

New York's Housing Prices

Alan James

May 31, 2020

Introduction

Description

New York City is the most densely populated places in the US. It has an estimated population of 8 million and so. As a person interested to settle in this city, I decided to use New York in my project. The city is divided into 5 boroughs in total. As a person generally interested in knowing where to search appropriate details, I wanted to segment New York Boroughs in a way to identify the least Housing Price.

Therefore, to get a better picture, I'd like to make those places in New York, where it'd be good to look out for a house. Also, a businessman can also get desired results opening a new venture where he could get the possible largest attention

Business Problem

The aim of this project is to identify boroughs or places with Average Housing Price, and thereby get better picture for client. Through this, a business-leading group can also make investments in places which are the top-most popular ones, and do business in the future.

Data

Data Sources

1. For spatial data, we can collect it from NYU Spatial Repository, where you could find New York .json file. But for now, I'd directly downloaded it from https://cocl.us/new_york_dataset.

This file contains the details regarding the no. of boroughs, latitude, longitude etc.

2. I also used Foursquare API for getting the top venues around a borough.
3. I created a table which listed the Average Housing Prices (\$) in different Boroughs, collecting data mainly from Google Search and just putting the data. You could get the info from Zillow website.

Methodology

Data Understanding

After getting the required data from the sources, we understand what business/client wants from the particular data. In our case, we have to determine places in New York City where Housing Price on Average is low. Thus, after understanding, we separate the unnecessary data from the out main data.

Data Preparation

In this stage, we sort the data. Mainly we clean the data for better analysis, we removed neighbourhoods' column for example to get a good output for each borough. Like that, we prepare data using data analysis tools and methods.

Data Cleaning

During Data Preparation, we clean the data and sort out the table for only boroughs, with their geographic coordinates. Also, we pull out some Housing Price data from the online resources to fit into our data. By this, we clean the data from unwanted things, leaving us the rest for further data analysis.

Data Modelling

DATA ANALYSIS

After getting enough data, we now use data wrangling to perfect our data analysis results. We also use Foursquare API to make calls which returns us the top venues for each location, based on their location parameters.

In [4]: newyork_data

```
Out[4]: {'type': 'FeatureCollection',
        'totalFeatures': 306,
        'features': [{'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]}},
                      {'type': 'Feature',
```

The data shown above is the overall data, from which I extracted only a handful for analysis purpose getting me this.

In [10]: neighborhoods

```
Out[10]:
```

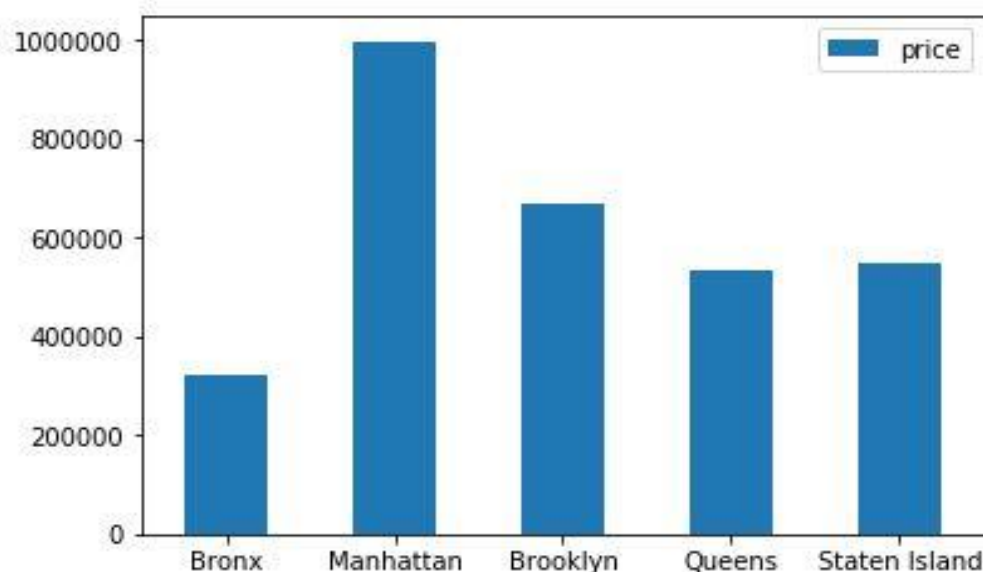
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756

After successfully getting this, now I call Foursquare API to get the top-most venues for each borough. That being said, I got the following table. The venues then were merged with the master data.

