House Sales Prices & Venues Data Analysis of New York City

Xia Zhang

June, 2020

Introduction

- New York City (NYC) is the most populous city in the United States.
- NYC has been described as the cultural, financial, and media capital of the world.
- Thus, NYC has attracted people over the world either to buy housing to live there or for investors to invest on properties.
- Thus, a comprehensive understanding of different areas in NYC with the corresponding statistics including housing prices, & venues information will be very critical for either a personal buyer or an investor.
- In this study, a comprehensive investigation and analysis of the mean unit housing prices in different neighborhoods of NYC was performed and the surrounding venues explored.
- The combined information of housing price and venues information was used to create a map.

Data Description

- The Department of Finance's Rolling Sales files of New York City, which lists properties that sold in the last twelve-month period in New York City for all tax classes. These files include information on:
  - The neighborhood
  - Building type
  - Square footage
  - Sale price
  - Other data
  - Data source : link1 and link2
- Geocoder.arcgis method was used to get the latitudes and longitudes pairs of each neighborhood in the housing sales file.
- Foursquare API was used to explore nearby venues for each neighborhood of NYC.

Data acquisition and cleaning

- The nyc-property-sales data from Kaggle was used as the housing sales information of New York City, which has 69987 rows and 13 columns : https://www.kaggle.com/new-york-city/nyc-property-sales
- Rows with $0 sales prices and 0 zip code values was dropped.
- A new column was created showing the building class category name, which will be used to display on the popup label in the map to be created.
- A new column “PRICE PER UNIT” was calculated and added to the dataframe.
- The cleaned dataframe was grouped by borough, neighborhood, zip code and building class category name and a new column of “Price/Unit Mean” was calculated and added to the new grouped dataframe.

ROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASE-MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	...	RESIDENTIAL UNITS	COMMERCIAL UNITS	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	TAX CLASS AT TIME OF SALE	BUILDING CLASS AT TIME OF SALE	SALE PRICE	SALE DATE
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	392	6		C2	153 AVENUE B		...	5	0	5	1633	6440	1900	2	C2	6625000	2017-07-19 00:00:00
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	26		C7	234 EAST 4TH STREET		...	28	3	31	4616	18690	1900	2	C7	-	2016-12-14 00:00:00
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	39		C7	197 EAST 3RD STREET		...	16	1	17	2212	7803	1900	2	C7	-	2016-12-09 00:00:00
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	402	21		C4	154 EAST 7TH STREET		...	10	0	10	2272	6794	1913	2	C4	3936272	2016-09-23 00:00:00
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	404	55		C2	301 EAST 10TH STREET		...	6	0	6	2369	4615	1900	2	C2	8000000	2016-11-17 00:00:00

Dataset downloaded from Kaggle website.

Dataset after cleaning and grouping.

	BOROUGH	NEIGHBORHOOD	ZIP CODE	BUILDING CLASS CATEGORY NAME	SALE PRICE MEAN	SALE PRICE MAX	SALE PRICE MIN	PRICE/UNIT MEAN	PRICE/GROSS SQFT MEAN	PRICE/LAND SQFT MEAN
0	Bronx	BATHGATE	10451	RENTALS - WALKUP APARTMENTS	4.461000e+06	4461000.0	4461000.0	1.715769e+05	244.000000	1323.000000
1	Bronx	BATHGATE	10456	RELIGIOUS FACILITIES	4.000000e+05	400000.0	400000.0	4.000000e+05	57.000000	160.000000
2	Bronx	BATHGATE	10457	COMMERCIAL GARAGES	2.100000e+06	2500000.0	1700000.0	1.675000e+06	inf	573.500000
3	Bronx	BATHGATE	10457	COOPS - ELEVATOR APARTMENTS	4.666667e+04	60000.0	40000.0	4.666667e+04	inf	inf
4	Bronx	BATHGATE	10457	FACTORIES	1.725000e+06	3000000.0	375000.0	1.725000e+06	153.666667	254.000000
5	Bronx	BATHGATE	10457	ONE FAMILY DWELLINGS	2.470067e+05	420000.0	127000.0	2.470067e+05	169.666667	119.333333
6	Bronx	BATHGATE	10457	RENTALS - WALKUP APARTMENTS	1.086667e+06	1980000.0	400000.0	1.260073e+05	155.000000	327.000000
7	Bronx	BATHGATE	10457	TAX CLASS 4 - OTHER	3.000000e+05	300000.0	300000.0	3.000000e+05	inf	97.000000
8	Bronx	BATHGATE	10457	THREE FAMILY DWELLINGS	4.826798e+05	689399.0	299000.0	1.608933e+05	173.400000	324.400000
9	Bronx	BATHGATE	10457	TWO FAMILY DWELLINGS	3.898750e+05	700000.0	200000.0	1.720208e+05	164.750000	199.750000
10	Bronx	BATHGATE	10457	WAREHOUSES	1.900000e+06	1900000.0	1900000.0	1.900000e+06	318.000000	217.000000

