

Post-pandemic: where to open a new Chinese restaurant

Introduction/Business Problem

I have been living in Manhattan for more than seven years. As a result of Covid-19 restrictions, a lot of my beloved restaurant/bakery/Café are permanently closed.

On Wednesday, June 9, 2021, New York Gov. Andrew Cuomo speaks at the opening ceremony of the Tribeca Film Festival. With 70% vaccinated, New York State is reopening from most Covid-19 restrictions.

I would like to use Foursquare location data to see which area in Manhattan has best potential to open a new Chinese restaurant.

Data Selection and Cleaning

Even though this course already has detailed guide for getting New York City Data, we'd like to use latest 2021 data obtained from [www.propertyshark.com](https://www.propertyshark.com/mason/market-trends/residential/nyc/manhattan) for two main reasons:

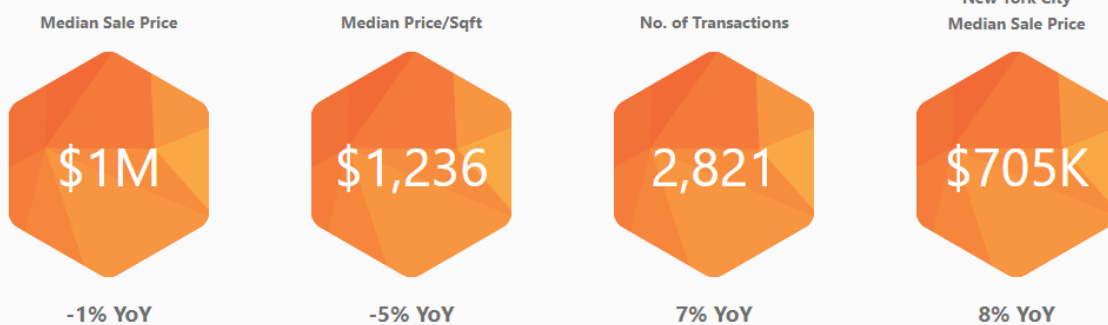
1. Up-to-date Neighborhood, focused for Manhattan only
2. Real Estate Median Sale Price could be an indicator of neighborhood purchasing power

url = <https://www.propertyshark.com/mason/market-trends/residential/nyc/manhattan>

Manhattan Real Estate Market Trends

Stats: Quarterly | Monthly

Market Overview for Quarter 1, 2021



In Quarter 1, the median home sale price in Manhattan was \$1M, a -1% change year-over-year. A total of 2821 properties changed hands, up 7% compared to the same month last year. In Quarter 1, the median price per square foot was \$1,236, a -5% YoY change. The median home sale price in New York City was \$705K.

Most Expensive Neighborhoods in Manhattan

Search:

Neighborhood	Borough	Median Sale Price
Hudson Yards	Manhattan	\$5,504,250
TriBeCa	Manhattan	\$3,550,000
SoHo	Manhattan	\$3,250,000
Hudson Square	Manhattan	\$2,906,012
Civic Center	Manhattan	\$2,760,000
NoLita	Manhattan	\$2,647,450
Central Park South	Manhattan	\$2,147,500
NoHo	Manhattan	\$1,935,000
Little Italy	Manhattan	\$1,925,000
Garment District	Manhattan	\$1,900,000
Flatiron District	Manhattan	\$1,865,000
Chinatown	Manhattan	\$1,733,179

Additionally, I use “**Restaurant**” as keyword to do the venues search to exclude useless venues. I set radius as 800 meters which is almost 1-mile diameter. Even I set limit to high number, I notice that foursquare can only return up to 50 venues according to the guide:

<https://developer.foursquare.com/docs/api-reference/venues/search/>



So, the returned number of venues is not real total number of restaurants, except neighborhoods which has relatively low real estate price, most neighborhood has around 50 restaurant info returned.