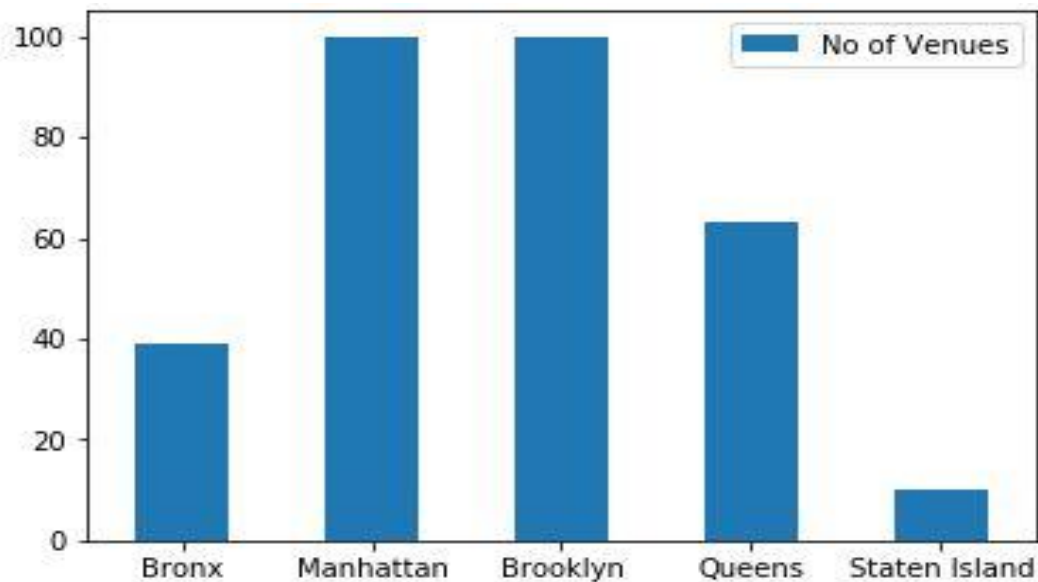
Borough	Housing Prices	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Bronx	320335	40.8448	-73.8648	Deli / Bodega	Pizza Place	Bus Station	Italian Restaurant	Restaurant	Spanish Restaurant	Playground	Home Service	Hookah Bar	Furniture / Home Store
Manhattan	998557	40.7831	-73.9712	Exhibit	Park	History Museum	Playground	Theater	Coffee Shop	Planetarium	Garden	Sushi Restaurant	Italian Restaurant
Brooklyn	669457	40.6782	-73.9442	Coffee Shop	Sandwich Place	Deli / Bodega	Southern / Soul Food Restaurant	Café	Playground	Seafood Restaurant	Chinese Restaurant	Caribbean Restaurant	Convenience Store
Queens	537350	40.7282	-73.7949	Pizza Place	Bus Station	Japanese Restaurant	Sandwich Place	Middle Eastern Restaurant	Clothing Store	Playground	Coffee Shop	Park	College Basketball Court
Staten Island	547792	40.5795	-74.1502	Golf Course	Trail	Bowling Alley	Yoga Studio	Food Truck	Department Store	Dessert Shop	Dog Run	Donut Shop	Exhibit

DATA VISUALISATION

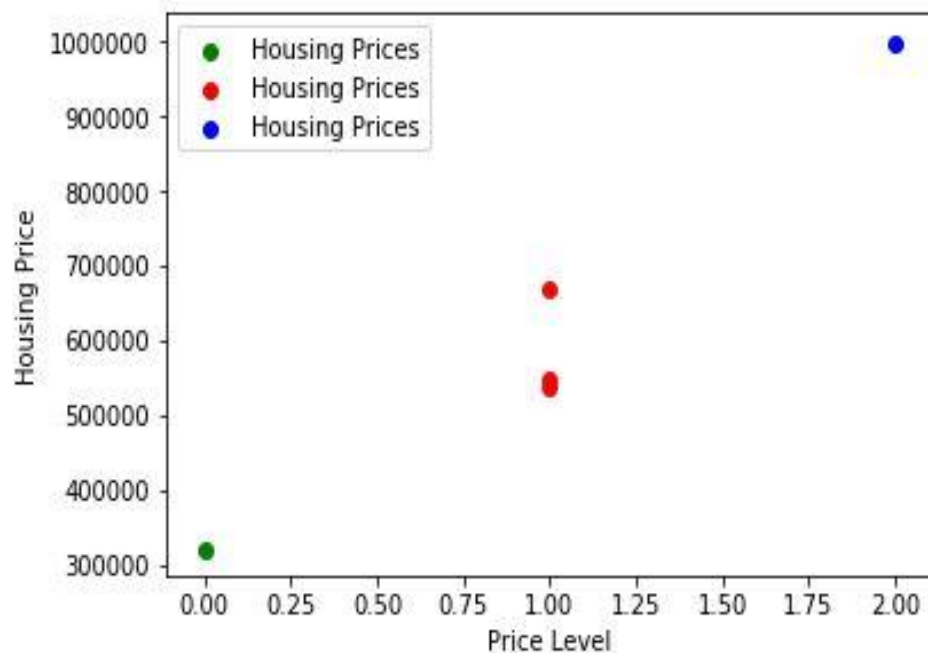
Using the housing prices data and boroughs data, I made a bar-graph to depict the categorising of places according to the housing prices.