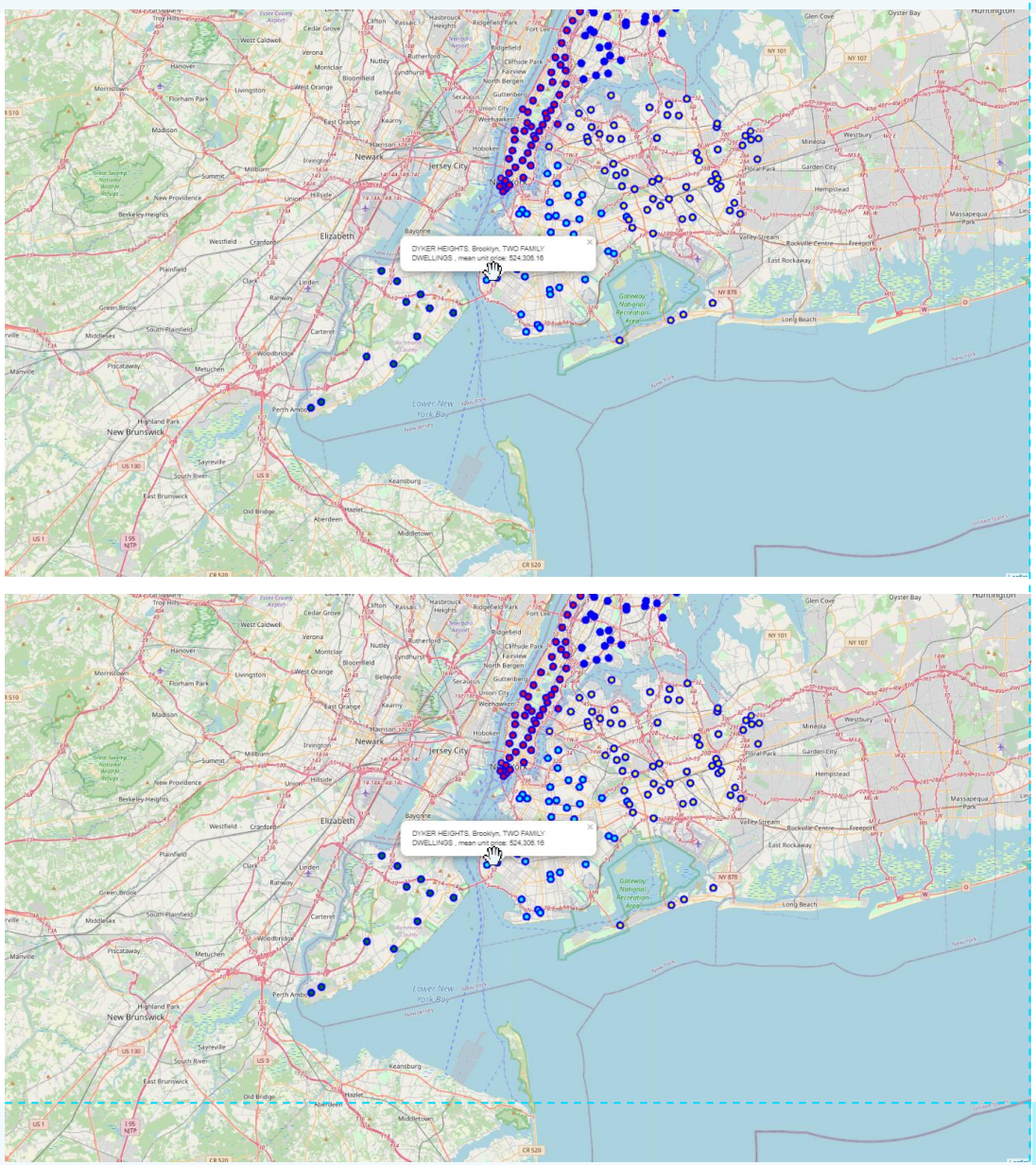
Explore the neighborhoods in New York City