Next, I need identify whether returned venue is a Chinese restaurant or not. I noticed that there are some more problems regarding foursquare api data

1. Some result is not restaurant; we need drop them.
2. Chinese restaurants are labeled as specific type restaurant, like 'Dim Sum Restaurant', 'Cantonese Restaurant', 'Szechuan Restaurant', we need combine them together.
3. Additionally, I need label other Asian restaurant as potential competitors, including 'Thai Restaurant', 'Soba Restaurant', 'Korean Restaurant', 'Japanese Restaurant', 'Sushi Restaurant', 'Vietnamese Restaurant', 'Ramen Restaurant', 'Asian Restaurant'. Other type of restaurants will be labeled as "Other"

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	New Category
0	Battery Park City	40.711017	-74.016937	O'Hara's Restaurant & Pub	40.709894	-74.012836	Pub	Other
1	Battery Park City	40.711017	-74.016937	Mudville Restaurant & Tap House	40.715209	-74.008923	Wings Joint	Other
2	Battery Park City	40.711017	-74.016937	Amore's Pizza Restaurant	40.715860	-74.009888	Pizza Place	Other
3	Battery Park City	40.711017	-74.016937	Yips Restaurant	40.704950	-74.012360	Chinese Restaurant	Chinese Restaurant
4	Battery Park City	40.711017	-74.016937	Su Hang Restaurant	40.708363	-74.014046	Chinese Restaurant	Chinese Restaurant

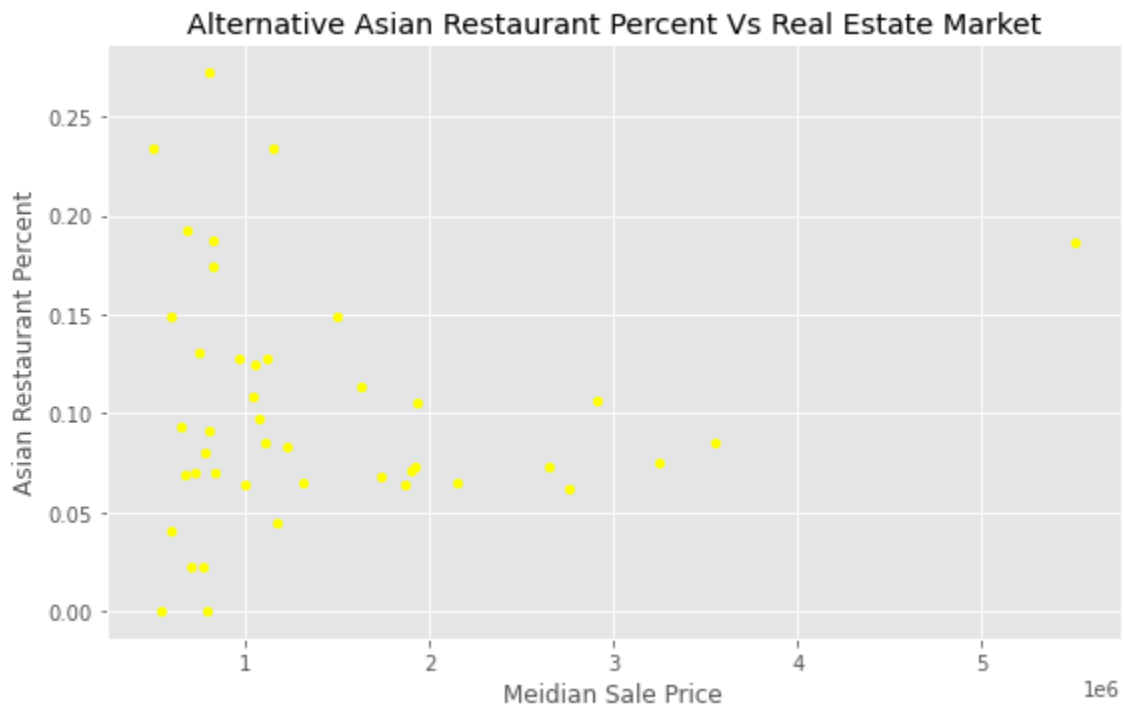
When new category is constructed successfully, instead of using absolute value, I would use restaurant type percent to do the analysis. Final table is like below. Alternative + Chinese Restaurant + Other =100%