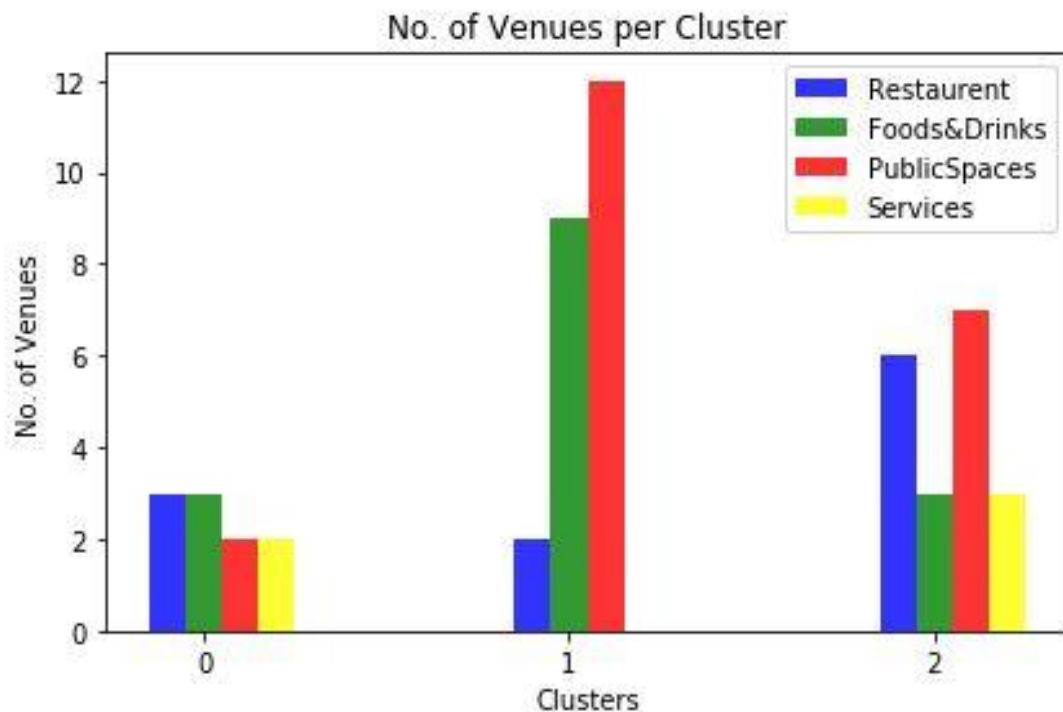
Our Foursquare API calls returned a set of number of venues for each borough. These boroughs and no. of venues were being visualised for a clear result.



After I used some ML Algorithms mainly K-Means Clustering to cluster the boroughs. After clustering the boroughs, I got this.



After Clustering, I merged the Cluster Labels and Price Level data to main data, creating another visualised dataset which shows generalised data output for things.



I used the below specified price range to divide the Housing Price

LOW LEVEL: 0-400000(\$)

MID LEVEL: 4000000-7000000(\$)

HIGH LEVEL: >7000000(\$)

ML ALGORITHMS USED: I used only K-Means to cluster the boroughs, for which I've already shown the data above.