- Geocoder.arcgis was used to obtain the latitudes and longitudes coordinates of NYC neighborhoods through the zip codes.
- Below shows the dataframe with latitudes and longitudes data.

	BOROUGH	NEIGHBORHOOD	ZIP CODE	BUILDING CLASS CATEGORY NAME	LONGITUDE	LATITUDE	SALE PRICE MEAN	SALE PRICE MAX	SALE PRICE MIN	PRICE/UNIT MEAN	PRICE/GROSS SQFT MEAN	PRICE/LAND SQFT MEAN
0	Bronx	BATHGATE	10451	RENTALS - WALKUP APARTMENTS	-73.918433	40.819986	4.461000e+06	4461000.0	4461000.0	1.715769e+05	244.000000	1323.000000
1	Bronx	BATHGATE	10456	RELIGIOUS FACILITIES	-73.896685	40.833955	4.000000e+05	400000.0	400000.0	4.000000e+05	57.000000	160.000000
2	Bronx	BATHGATE	10457	COMMERCIAL GARAGES	-73.903813	40.848111	2.100000e+06	2500000.0	1700000.0	1.675000e+06	inf	573.500000
3	Bronx	BATHGATE	10457	COOPS - ELEVATOR APARTMENTS	-73.903813	40.848111	4.666667e+04	60000.0	40000.0	4.666667e+04	inf	inf
4	Bronx	BATHGATE	10457	FACTORIES	-73.903813	40.848111	1.725000e+06	3000000.0	375000.0	1.725000e+06	153.666667	254.000000
5	Bronx	BATHGATE	10457	ONE FAMILY DWELLINGS	-73.903813	40.848111	2.470067e+05	420000.0	127000.0	2.470067e+05	169.666667	119.333333
6	Bronx	BATHGATE	10457	RENTALS - WALKUP APARTMENTS	-73.903813	40.848111	1.086667e+06	1980000.0	400000.0	1.260073e+05	155.000000	327.000000

Create the 1st Map

- Map of housing located at different neighborhood of New York City for the five boroughs, each borough shown with a different color.
- A popup label was created to display some key information including the neighborhood, borough name, building class category name, and the mean unit price sold for this type of building.



Create the 1st Map

- This map can be very useful for people who are looking for buying a either an apartment, a house, or investing apartments.
- By clicking on the map, one can easily find the mean unit sale prices as well as its neighborhood and how the price is compared to similar or different building in the same or different neighborhoods.

Explore the nearby venues of NYC to combine with house sales information

- Foursquare API was used to explore the neighborhoods and segment them.
- The dataset again by only the borough, neighborhood and zip code.
- Searching criteria: radius: 180 meters , limit: 30 venues
- The two files on the right shows the file used for searching nearby venues. The lower one has latitudes and longitudes coordinates.

```
nyc_grouped2.head(10)
```

```
(613, 7)
```

