Examining the residential profile of NYC neighborhoods

Introduction

New York City is one of the most populous cities in the United States and has emerged as one of the world's financial capitals. Aside from finance, the city is also a major tourist centre in the USA. The city is very diverse and is home to many kinds of people. Because the city serves so many different sectors, it can be expected that the residential profiles of the neighbourhoods differ significantly from each other. The residential profile reflects the kind of people that live in the neighbourhood and can be an interesting variable for businesses, prospective residents and tourists.

Residential profile of the neighbourhood

Different sectors in NYC are often reflected in the residential profile of a city. You can expect that Hotels will be mostly located near the big tourist sights and will be mostly located on Manhattan, while hostels will also be located close to tourist sights, but a bit further away to keep the prices low. Similarly, local residents are not interested in the tourist attraction, and as a result will live in neighbourhoods that are not close to tourist attractions. Having a spatial overview of the kinds of residential buildings and entities will provide a good idea about the kind of people that live in each neighbourhood. The Foursquare API offers an overview of all kinds of residential buildings including apartments, hotels, student resident halls and many others. The proportion of each residential type per neighbourhood can be used to cluster the neighbourhoods as a representation of different sectors.

Stakeholders

The stakeholders for this project are essentially all people that are coming to NYC and are interested in the make up of different neighbourhoods. Imagine you're a tourist coming to NYC and you're looking to find a neighbourhood to stay in, you would want to know where most of the hotels are as they tend to be near the major tourist sites. At the same time, someone that is looking to move to NYC for work isn't interested in the tourist sites, but wants to find a place that's more suited for the average resident. Business that want to cater to a specific type of person would be interested in knowing the profile of each neighbourhood to make an informed deciscion on where to place their business

Data

To find the residential profile of each neighbourhood, I will use the following sources of information: NYC has a total of 5 boroughs and 306 neighborhoods. In order to cluster the neighborhoods and explore them, we will 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 and can be found at: https://geo.nyu.edu/catalog/nyu_2451_34572
The residential data can be sourced from the Foursquare API. Example codes for residential types are 4e67e38e036454776db1fb3a, 5032891291d4c4b30a586d68, 4bf58dd8d48988d103941735,

4f2a210c4b9023bd5841ed28, 4d954b06a243a5684965b473, 52f2ab2ebcbc57f1066b8b55 and 4bf58dd8d48988d1d5941735, which range from Hotels to trailer parks.

Methodology

The first step is to load in the data from the NYC borough and their latitude and longitude values as it's important for the location clustering further on. The resulting data frame looks as follows:

	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

Afterwards, the foursquare data can be loaded into this data frame. The exact data has been described in the data section, but they are mostly different kind of residential building types ranging from hotels to trailer parks. After loading in the data, we can figure out how many venues and venues types were loaded in.

After the raw data is loaded in, we clean the data by taking out all the data that doesn't have a borough or city assigned to it. Furthermore, a lot of venues that are loaded in have nothing to do with residential types, examples are ''scenic lookout" and "pizza place". After cleaning data, we end up with only the following residential types: 'Residential Building (Apartment / Condo)', 'Hotel', 'Housing Development', 'Assisted Living', 'Bed & Breakfast', 'Motel', 'Hostel', 'College Residence Hall', 'Trailer Park', 'Vacation Rental'. The cleaned data set looks as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Name	Venue Category	Venue Latitude	Venue Longitude	Venue City	Venue State
1	Wakefield	40.894705	-73.847201	Edenwald Houses - NYCHA	Housing Development	40.886606	-73.842116	Bronx	NY
2	Wakefield	40.894705	-73.847201	The Chamber	Bed & Breakfast	40.891560	-73.844467	Bronx	NY
3	Wakefield	40.894705	-73.847201	741 Hunts Point Avenue	Residential Building (Apartment / Condo)	40.889555	-73.843807	Bronx	NY
5	Wakefield	40.894705	-73.847201	Andre Camp David (Edendwald Houses Southside)	Housing Development	40.885182	-73.844488	Bronx	NY
6	Wakefield	40.894705	-73.847201	East 238th Street	Residential Building (Apartment / Condo)	40.899167	-73.856684	Bronx	NY

After the data is cleaned, we can start counting all the venues per neighborhood, and by dividing this number by the total amount we get a mean value per residential type per neighborhood. This mean can be used to create a top 5 of the relatively most important venues per neighborhood.

