Capstone Project

The Battle of the Neighborhoods (Week 2)

Recommending initial target neighborhood for product sales

Harsh Rahamatkar

October 05, 2019

1. Introduction

1.1 Background and Problem

Any company who wants to sell a product specific to particular business needs target business venues or locations to send their salesperson where they could find relevant businesses for best response.

For example, a company developed a new much efficient kind of coffee dispenser. Now it wants to contact or sell the product to businesses in optimal neighborhoods where most of the coffee shops are available and is sufficiently populated to reduce the transportation cost, increase the number of target shops and increasing the number of consumers directly affected.

Similarly, we can determine these initial target areas for the for other companies with new products.

1.2 Interest

There are many innovators and companies that relies on innovation who want to show or sell their limited products to specific business and influence higher number of people with lower investments on transportation costs.

2. Data Acquisition and Cleaning

2.1 Data Sources

Based on definition of our problem, factors that will influence our decision are:

- number of existing business of specific category in the neighborhood (high weightage)
- population of the neighborhoods (normal weightage)

Following data sources will be needed to extract/generate the required information:

- demographic data of New York from **Kaggle**
- number of business/venues and their type and location in every neighborhood will be obtained using **Foursquare API**

2.2 Data Cleaning and Feature Selection

Data downloaded from Kaggle contains New York's demographic data of 2000 and 2010. We will take the latest data. There are five boroughs in the table. We will take the latest data of Manhattan Borough i.e. 2010 data.

The table does not have the location coordinates of the neighborhoods. We used geopy library to get the latitude-longitude coordinates of the neighborhoods

For getting the venues and business details (name and coordinates) we used the Foursqure API on the neighborhoods available in the Manhattan data of year 2010. For now, we keep the radius as 500 units and limit the results to 100 records per neighborhood.

We now have population of each neighborhood and at most 100 venues in radius of 500 units in every neighborhood with their names, category and latitude-longitude coordinates.

Population Data

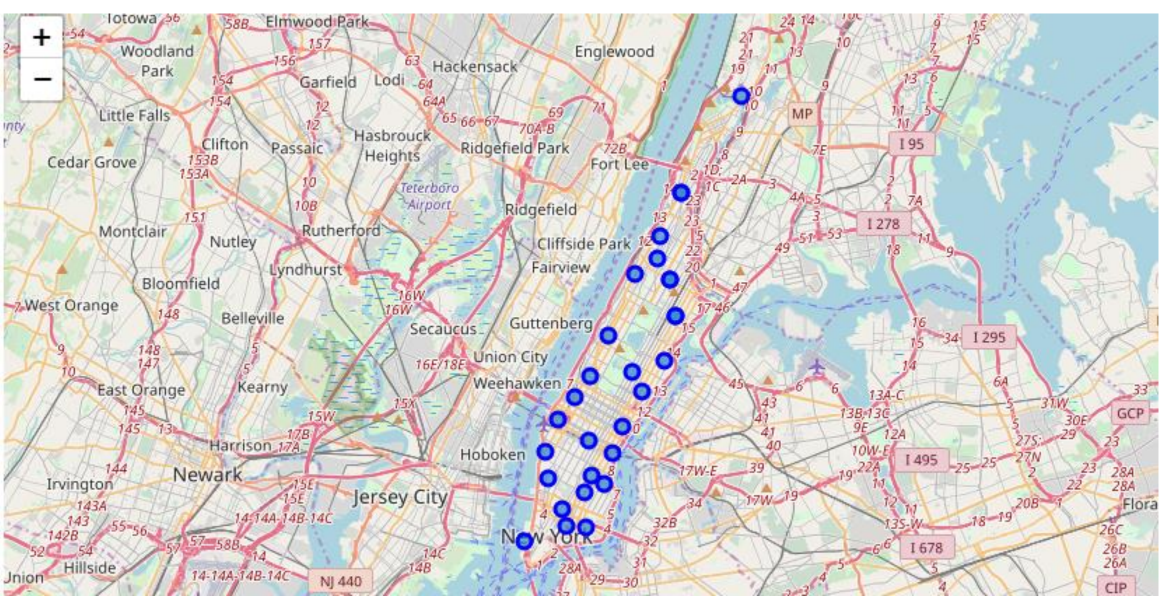
	Borough	NTA Name	Population	Latitude	Longitude
0	Manhattan	Marble Hill-Inwood	46746	40.876298	-73.910429
1	Manhattan	Central Harlem North-Polo Grounds	75282	40.807879	-73.945415
2	Manhattan	Hamilton Heights	48520	40.824145	-73.950062
3	Manhattan	Manhattanville	22950	40.815778	-73.951554
4	Manhattan	Morningside Heights	55929	40.810000	-73.962500

Venues Data

	NTA Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill-Inwood	40.876298	-73.910429	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill-Inwood	40.876298	-73.910429	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill-Inwood	40.876298	-73.910429	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill-Inwood	40.876298	-73.910429	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill-Inwood	40.876298	-73.910429	Dunkin'	40.877136	-73.906666	Donut Shop
...	...	...	...	...	...	...	...
2372	park-cemetery-etc-Manhattan	40.744249	-74.006425	Chelsea Piers Fitness	40.746645	-74.010057	Gym / Fitness Center
2373	park-cemetery-etc-Manhattan	40.744249	-74.006425	Top of The Standard	40.740818	-74.008116	Roof Deck
2374	park-cemetery-etc-Manhattan	40.744249	-74.006425	Apple West 14th Street	40.741270	-74.005389	Electronics Store
2375	park-cemetery-etc-Manhattan	40.744249	-74.006425	Sikkema Jenkins	40.747592	-74.005983	Art Gallery
2376	park-cemetery-etc-Manhattan	40.744249	-74.006425	PH-D at Dream Downtown	40.742347	-74.003356	Nightclub

2377 rows × 7 columns

Map of Manhattan New York



3. Methodology

We now must find out how many venues are available in the neighborhood from each of 288 categories. We will use one hot encoding for this purpose, and then group the resulting dataframe by sum method. The resulting dataframe will have the number of venues from each category in the neighborhood.

