Finding the best location in Manhattan, NY for an Italian Restaurant Supply business

An analysis using machine learning, Foursquare, python and open source tools

JOHN SECKMAN

APRIL 23, 2020

Introduction

Background

Everyone knows New York has the best pizza and Italian restaurants in the world, but where do the ingredients come from? We have all heard that it's the "hard" water of New York City that is the secret ingredient, but that is only fraction of the entire picture. The dough, noodles, spices, meat, sausage and other ingredients come from restaurant suppliers. Although each restaurant has their own special recipes and ambiance, many restaurants are supplied by the same supplier. With a single supplier making daily deliveries to multiple restaurants, it is important that the supplier be closely and centrally located to the restaurants it is serving.

Interest and Target Audience

If an Italian restaurant supplier were interested in locating to Manhattan, NY, they would be interested in finding a location that minimized delivery distances to group of close by customers. In this case, it would be Italian and pizza restaurants but the methodology could be applied to any supply business.

Problem

Find a cluster of potential customers and then find the central location from which to serve them.

Approach and Methodology

Two approaches are used.

The first is to find the Manhattan neighborhood with the most Italian or Pizza restaurants. This can be done by combining neighborhood information with restaurant location data and identifying the neighborhood with the most potential customers.

The second is to use location data to locate all potential customers in Manhattan and then to locate central locations for those clusters as potential site targets for a supplier to locate. This can be done by using K means analysis to identify customer clusters and then identify a central location that minimizes the delivery distance to each of those customers.

Data Sources

Venue type and location data – Foursquare, an online venue location and reviewer tool which has both free and upgraded accounts available.

Neighborhood data – An online file with names and geo coordinates for Manhattan neighborhoods namely https://geo.nyu.edu/catalog/nyu_2451_34572

Data Tools

Mapping and data visualization – Geopy libraries from Nominatim for translating address and geo locations and Folium an open source tool for mapping.

Source Code – Python

Source Code Storage – Github, an online source code repository that also provides branching and merging for code.

Data manipulation – Pandas and Numpy, both open source libraries for Python.

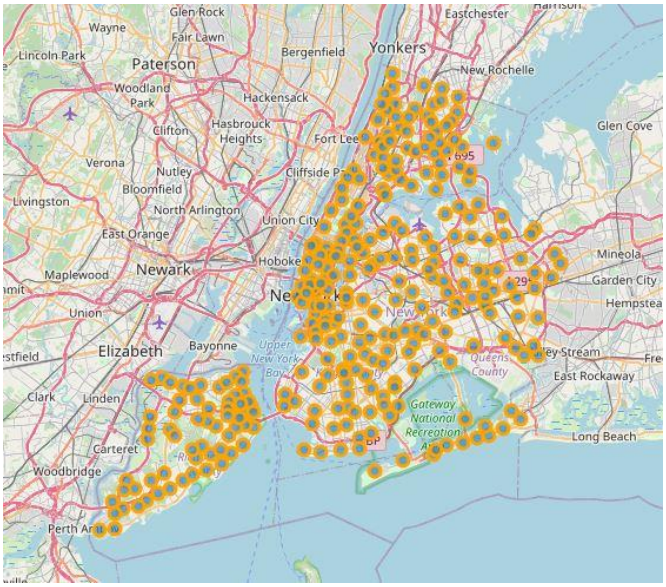
K means analysis – Sklearn, another open source library for Python.

Process for the First Approach

Neighborhood data was read from the online neighborhood file. It was in JSON format and read into a Pandas data frame. Below is an example:

	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

Below is a map of the neighborhoods:



Data was then limited to Manhattan and Foursquare data was added. The location data and venue data was then combined. Below is a sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

The venue data was then filtered to only include Italian and Pizza restaurants. Below is a sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category
0	Marble Hill	40.876551	-73.910660	Arturo's	40.874412	-73.910271	Pizza Place
30	Chinatown	40.715618	-73.994279	Scarr's Pizza	40.715335	-73.991649	Pizza Place
128	Washington Heights	40.851903	-73.936900	Saggio Restaurant	40.851423	-73.939761	Italian Restaurant
148	Washington Heights	40.851903	-73.936900	Fresco's Pizzeria	40.855202	-73.937216	Pizza Place
190	Washington Heights	40.851903	-73.936900	Exclusive Pizza	40.850989	-73.938635	Pizza Place

The data was then grouped and counted to see which neighborhoods had the most Italian and Pizza restaurants. Below is a sample:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category
Battery Park City	3	3	3	3	3	3
Carnegie Hill	6	6	6	6	6	6
Central Harlem	1	1	1	1	1	1
Chelsea	4	4	4	4	4	4
Chinatown	1	1	1	1	1	1
Civic Center	1	1	1	1	1	1
Clinton	5	5	5	5	5	5
East Harlem	2	2	2	2	2	2
East Village	8	8	8	8	8	8

Process for the Second Approach

The geo data for the Italian and Pizza restaurants was supplied as an input to K means analysis to generate clusters of customers. Six clusters were requested. Buyers investing in Manhattan real estate would want to have a number of alternatives. The clustering data was added to the geo data. Below is a sample:

	Cluster Labels	Venue Latitude	Venue Longitude
0	2	40.874412	-73.910271
30	4	40.715335	-73.991649
128	2	40.851423	-73.939761
148	2	40.855202	-73.937216
190	2	40.850989	-73.938635

