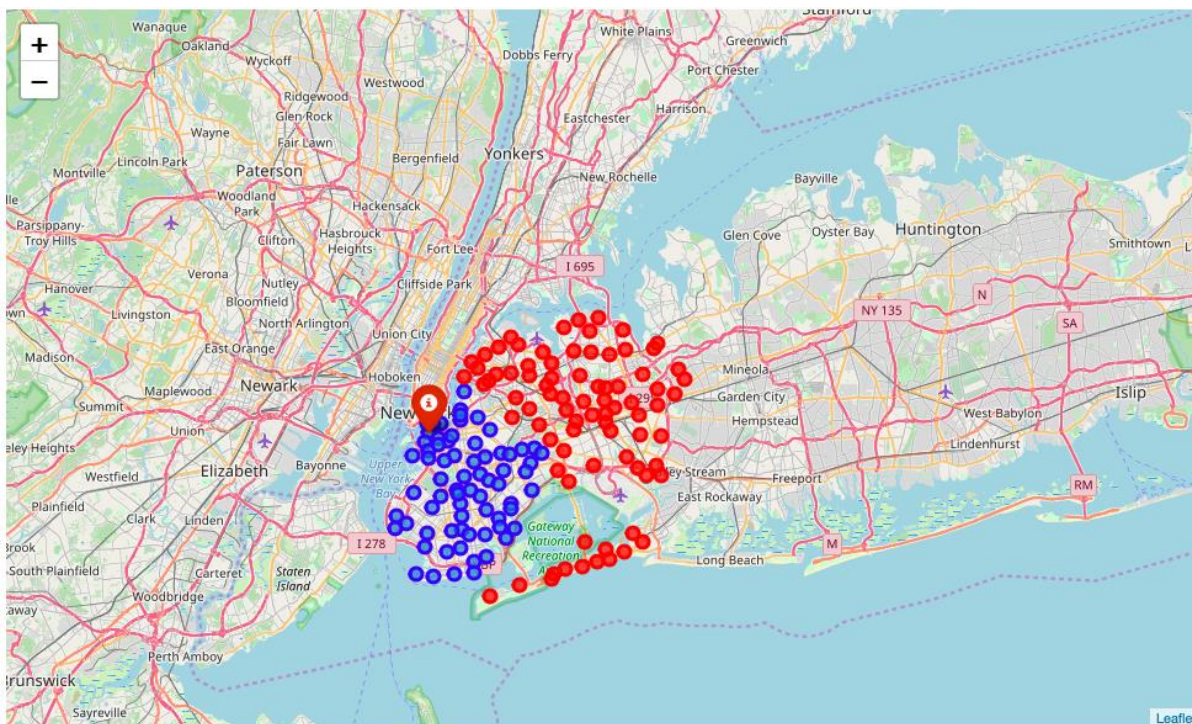


Figure 8. Visual representation of Brooklyn and Queens

Further, to analyse the two neighbourhoods we use Four Square API to obtain the various types of venues in each neighbourhoods. For instance, below is a sample the venue name, from the Four Square API^[3]

```
'venue': {
'id': '5012043564a4944f5c47738a',
'name': "Dellarocco's",
'location': {
'address': '214 Hicks St',
'crossStreet': 'btwn Montague & Remsen St.',
'lat': 40.69499209990746,
'lng': -73.99592429396137,
'labeledLatLngs': [{
'label': 'display',
'lat': 40.69499209990746,
'lng': -73.99592429396137}]}
```

Along with the venue coordinates the API also sends across the ratings for each venue, its category, reviews etc.

`nearby_venues.head()`

Out [32]:

	name	categories	lat	lng
0	Brooklyn Historical Society	History Museum	40.694942	-73.992333
1	Dellarocco's	Pizza Place	40.694992	-73.995924
2	Sushi Gallery	Sushi Restaurant	40.697595	-73.993236
3	Emack & Bolio's	Ice Cream Shop	40.694987	-73.994946
4	Area Yoga & Spa	Yoga Studio	40.694568	-73.993514

Figure 9. Visual representation of Brooklyn and Queens

4 Comparing The Data With Venues obtained from four Square API & Conclusion

Venue details for Queens is obtained as described above and the same is combined with the Venues of Brooklyn Heights. This is ensure the complete data set is in one data frame for us to cluster and identify the Neighbourhoods of Queens having similar venues as of Brooklyn Heights.