Results

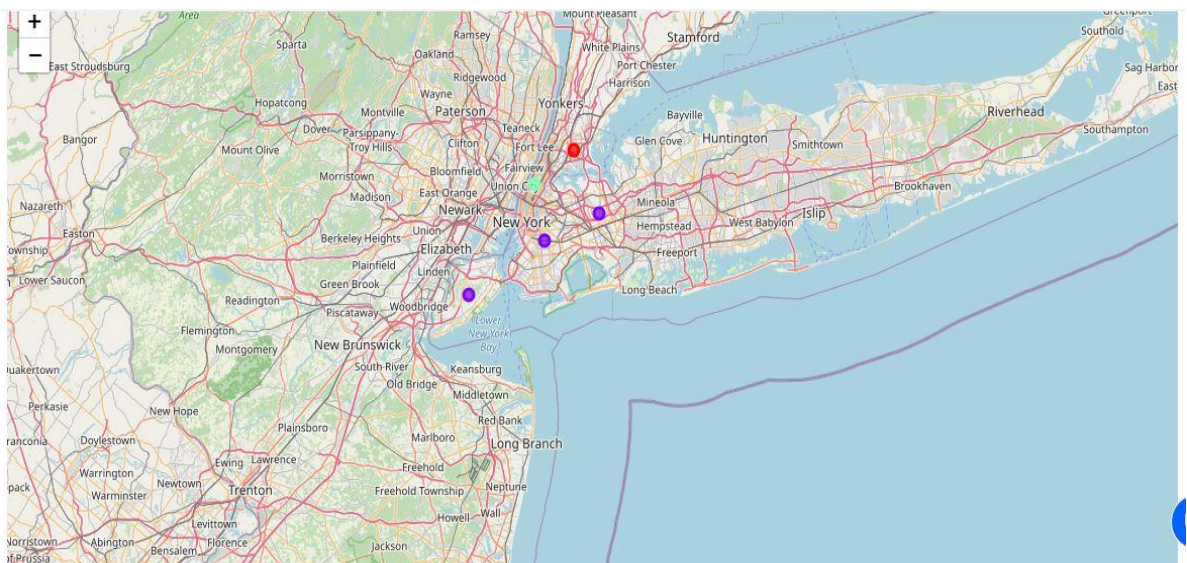
After depicting all these data, I made a dataset containing the cluster labels, price levels with the rest of the data.

Borough	Housing Prices	Latitude	Longitude	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	Price Level	Price Level Num	cluster	T
Bronx	320335	40.8448	-73.8648	Deli / Bodega	Pizza Place	Bus Station	Italian Restaurant	Restaurant	Spanish Restaurant	Playground	Home Service	Hookah Bar	Furniture / Home Store	Low	0	0	Deli / Bk Place
Manhattan	998557	40.7831	-73.9712	Exhibit	Park	History Museum	Playground	Theater	Coffee Shop	Planetarium	Garden	Sushi Restaurant	Italian Restaurant	High	2	2	Exhibiti
Brooklyn	669457	40.6782	-73.9442	Coffee Shop	Sandwich Place	Deli / Bodega	Southern / Soul Food Restaurant	Café	Playground	Seafood Restaurant	Chinese Restaurant	Caribbean Restaurant	Convenience Store	Mid	1	1	Sho PlaceDe
Queens	537350	40.7282	-73.7949	Pizza Place	Bus Station	Japanese Restaurant	Sandwich Place	Middle Eastern Restaurant	Clothing Store	Playground	Coffee Shop	Park	College Basketball Court	Mid	1	1	Pizz: Statio
Staten Island	547792	40.5795	-74.1502	Golf Course	Trail	Bowling Alley	Yoga Studio	Food Truck	Department Store	Dessert Shop	Dog Run	Donut Shop	Exhibit	Mid	1	1	CourseT

Using this above data, I also found out the top venues which could suggest a new place for business venture.

Cluster	Venues	Average Housing Price
0	3Restaurants 3Food&Drinks 2PublicSpaces 2Services	320335
1	2Restaurants 9Food&Drinks 12PublicSpaces 0Serv...	998557
2	6Restaurants 3Food&Drinks 7PublicSpaces 3Services	584866

After charting and dividing the data, we now plot it on Folium Map.



The above map has 3 clusters, divided according to their housing price.

Bluish Green Colour: Low Level Housing Price.

Violet Colour: Mid-Level HP

Red Colour: High-Level HP

Thus, we differentiated boroughs with respect to their housing prices. Also, we could use above venues data, to place a better bait for business/client side interested in the investing out.

Discussions

Through the above results, I could recommend users to look out for places which have affordable Housing Price. We could also modify this project by including the neighbourhoods' profile. But it's a tedious task, as within a single borough different neighbourhoods' maybe or maybe not have the same Housing Price. So, I decided to average out, and make out Folium Map less crowded by removing the neighbourhoods' data. I would also recommend investors to see the top-most venue in each borough and decide a venture out for them. I'd limited the venues' list, and radius from the main borough location to put enough understanding for investors. Therefore, it would definitely help in the business outlook that way.

Conclusion

Thus, through this project we found out that, we could easily point out places on the basis of Housing Price range, in the New York City. The division of boroughs into neighbourhood would affect slight variations. Besides that, given data would suffice that Housing Price would remain the same within that borough. Also, Foursquare API calls made it easier to locate the top-most venues in each borough, further showing clients what they should look for.