	Borough	Neighborhood	Latitude	Longitude	Median Sale Price	Venue	Alternative	Chinese Restaurant	Other
0	Manhattan	Battery Park City	40.711017	-74.016937	762900.0	44	0.022727	0.113636	0.863636
1	Manhattan	Carnegie Hill	40.784197	-73.954339	1500000.0	47	0.148936	0.042553	0.808511
2	Manhattan	Central Midtown	40.756182	-73.981039	830000.0	43	0.069767	0.046512	0.883721
3	Manhattan	Central Park South	40.767814	-73.981232	2147500.0	46	0.065217	0.043478	0.891304
4	Manhattan	Chelsea	40.746491	-74.001528	1077600.0	41	0.097561	0.146341	0.756098
5	Manhattan	Chinatown	40.716491	-73.996250	1733179.0	44	0.068182	0.636364	0.295455

Methodology

We can start from simple data visualization – scattered plot. Based on the plotting, I get two clues:

1. The neighborhood could be separated into different groups, mainly based on real estate market median sale price
2. Within same group, neighborhood with relative low Chinese restaurant/Asian restaurant ratio could be a reasonable choice to open a new Chinese restaurant



Using K-means clustering methodology, all Manhattan neighborhoods could be categorized as five groups (m stand for million dollar):

0.9-m club

Cluster Labels	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other	
0	4	Manhattan	Battery Park City	40.711017	-74.016937	762900.0	44	0.022727	0.113636	0.863636
2	4	Manhattan	Central Midtown	40.756182	-73.981039	830000.0	43	0.093023	0.023256	0.883721
7	4	Manhattan	Clinton - Hell's Kitchen	40.768228	-73.994269	779500.0	50	0.060000	0.000000	0.940000
8	4	Manhattan	East Harlem	40.794722	-73.942500	790000.0	48	0.000000	0.166667	0.833333
15	4	Manhattan	Harlem	40.807879	-73.945415	705000.0	45	0.022222	0.088889	0.888889
18	4	Manhattan	Inwood	40.869258	-73.920495	535000.0	33	0.000000	0.181818	0.818182
19	4	Manhattan	Kips Bay	40.739546	-73.977083	597000.0	47	0.127660	0.191489	0.680851
20	4	Manhattan	Koreatown	40.747642	-73.986516	645000.0	43	0.093023	0.139535	0.767442
24	4	Manhattan	Lower East Side	40.715936	-73.986806	728000.0	43	0.069767	0.395349	0.534884
25	4	Manhattan	Manhattan Valley	40.799776	-73.967772	800000.0	44	0.090909	0.136364	0.772727
26	4	Manhattan	Morningside Heights	40.810000	-73.962500	670000.0	29	0.068966	0.241379	0.689655
27	4	Manhattan	Murray Hill	40.748157	-73.978750	750000.0	46	0.130435	0.173913	0.695652
31	4	Manhattan	Roosevelt Island	40.761418	-73.950228	676551.0	26	0.192308	0.115385	0.692308
33	4	Manhattan	Sutton Place	40.758037	-73.960182	803750.0	44	0.272727	0.022727	0.704545
36	4	Manhattan	Tudor City	40.748623	-73.971389	492500.0	47	0.255319	0.170213	0.574468
37	4	Manhattan	Turtle Bay	40.753467	-73.968866	820000.0	48	0.208333	0.062500	0.729167
40	4	Manhattan	Washington Heights	40.840198	-73.940221	590000.0	49	0.040816	0.122449	0.836735
42	4	Manhattan	Yorkville	40.778007	-73.948202	824000.0	46	0.152174	0.065217	0.782609