This could then be mapped (see results section). The centroids were derived. Below is a sample of the data:

	0	1
0	40.724716	-74.003523
1	40.774666	-73.957426
2	40.861437	-73.927276
3	40.754432	-73.983737
4	40.728416	-73.985932
5	40.816091	-73.953878

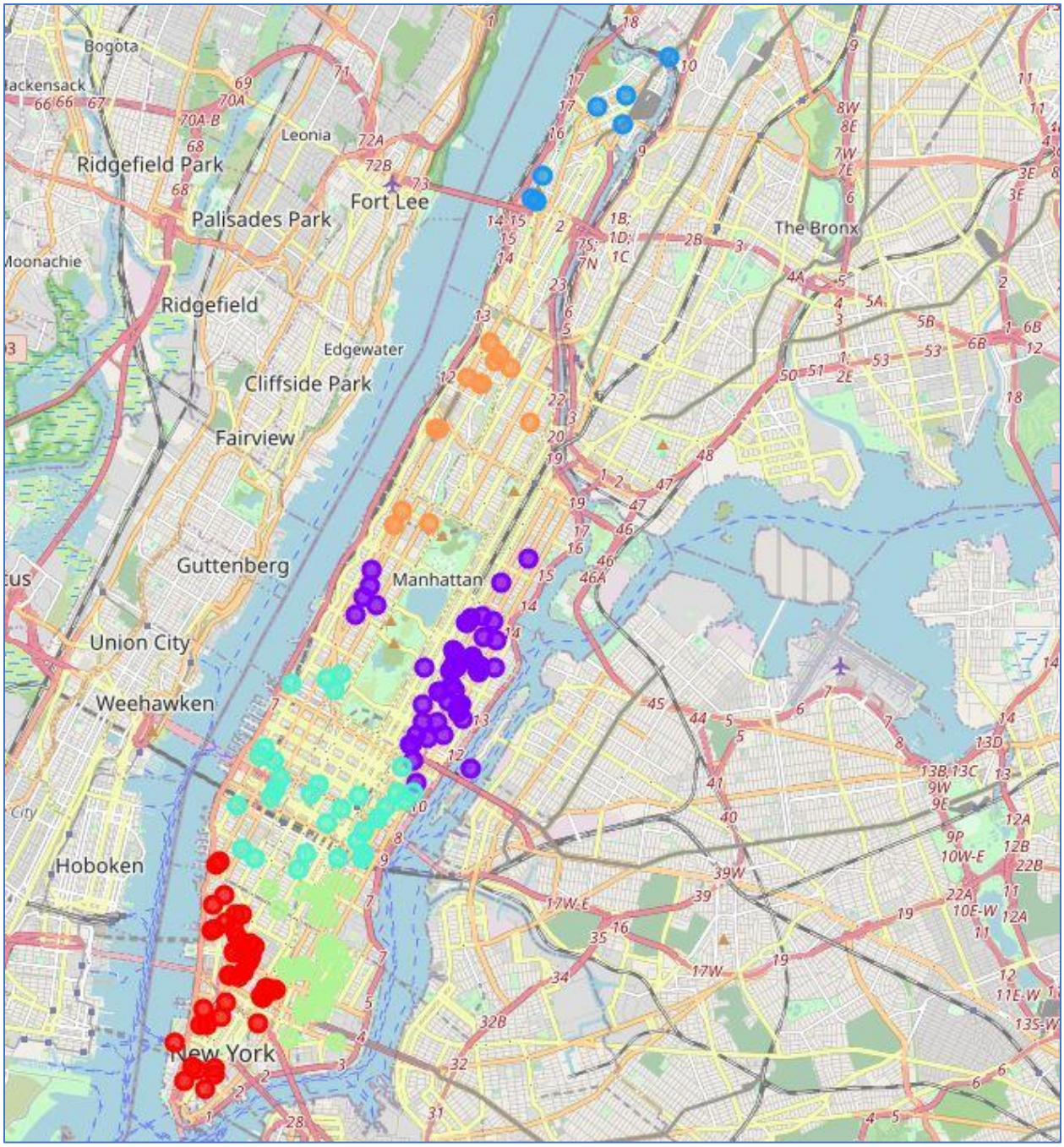
This allowed mapping of the centroids (see results section). This provided the most central location to serve each one of the identified clusters.

Results

First Approach: The top 6 neighborhoods with the most Italian or Pizza restaurants are as follows:

NoHo	14
Greenwich Village	12
Lenox Hill	12
Upper East Side	11
Yorkville	10
West Village	9

Second Approach: The map of customer clusters appears below:



The location of the centroids for these clusters are mapped below:

Discussion

There are many more factors that would go into a site analysis and acquisition. Cost / benefit might be the most pervasive given the expense of Manhattan real estate. Another consideration would be how the supplier is being supplied. Would it make business sense to be closer to the wharfs or transportation

arteries? While the centroids identify the mathematically optimized central location, they do not take into account transportation networks or real estate availability or suitability.

Conclusion

While the data and tools used for this analysis don't provide a final answer for the ultimate site location, they would help supply chain operators visualize and consider future logistics and would enable more informed discussions with real estate brokers.