Brooklyn Heights	100	100	100	100	100	100	100
------------------	-----	-----	-----	-----	-----	-----	-----

```
In [50]: # Merging the Data To Compare Queens and Brooklyn Venues
merged_neighborhoods=pd.concat([queens_venues, brooklyn_venues], ignore_index=True)

In [51]: merged_neighborhoods.groupby('Neighborhood').count()
```

Figure 10. Merging of Brooklyn Heights Venues with Queens venues to compare

In the process we obtain the Top five venues in each neighbourhoods in the below format

---Brooklyn Heights---

	venue	freq
0	Yoga Studio	0.04
1	Deli / Bodega	0.04
2	Park	0.04
3	Pizza Place	0.03
4	Pharmacy	0.03

Upon analysis, below is the visual Representation of Neighbourhoods that are similar to Brooklyn Heights in Queens. Person X, Can buy real estates in any of the below Neighbourhoods of Queens

```
['Astoria' 'Woodside' 'Jackson Heights' 'Elmhurst' 'Howard Beach' 'Corona'
'Forest Hills' 'Kew Gardens' 'Richmond Hill' 'Flushing'
'Long Island City' 'Sunnyside' 'East Elmhurst' 'Maspeth' 'Ridgewood'
'Glendale' 'Rego Park' 'Woodhaven' 'Ozone Park' 'South Ozone Park'
'College Point' 'Whitestone' 'Bayside' 'Auburndale' 'Little Neck'
'Douglaston' 'Glen Oaks' 'Bellerose' 'Kew Gardens Hills' 'Fresh Meadows'
'Briarwood' 'Jamaica Center' 'Oakland Gardens' 'Queens Village' 'Hollis'
'South Jamaica' 'St. Albans' 'Rochdale' 'Springfield Gardens'
'Cambria Heights' 'Rosedale' 'Far Rockaway' 'Broad Channel' 'Steinway'
'Beechhurst' 'Bay Terrace' 'Edgemere' 'Arverne' 'Murray Hill'
'Floral Park' 'Holliswood' 'Jamaica Estates' 'Queensboro Hill'
'Hillcrest' 'Ravenswood' 'Lindenwood' 'Lefrak City' 'Rockaway Park'
'Bellaire' 'North Corona' 'Forest Hills Gardens' 'Jamaica Hills' 'Utopia'
'Pommonok' 'Astoria Heights' 'Hunters Point' 'Sunnyside Gardens'
'Blissville' 'Middle Village' 'Malba' 'Queensbridge']
```

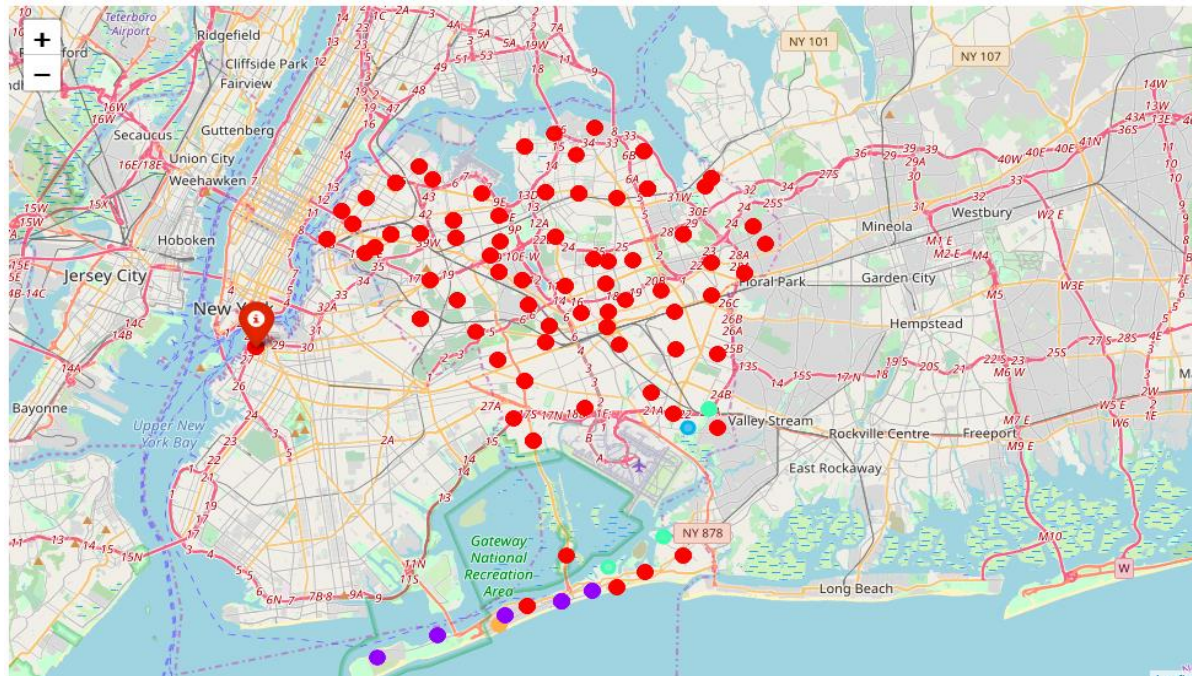


Figure 10. Visual Representations of Queens Neighborhoods similar to Brooklyn Heights.

5 Reference

1. Description of NY from Wikipedia
2. Data source Identified from IBM Course Workbooks
3. Four Square API – Free Account