

Enhanced Manhattan Property Search

See associated report for further details

The Problem

Our client wants to find an apartment in Manhattan with particular venues available.

They are totally unfamiliar with Manhattan and find it tedious to have to analyze each available property address to find if it has the venues they require.

The Data

- The fundamental neighborhood data was provided in the earlier exercise in the file `newyork_data.json`. Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.
- As in the earlier exercise, venue information for the neighborhoods was acquired from the Foursquare api at “api.foursquare.com”.
- Candidate property data was acquired from New York City sales data at <https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page>.

The Methodology

- Acquisition and processing of New York neighborhood venue data
- *Used neighborhood location information to access venue information from Foursquare.*
- *Then process that information to establish a neighborhood/venue-type frequency dataframe*



The Manhattan dataframe has 40 neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

	Neighborhood	Accessories Store	Acupuncturist	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant
0	Battery Park City	0.0	0.0	0.0	0.0	0.000000	0.010753
1	Carnegie Hill	0.0	0.0	0.0	0.0	0.000000	0.011111
2	Central Harlem	0.0	0.0	0.0	0.0	0.068182	0.045455
3	Chelsea	0.0	0.0	0.0	0.0	0.000000	0.040000
4	Chinatown	0.0	0.0	0.0	0.0	0.000000	0.040000

- Acquisition and processing of available property data

- *Used Manhattan property sales data, as free resource of real available properties not found*
- *Needed to add latitude and longitude with geocode.*
- *Needed to assign closest “Foursquare Neighborhood” based on property latitude and longitude, since there was a poor match between the data sources and property data often appeared inaccurate.*

	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	ADDRESS	ZIP CODE	YEAR BUILT	SALE PRICE	latitude	longitude	Foursquare Neighborhood
3	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	233 EAST 3 STREET	10009	1910	7,500,000	43.150612	-77.597139	Inwood
36	CLINTON	07 RENTALS - WALKUP APARTMENTS	2	446 WEST 55TH STREET	10019	1901	4,550,000	40.767257	-73.988725	Lincoln Square
70	FASHION	07 RENTALS - WALKUP APARTMENTS	2	532 9 AVENUE	10018	1901	4,000,000	40.808930	-73.953850	Morningside Heights
71	GRAMERCY	07 RENTALS - WALKUP APARTMENTS	2A	129 EAST 17TH STREET	10003	1900	0	40.735814	-73.986666	Gramercy
79	GREENWICH VILLAGE-CENTRAL	07 RENTALS - WALKUP APARTMENTS	2	30 EAST 14TH STREET	10003	1910	23,500,000	40.735139	-73.992286	Flatiron

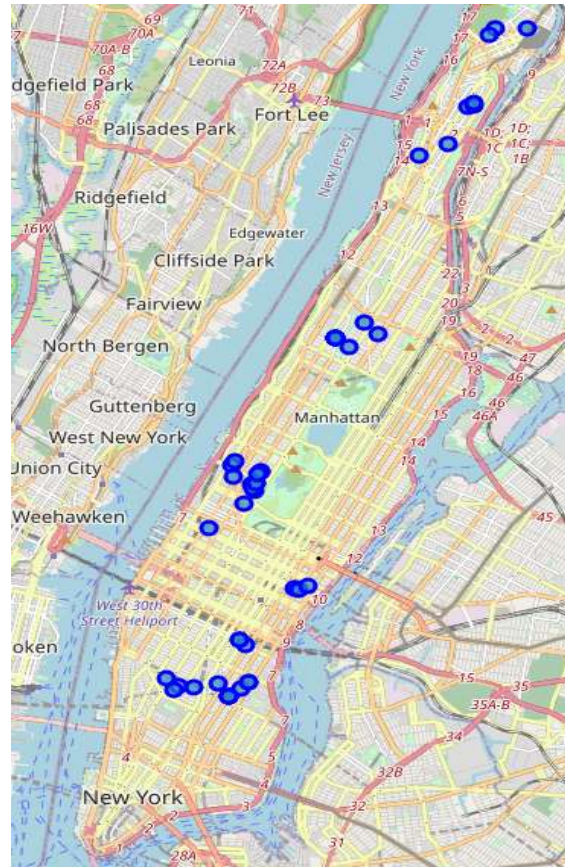
- Utilize the venue and property data to provide a client with properties meeting their criteria
 - *Started with the assumption the client wants to require a neighborhood to include a Grocery Store, Pharmacy, and Coffee Shop*
 - *Created a “scored” neighborhood dataframe that includes the required venue scores, the total of those scores, and a Boolean to indicate if all the required venues were present.*
 - *Reduced that dataframe to only those neighborhoods that included all the require venues..*

	Neighborhood	Coffee Shop	Grocery Store	Pharmacy	Score	All Present
10	Flatiron	0.030000	0.010000	0.010000	0.050000	True
11	Gramercy	0.042553	0.021277	0.010638	0.074468	True
15	Inwood	0.018868	0.018868	0.018868	0.056604	True
17	Lincoln Square	0.030612	0.010204	0.010204	0.051020	True
25	Morningside Heights	0.073171	0.024390	0.024390	0.121951	True
26	Murray Hill	0.040000	0.020000	0.010000	0.070000	True
34	Turtle Bay	0.050000	0.010000	0.010000	0.070000	True
37	Washington Heights	0.022727	0.022727	0.011364	0.056818	True

- I then joined the reduced property dataframe with “scored” neighborhood dataframe to have the scores directly available with each eligible property. This sorted list now provided one of the desired outputs of a venue-scored ordered list of properties.

ADDRESS	ZIP CODE	YEAR BUILT	SALE PRICE	latitude	longitude	Foursquare Neighborhood	Neighborhood	Coffee Shop	Grocery Store	Pharmacy	Score	All Present
353 WEST 115 STREET	10026	1900	11,000,000	40.805649	-73.960339	Morningside Heights	Morningside Heights	0.073171	0.024390	0.024390	0.121951	True
344 MANHATTAN AVENUE	10026	1900	1,623,468	40.803910	-73.957328	Morningside Heights	Morningside Heights	0.073171	0.024390	0.024390	0.121951	True
204 WEST 121ST STREET	10027	1910	4	40.806486	-73.950625	Morningside Heights	Morningside Heights	0.073171	0.024390	0.024390	0.121951	True
213 WEST 115 STREET	10026	1900	810,630	40.805649	-73.960339	Morningside Heights	Morningside Heights	0.073171	0.024390	0.024390	0.121951	True

- Provided a map of the eligible properties.



Conclusion

This project was able to successfully demonstrate that data (Foursquare venue data) could be used to solve an interesting problem (Client desire to filter property search results based on venue information).