	BOROUGH	NEIGHBORHOOD	ZIP CODE	SALE PRICE MEAN	SALE PRICE MAX	SALE PRICE MIN	PRICE/UNIT MEAN
0	Bronx	BATHGATE	10451	4.461000e+06	4461000.0	4461000.0	171576.920000
1	Bronx	BATHGATE	10456	4.000000e+05	400000.0	400000.0	400000.000000
2	Bronx	BATHGATE	10457	6.871700e+05	3000000.0	40000.0	465802.968125
3	Bronx	BATHGATE	10458	9.413929e+05	4052000.0	18000.0	229476.034167
4	Bronx	BAYCHESTER	10466	4.748387e+05	3000000.0	10.0	328506.865736
5	Bronx	BAYCHESTER	10469	3.938067e+05	4750000.0	1.0	281702.618589
6	Bronx	BAYCHESTER	10475	8.860018e+05	11000000.0	123000.0	722394.645417
7	Bronx	BEDFORD PARK/NORWOOD	10458	6.830143e+05	7000000.0	10.0	259213.154110
8	Bronx	BEDFORD PARK/NORWOOD	10467	9.544206e+05	7375000.0	10.0	438735.046731
9	Bronx	BEDFORD PARK/NORWOOD	10468	9.497747e+05	11118000.0	72000.0	275719.651379

```
nyc_grouped2.head(3)
```

	BOROUGH	NEIGHBORHOOD	ZIP CODE	LATITUDE	LONGITUDE	SALE PRICE MEAN	SALE PRICE MAX	SALE PRICE MIN	PRICE/UNIT MEAN
0	Bronx	BATHGATE	10451	40.819986	-73.918433	4.461000e+06	4461000.0	4461000.0	171576.920000
1	Bronx	BATHGATE	10456	40.833955	-73.896685	4.000000e+05	400000.0	400000.0	400000.000000
2	Bronx	BATHGATE	10457	40.848111	-73.903813	6.871700e+05	3000000.0	40000.0	465802.968125

Explore the nearby venues of NYC to combine with house sales information

- Below shows the output file from searching nearby venues
- 247 unique venue categories was returned

```
nyc_venues.shape
```

```
(4621, 9)
```

```
nyc_venues.head()
```

	Borough	Neighborhood	Zip Code	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	BATHGATE	10451	40.819986	-73.918433	Pepe Joes	40.820838	-73.917523	Pizza Place
1	Bronx	BATHGATE	10456	40.833955	-73.896685	JC Wines & liqours	40.834490	-73.896980	Liquor Store
2	Bronx	BATHGATE	10456	40.833955	-73.896685	Prospect Gourmet Deli	40.833674	-73.896681	Deli / Bodega
3	Bronx	BATHGATE	10456	40.833955	-73.896685	Fine Fare Supermarket	40.834043	-73.894640	Supermarket
4	Bronx	BATHGATE	10456	40.833955	-73.896685	FOOT LOCKER - CONCOURSE PLAZA	40.834428	-73.895602	Shoe Store