Let’s take the required business as coffee shop (for new efficient coffee dispenser machine), we will require to sell more coffee machine as well as a greater number of people should be influenced.

So first, we collect all the neighborhoods with required businesses and create a business score column for them. This score will prefer number of coffee shops over the population of the neighborhood. Thus, we will calculate it as:

- business score of the neighborhood = Population x (No. of coffee shops) ^2

So, generalizing this method, business score of neighborhoods for any category of business will be:

- business score of the neighborhood = Population x (No. of business of the category) ^2

We then select the top n number of neighborhoods sorted by business score from the list according to requirement n.

4. Analysis

Encoded list of neighborhoods with number of different venues in it.

	NTA Name	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Arepa Restaurant	Art Gallery	...	Video Game Store	Video Store	Viet Res
0	Battery Park City-Lower Manhattan	0	0	0	1	0	0	0	0	0	...	0	0	
1	Central Harlem North-Polo Grounds	0	0	3	1	1	0	0	0	1	...	1	0	
2	Central Harlem South	0	0	3	1	1	0	0	0	1	...	1	0	
3	Chinatown	0	0	0	0	0	1	0	0	0	...	0	0	
4	Clinton	0	0	0	2	0	0	0	0	0	...	0	0	
5	East Harlem North	0	0	0	0	0	0	0	0	0	...	0	0	
6	East Harlem South	0	0	0	0	0	0	0	0	0	...	0	0	
7	East Village	0	0	0	1	0	0	0	1	0	...	1	0	

Using the methodology, we can get recommended neighborhoods following for coffee shops.

```
recommendations=recommend_location('Coffee Shop').reset_index(drop=True)
recommendations
```

	Borough	NTA Name	Population	Latitude	Longitude	Coffee Shop	business_score
0	Manhattan	Lenox Hill-Roosevelt Island	80771	40.766437	-73.959017	7	3957779
1	Manhattan	Yorkville	77942	40.778007	-73.948202	5	1948550
2	Manhattan	Battery Park City-Lower Manhattan	39699	40.711017	-74.016937	7	1945251
3	Manhattan	Lower East Side	72957	40.715936	-73.986806	5	1823925
4	Manhattan	Upper East Side-Carnegie Hill	61207	40.773702	-73.964120	5	1530175

Using the venues data, we can list all the venues in recommended neighborhood

```
recommended_venues=manhattan_venues[manhattan_venues['NTA Name'].apply(lambda x: x in recommended_neighborhoods) * manhatta
recommended_venues
```

	NTA Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1430	Battery Park City-Lower Manhattan	40.711017	-74.016937	Le District Coffee	40.713284	-74.015854	Coffee Shop
1451	Battery Park City-Lower Manhattan	40.711017	-74.016937	Blue Bottle Coffee	40.710589	-74.012371	Coffee Shop
1459	Battery Park City-Lower Manhattan	40.711017	-74.016937	Starbucks Reserve	40.714170	-74.015434	Coffee Shop
1464	Battery Park City-Lower Manhattan	40.711017	-74.016937	Joe Coffee	40.712526	-74.013137	Coffee Shop
1493	Battery Park City-Lower Manhattan	40.711017	-74.016937	Boundless Plains Espresso	40.707990	-74.013894	Coffee Shop
1500	Battery Park City-Lower Manhattan	40.711017	-74.016937	Laughing Man Coffee & Tea	40.714754	-74.017368	Coffee Shop
1507	Battery Park City-Lower Manhattan	40.711017	-74.016937	Nobletree Coffee	40.710037	-74.011839	Coffee Shop
1636	Lower East Side	40.715936	-73.986806	Cafe Grumpy	40.715069	-73.989952	Coffee Shop
1639	Lower East Side	40.715936	-73.986806	Little Canal	40.714317	-73.990361	Coffee Shop
1649	Lower East Side	40.715936	-73.986806	Blue Bottle Coffee	40.719140	-73.985224	Coffee Shop
1667	Lower East Side	40.715936	-73.986806	GrandLo Cafe	40.716885	-73.985680	Coffee Shop
1713	Lower East Side	40.715936	-73.986806	Caffe Vita Coffee Roasting Co.	40.719752	-73.988529	Coffee Shop
1741	Lenox Hill-Roosevelt Island	40.766437	-73.959017	The Coffee Inn	40.766767	-73.956515	Coffee Shop

Using folium library to visualize the venues in the map of Manhattan

5. Results and Discussion

So, for the new coffee dispenser machine manufactured by the company, to have large influence on the basis of both number of coffee shops in and the population of the neighborhood, we came to conclusion that Lenox Hill-Roosevelt Island, Yorkville, Battery Park City-Lower Manhattan, Lower East Side and Upper East Side-Carnegie Hill will be the best five.

We can use the same for other categories of businesses like gym, video store, restaurants etc. to get the target locations for initial sales of the good to have wide influence and least transportation costs.

6. Conclusion

We were required to get the recommendations for initial target neighborhoods for sending the new product manufactured/created by any company for most sales and influence. Following the given methodology of weighted products with population and number of businesses/venues in neighborhood, we successfully got recommendations for a new coffee dispenser machine.