1+m club

	Cluster Labels	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other
4	1	Manhattan	Chelsea	40.746491	-74.001528	1077600.0	41	0.097561	0.146341	0.756098
9	1	Manhattan	East Village	40.729269	-73.987361	1049500.0	40	0.125000	0.025000	0.850000
10	1	Manhattan	Financial District	40.707668	-74.009271	967500.0	47	0.106383	0.085106	0.808511
13	1	Manhattan	Gramercy Park	40.737939	-73.985922	1119375.0	47	0.127660	0.063830	0.808511
14	1	Manhattan	Greenwich Village	40.733584	-74.002817	1165000.0	45	0.044444	0.022222	0.933333
21	1	Manhattan	Lenox Hill	40.766437	-73.959017	1150000.0	47	0.234043	0.021277	0.744681
22	1	Manhattan	Lincoln Square	40.772319	-73.984401	1040000.0	46	0.108696	0.021739	0.869565
34	1	Manhattan	Theatre District - Times Square	40.763398	-73.983355	997500.0	47	0.063830	0.042553	0.893617
38	1	Manhattan	Two Bridges	40.711288	-73.992233	1225900.0	48	0.062500	0.645833	0.291667
39	1	Manhattan	Upper West Side	40.787045	-73.975416	1100000.0	47	0.085106	0.106383	0.808511
41	1	Manhattan	West Village	40.734186	-74.005580	1312500.0	45	0.022222	0.022222	0.955556

1.5m-2m club

	Cluster Labels	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other
1	3	Manhattan	Carnegie Hill	40.784197	-73.954339	1500000.0	47	0.127660	0.042553	0.829787
3	3	Manhattan	Central Park South	40.767814	-73.981232	2147500.0	46	0.065217	0.043478	0.891304
5	3	Manhattan	Chinatown	40.716491	-73.996250	1733179.0	44	0.068182	0.613636	0.318182
11	3	Manhattan	Flatiron District	40.741072	-73.989653	1865000.0	47	0.063830	0.063830	0.872340
12	3	Manhattan	Garment District	40.753694	-73.990517	1900000.0	42	0.095238	0.119048	0.785714
23	3	Manhattan	Little Italy	40.719273	-73.998215	1925000.0	41	0.073171	0.560976	0.365854
28	3	Manhattan	NoHo	40.725875	-73.993957	1935000.0	38	0.105263	0.157895	0.736842
30	3	Manhattan	NoMad	40.744688	-73.988285	1623500.0	43	0.116279	0.139535	0.744186

2.5m-3.5m club

	Cluster Labels	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other
6	0	Manhattan	Civic Center	40.713679	-74.002404	2760000.0	48	0.041667	0.583333	0.375000
16	0	Manhattan	Hudson Square	40.726804	-74.007983	2906012.0	47	0.085106	0.000000	0.914894
29	0	Manhattan	NoLiTa	40.722989	-73.994558	2647450.0	40	0.075000	0.375000	0.550000
32	0	Manhattan	SoHo	40.722880	-73.998750	3250000.0	40	0.050000	0.375000	0.575000
35	0	Manhattan	TriBeCa	40.715380	-74.009306	3550000.0	47	0.085106	0.319149	0.595745

5m club

	Cluster Labels	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other
17	2	Manhattan	Hudson Yards	40.755906	-74.000532	5504250.0	43	0.186047	0.186047	0.627907

Conclusion

Desired location should be a rich neighborhood, I focus on 2m+ club. I set 5% as a Chinese restaurant ratio threshold and get below result.

	Borough	Neighborhood	Latitude	Longitude	Meidian Sale Price	Venue	Alternative	Chinese Restaurant	Other
3	Manhattan	Central Park South	40.767814	-73.981232	2147500.0	46	0.065217	0.043478	0.891304
16	Manhattan	Hudson Square	40.726804	-74.007983	2906012.0	47	0.085106	0.000000	0.914894

In sum, **Hudson square** would be my recommended choice.

[Hudson Square, New York, NY Real Estate Market, Home Prices & Sales Trends \(propertyshark.com\)](https://www.propertyshark.com/)