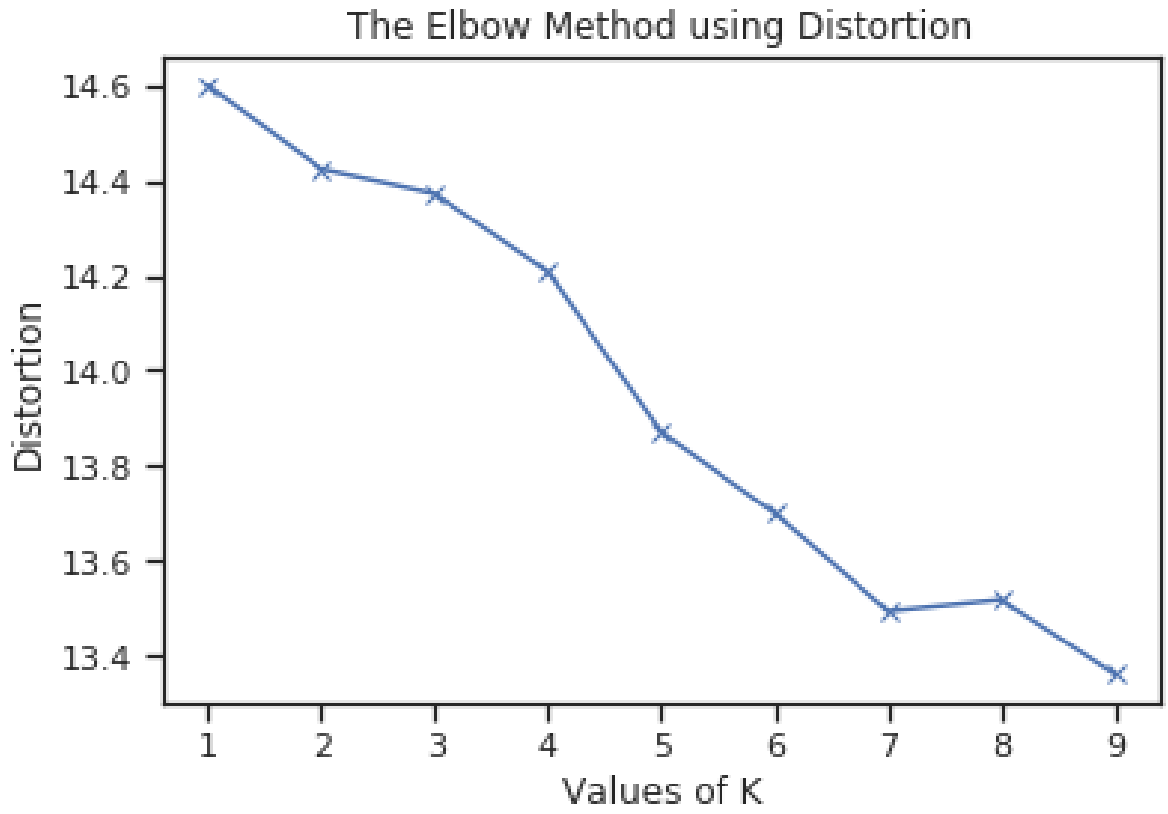
Analyze each neighborhood

- By applying pandas.get_dummies method, a file (nyc_onehot) containing the encoded venue category information was obtained.
- Then the venues were sorted to return the most common venues for each neighborhood.

	Borough	Neighborhood	Zip_Code	Location	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
0	Bronx	BATHGATE	10451	Bronx,BATHGATE,10451	Pizza Place	Yoga Studio	Entertainment Service	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market
1	Bronx	BATHGATE	10456	Bronx,BATHGATE,10456	Deli / Bodega	Liquor Store	Shoe Store	Supermarket	Bus Station	Yoga Studio	Fast Food Restaurant	Exhibit
2	Bronx	BATHGATE	10457	Bronx,BATHGATE,10457	Deli / Bodega	Sandwich Place	Grocery Store	Yoga Studio	Entertainment Service	Food & Drink Shop	Food	Flower Shop
3	Bronx	BATHGATE	10458	Bronx,BATHGATE,10458	Coffee Shop	Sandwich Place	Wine Bar	Yoga Studio	Fast Food Restaurant	Event Space	Exhibit	Falafel Restaurant
4	Bronx	BAYCHESTER	10466	Bronx,BAYCHESTER,10466	Diner	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flower Shop