Based on the mean neighborhood data, we can apply the k-means machine learning technique to cluster the neighborhoods in different clusters based on their respective mean residential type values. The clusters can then be visualized on top of a map of NYC.

Results

The amount of venues per neighborhood looks as follows:

Neighborhood	Assisted Living	Bed & Breakfast	College Residence Hall	Hostel	Hotel	Housing Development	Motel	Residential Building (Apartment / Condo)	Trailer Park	Vacation Rental
Allerton	4	2	0	0	0	13	1	26	0	0
Arden Heights	0	0	0	0	0	2	0	0	0	0
Arlington	0	0	0	0	0	2	0	1	0	0
Arrochar	0	0	0	0	1	0	0	5	0	0
Arverne	0	1	0	0	1	4	0	8	0	2
Astoria	0	0	0	0	1	2	0	44	0	0
Astoria Heights	0	0	0	2	4	0	1	12	0	0
Auburndale	3	0	0	0	0	1	0	6	0	0
Bath Beach	2	0	0	0	1	0	0	35	0	0
Battery Park City	0	0	0	0	28	0	0	18	0	0

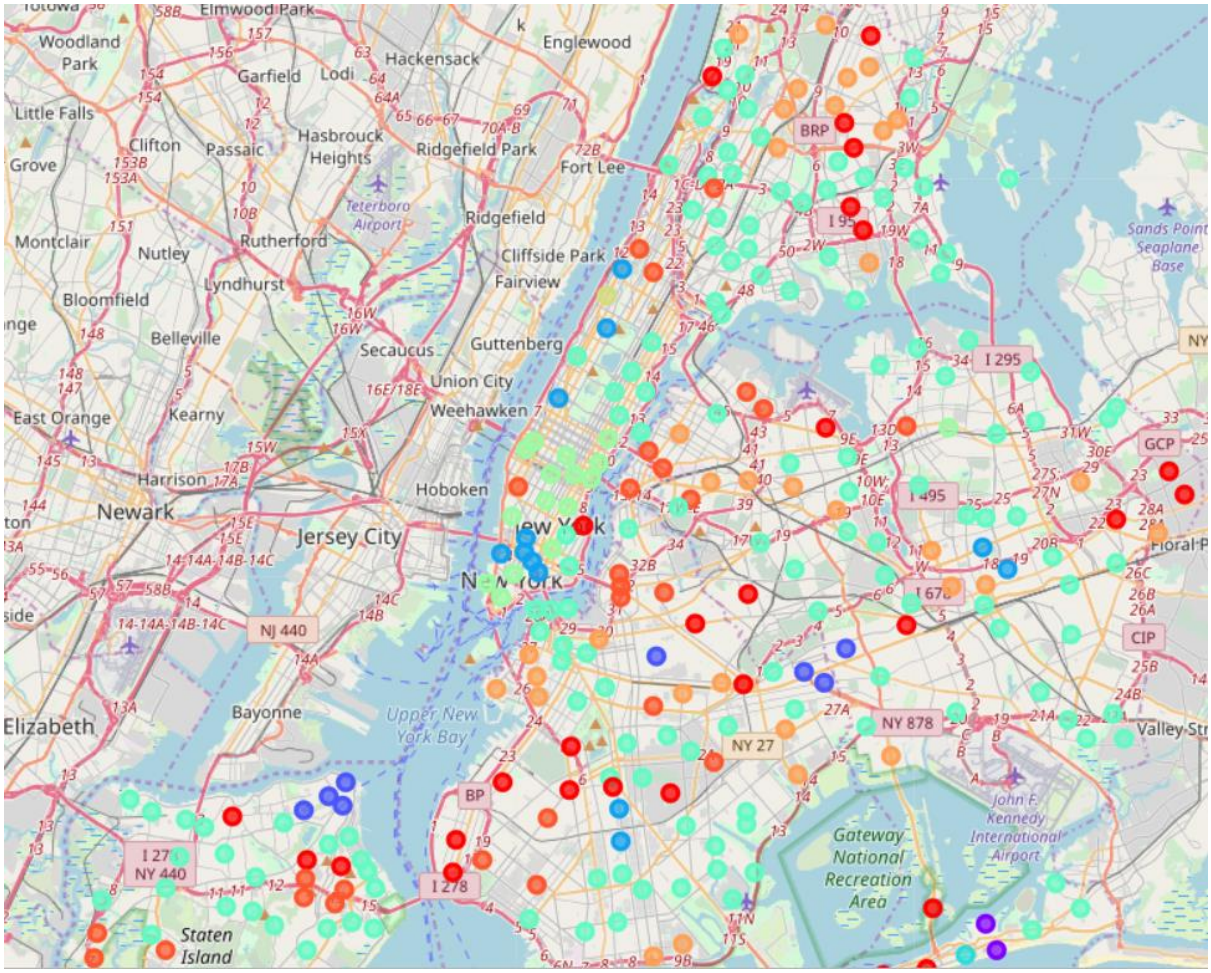
After dividing the integer numbers by their respective totals, we get a similar looking table with all the mean values for each residential building type per neighborhood:

	Neighborhood	Assisted Living	Bed & Breakfast	College Residence Hall	Hostel	Hotel	Housing Development	Motel	Residential Building (Apartment / Condo)	Trailer Park	Vacation Rental
0	Allerton	0.018519	0.027778	0.0	0.0	0.000000	0.020767	0.018182	0.004459	0.0	0.000000
1	Arden Heights	0.000000	0.000000	0.0	0.0	0.000000	0.003195	0.000000	0.000000	0.0	0.000000
2	Arlington	0.000000	0.000000	0.0	0.0	0.000000	0.003195	0.000000	0.000171	0.0	0.000000
3	Arrochar	0.000000	0.000000	0.0	0.0	0.000784	0.000000	0.000000	0.000857	0.0	0.000000
4	Arverne	0.000000	0.013889	0.0	0.0	0.000784	0.006390	0.000000	0.001372	0.0	0.333333

Based on these values we get the top 5 residential type per neighborhood:

	Neighborhood	1st Top Venue Category	2nd Top Venue Category	3rd Top Venue Category	4th Top Venue Category	5th Top Venue Category
0	Allerton	Bed & Breakfast	Housing Development	Assisted Living	Motel	Residential Building (Apartment / Condo)
1	Arden Heights	Housing Development	Vacation Rental	Trailer Park	Residential Building (Apartment / Condo)	Motel
2	Arlington	Housing Development	Residential Building (Apartment / Condo)	Vacation Rental	Trailer Park	Motel
3	Arrochar	Residential Building (Apartment / Condo)	Hotel	Vacation Rental	Trailer Park	Motel
4	Arverne	Vacation Rental	Bed & Breakfast	Housing Development	Residential Building (Apartment / Condo)	Hotel

Finally, the mean value per residential building type is used to cluster every neighborhood based on their similarity in neighborhood residential type, these can be visualized as markers on a folium map. The resulting map looks as follows:



For the analysis of each cluster I will refer you to the code as it is too long to post in this small report, during the discussion I will touch upon some of the details of the clusters.

Discussion

There are a few interesting findings that can be extracted from the results. First, we see that many clusters are based on the similarity in the first residential type category, while the rest of the top 5 seems rather random. The exception is cluster 5, which have differing number 1 residential type. We can explain this by considering this as the 'average' neighborhood where not a single economic sector such as tourism is dominant. Cluster 5 can therefore be considered the average neighborhood for local residents. In mid to lower Manhattan we see a cluster with 'hotels' as number 1 type, this can be explained by the fact that this part of NYC has the most tourist attractions, we also see student residential halls as being an important hallmark, which is logical when you consider the amount of universities in the neighborhood. In western queens and the Bronx, we see a cluster with 'hostels' as most important residential type. Hostels tend to be for budget travelers that want to be near the touristy sites, but can't afford to stay on Manhattan, explaining the popularity of these regions for hostels. Another interesting finding is that the clusters where 'Motel' is important are all near major highways, which is what defines motels.

An important note to make is that ''apartments/condo'' is the residential type is actually the most common for every neighborhood, but because every neighborhood has many of this type, it's not seen as the deciding factor in dividing the clusters. In reality, it is the most important type in every

district, e.g. you will still find more apartments than hotels in the ‘hotel’ clusters. This analysis is really based on the relative distribution of residential types.

Conclusion

For many people it would be useful to know the residential profile of the neighborhood they will stay in as it can often say something about the surround part of the city. NYC can be divided in about 10 clusters, each with a distinct profile.

- Clusters where hotels are an important part of the profile are concentrated on the mid to lower side of Manhattan.
- Hostels are found close to Manhattan, but further from the tourist location than the hotels.
- University residential halls are also found on Manhattan, near the many universities.
- Motels are found in neighborhoods that are near the major highways as can be expected.
- The teal colored cluster does not have a dominant residential building type, which means it’s most likely an average residential neighborhood, and not a neighborhood associated with a specific sector such as tourism.