**Capstone Project**

**The Battle of the Neighborhoods (Week 2)**

Recommending initial target neighborhood for product sales

Harsh Rahamatkar

October 05, 2019

1. **Introduction**
   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. **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.

1. **Data Acquisition and Cleaning**
   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**
  1. **Data Cleaning and Feature Selection**

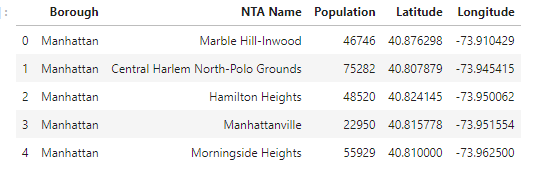
Data downloaded from Kaggle contains New York’s demographic data of 2000 and 2010. We will take the latest data. There are five boroughs in the table. We will take the latest data of Manhattan Borough i.e. 2010 data.

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

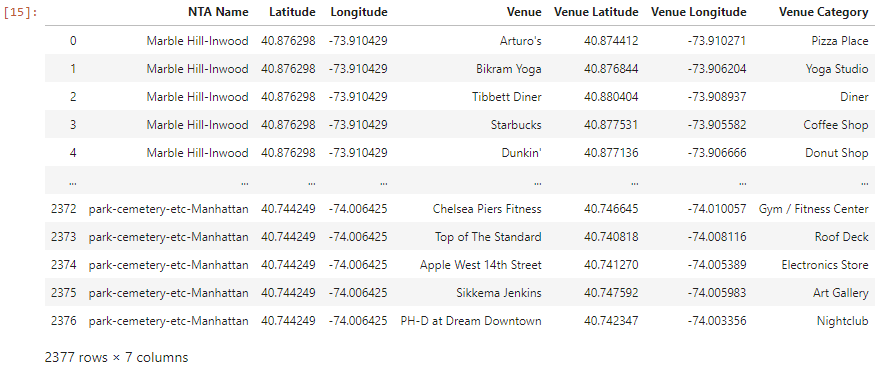
For getting the venues and business details (name and coordinates) we used the Foursqure API on the neighborhoods available in the Manhattan data of year 2010. For now, we keep the radius as 500 units and limit the results to 100 records per neighborhood.

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.

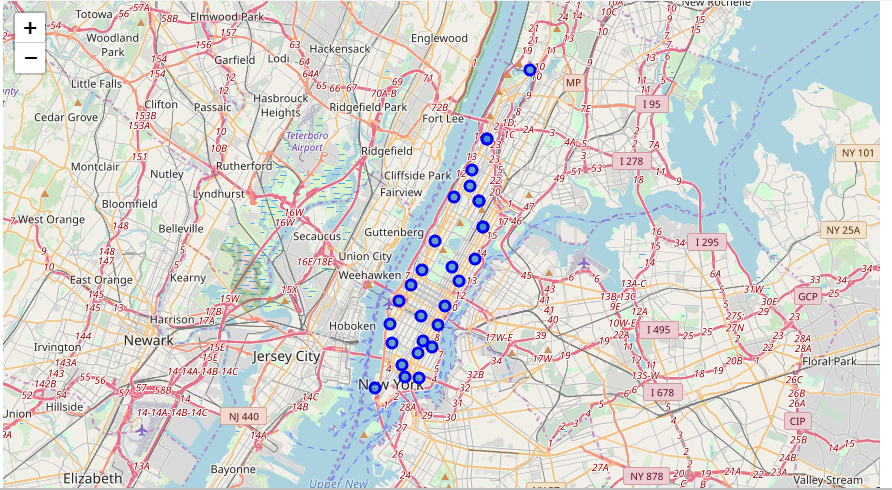
*Population Data*



*Venues Data*



*Map of Manhattan New York*



1. **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

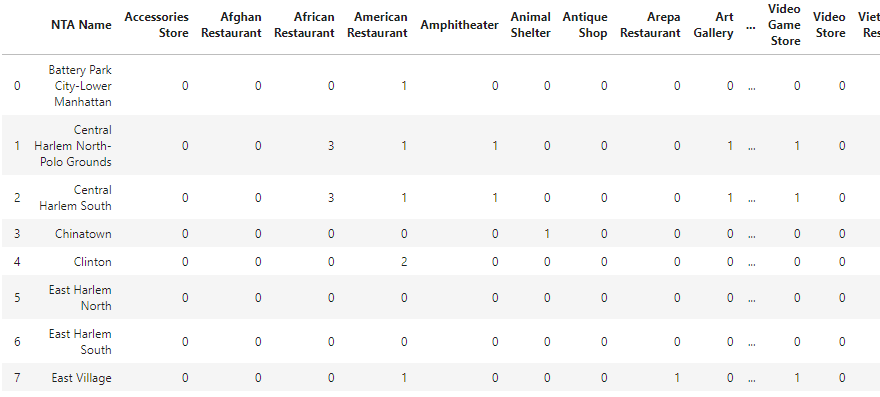
So, generalizing this method, business score of neighborhoods for any category of business will be:

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

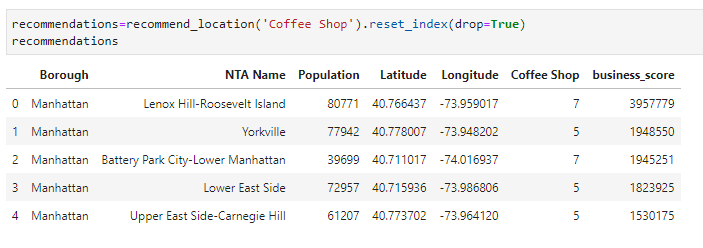
We then select the top n number of neighborhoods sorted by business score from the list according to requirement n.

1. **Analysis**

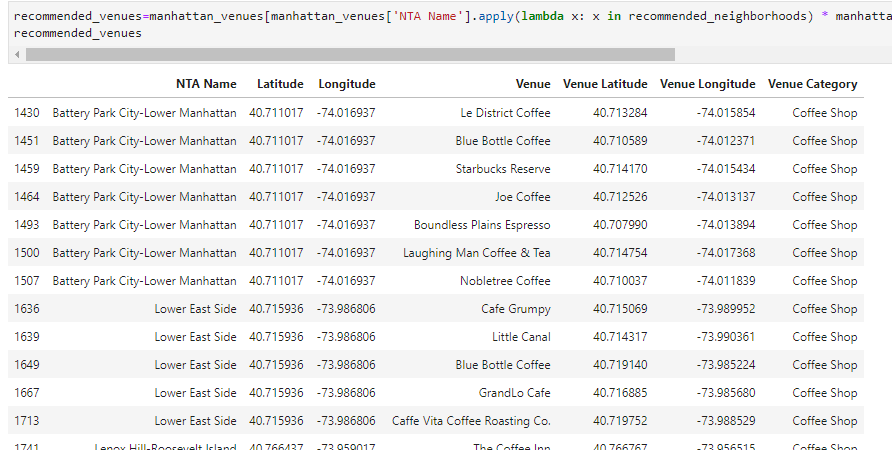
Encoded list of neighborhoods with number of different venues in it.



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



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



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



1. **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.

1. **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.