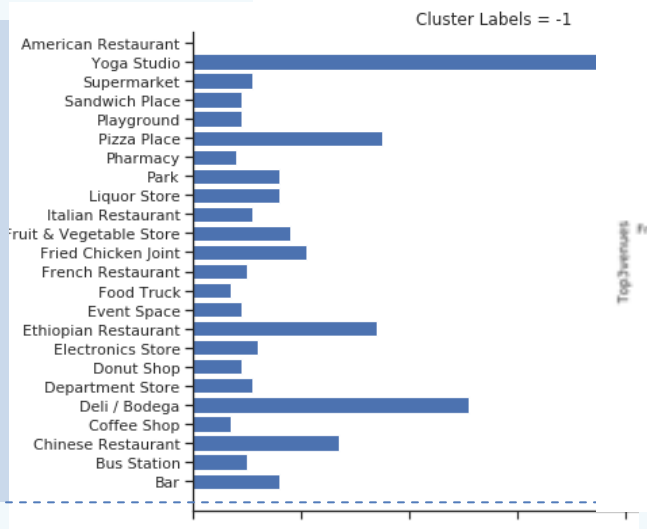
Cluster the neighborhoods: K-Means

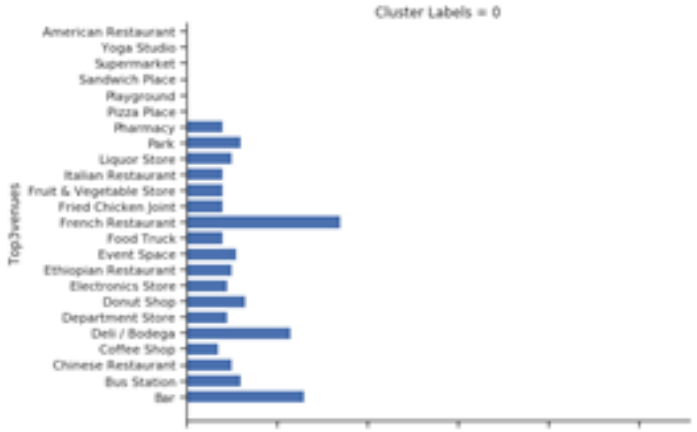
- To determine the best number of K, the elbow method using distortion was performed.
- However, no elbow was found from the graph.
- No elbow means that the algorithm used cannot separate clusters.
- Thus, the DBSCAN method was used instead.

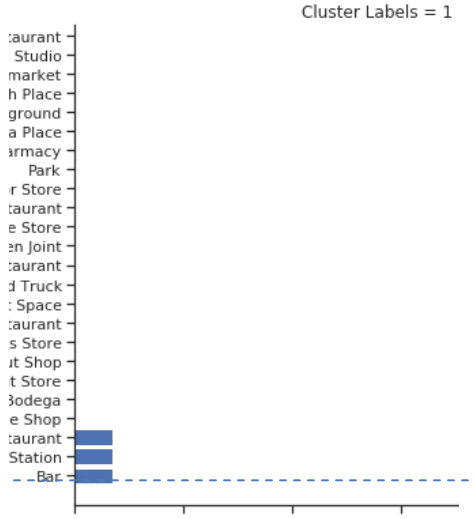


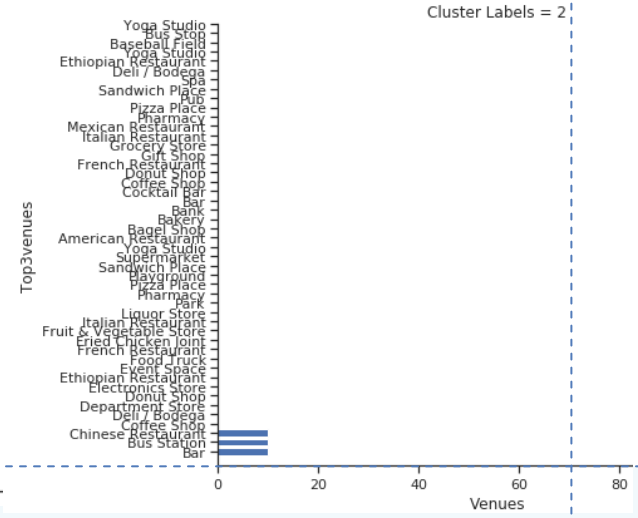
Cluster the neighborhoods: DBSCAN

- A total of 4 clusters was obtained from DBSCAN:
- To show the cluster name based on venues frequencies, bar charts of venues frequency for the top 3 most common venues were created for each cluster.
- Clusters were named as follows:
  - Cluster -1: Social, Shopping & Entertainment Complex
  - Cluster 0: French Restaurant, Bar & Misc
  - Cluster 1: Chinese Restaurant & Bar_1
  - Cluster -1: Chinese Restaurant & Bar_2









Create the 2nd Map

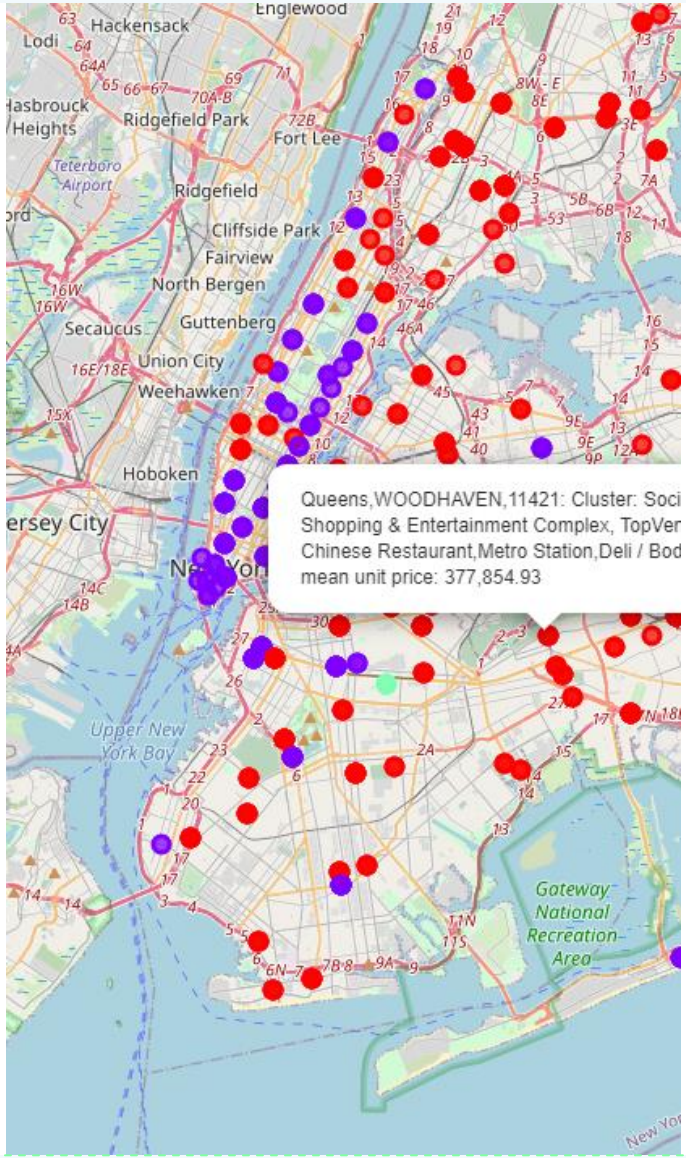
- The venues data searched was merged with the nyc_grouped 2 data, which includes the mean unit building sales price info.
- A new column was created which shows the joined name of the top 3 most common venues, named “Top_Common_Venues”

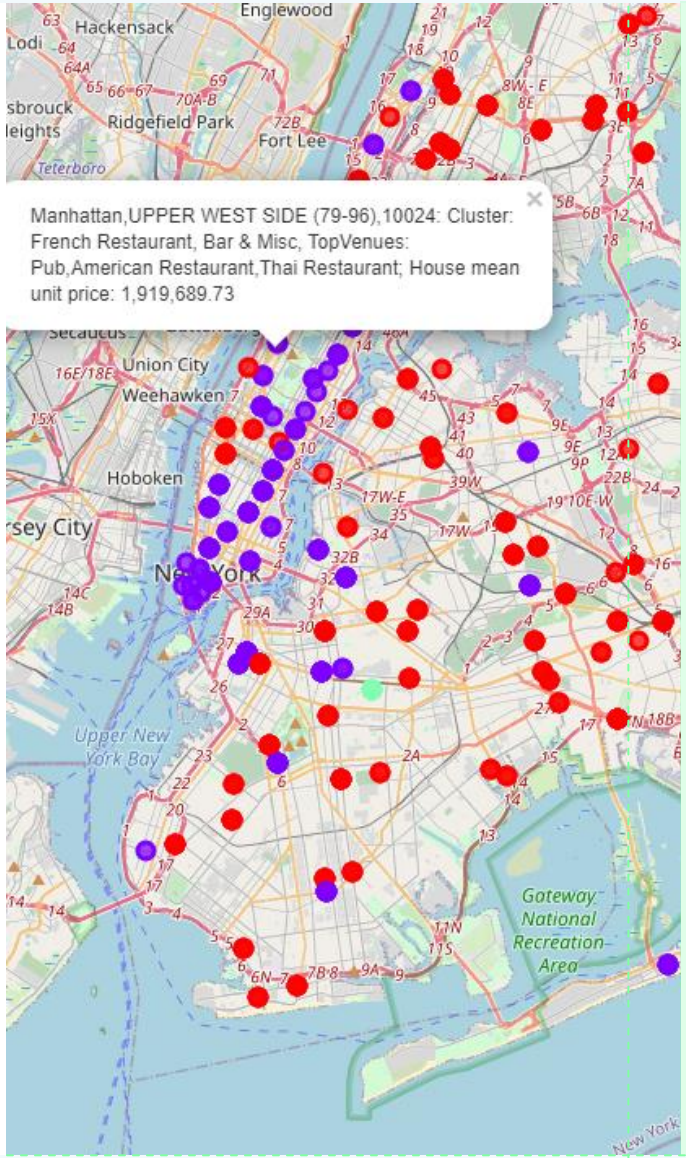
	Borough	Neighborhood	Zip_Code	Location	Latitude	Longitude	Sale Price Mean	Sale Price Max	Sale Price Min	Price/Unit Mean	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6 C
0	Bronx	BATHGATE	10451	Bronx,BATHGATE,10451	40.819986	-73.918433	4.461000e+06	4461000.0	4461000.0	171576.920000	Social, Shopping & Entertainment Complex	Pizza Place	Yoga Studio	Ethiopian Restaurant	French Restaurant	Food Truck	
1	Bronx	BATHGATE	10456	Bronx,BATHGATE,10456	40.833955	-73.896685	4.000000e+05	400000.0	400000.0	400000.000000	Social, Shopping & Entertainment Complex	Liquor Store	Playground	Deli / Bodega	Supermarket	Shoe Store	
2	Bronx	BATHGATE	10457	Bronx,BATHGATE,10457	40.848111	-73.903813	6.871700e+05	3000000.0	40000.0	465802.968125	Social, Shopping & Entertainment Complex	Deli / Bodega	Grocery Store	Sandwich Place	Yoga Studio	Ethiopian Restaurant	
3	Bronx	BATHGATE	10458	Bronx,BATHGATE,10458	40.862059	-73.887575	9.413929e+05	4052000.0	18000.0	229476.034167	Social, Shopping & Entertainment Complex	Coffee Shop	Theater	Sandwich Place	Café	Yoga Studio	
4	Bronx	BAYCHESTER	10466	Bronx,BAYCHESTER,10466	40.887857	-73.827943	4.748387e+05	3000000.0	10.0	328506.865736	Social, Shopping & Entertainment Complex	Diner	Business Service	Yoga Studio	Event Space	French Restaurant	

Most mon /enue	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	Top_Common_Venues
opian urant	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Filipino Restaurant	Pizza Place,Yoga Studio,Ethiopian Restaurant
Deli / odega	Supermarket	Shoe Store	Bus Station	Yoga Studio	Fast Food Restaurant	Exhibit	Falafel Restaurant	Liquor Store,Playground,Deli / Bodega
dwich Place	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Deli / Bodega,Grocery Store,Sandwich Place
dwich Place	Café	Yoga Studio	Farmers Market	Exhibit	Falafel Restaurant	Fast Food Restaurant	Ethiopian Restaurant	Coffee Shop,Theater,Sandwich Place
Yoga tudio	Event Space	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Diner,Business Service,Yoga Studio

Create the 2nd Map

- Finally, we can visualize the data in a map using Folium:
- The map shows clustered neighborhoods of NYC with combined information of mean unit housing price and venues information.
- The popup labels display the borough, neighborhood, top surrounding venues and the mean unit housing prices.
- This map will be very useful for house buyer or investors to review and compare housing sale prices as well as what is the characteristic venues nearby.





Conclusion

- Two maps were created for New York City neighborhoods.
  - 1st Map:
    - shows the NYC housing sold at different neighborhoods in the 5 boroughs with different color notations.
    - Each house sold was shown on the map with a popup label with information including: neighborhood, borough name, building class category name, and the mean unit price sold for this type of building.
  - 2nd Map:
    - Nearby venues information was combined with the average unit sale price for housing in that neighborhood.
    - The popup labels display the borough, neighborhood, top surrounding venues and the mean unit housing prices.
- Combined use of Map 1 and Map 2 created in this study will be even more powerful for house buyers or investors to get more comprehensive information on average housing price, building class category names, neighborhood and top venues nearby.