**CAPSTONE PROJECT- BATTLE OF THE**

**NEIGHBORHOODS**

**City comparison and Restaurant Venue Recommendation**

**Purpose:**

This document provides the details of my final peer reviewed assignment for the IBM

Data Science Professional Certificate program.

**INTRODUCTION:**

New York City is the biggest city in United States and Toronto is the capital city of the province of Ontario in Canada. Both cities have some similarities and also differences. Both cities are quite safe. Both are the economic, cultural, and fashion capitals of their countries.

New York City's architecture is far more impressive.  Toronto's a lovely city, but its streetscapes are made beautiful by the people who occupy it--the cafes and such.  The architecture itself leaves much to be desired. New York is brilliantly planned. If anyone in New York for a three-day weekend, get a hotel in Midtown Manhattan and nearly everything able to see and visit within walking distance. Toronto is much more sprawled out and need to use the subway lots. (3 of TO's biggest attractions: the CN Tower, the Royal Ontario Museum, the Ontario Science Centre and Casa Loma, are nowhere near each other).

New York has AMAZING pizza and the best Italian restaurants in North America; Toronto has quite possibly the worst pizza on the planet. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Italian, Chinese, Indian, French etc.

So, as part of this project, I will list and visualize all recommended place of New York City where someone can open restaurant and perfect location to setup the office.

The sample recommender in this notebook will provide the following use case scenario:

* Populated area like shopping complex, school, college, office is placed, there will restaurant suitable.
* Area wise target customer.
* Customer wise food variant.
* Based on peoples earning define food cost.

**DATA ACQUISITION:**

New York data set will retrieve from this link <https://cocl.us/new_york_dataset> . Then will explore Neighborhoods in New York City and analyze each neighborhood.

I will be using the FourSquare API (FourSquare website: www.foursquare.com) to explore neighborhoods in selected area. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

* Venue ID
* Venue Name
* Coordinates: Latitude and Longitude

Another venue query will be performed to retrieve venue ratings (based on selected parameters) for each location.

I will use GeoSpace data from <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm> to get New York Borough boundaries that will help us visualize choropleth map.

**Approach:**

* Collect the New York City data from <https://cocl.us/new_york_dataset>
* Using FourSquare API, we will find all venues for each neighborhood.
* Find out all venues for restaurants.
* Find rating, tips and like count and preferable place for restaurants using FourSquare API.
* Using rating for each restaurant, we will sort that data.
* Visualize the Ranking of neighborhoods using folium library(python)

**Questions based on datasets**

* What is best location in New York City for restaurant business?
* Which areas have potential Restaurant market?
* Which all areas lack of restaurants?
* Which is the best place to setup the office?

**METHODOLOGY**

**Restaurant Venue Recommendation:**

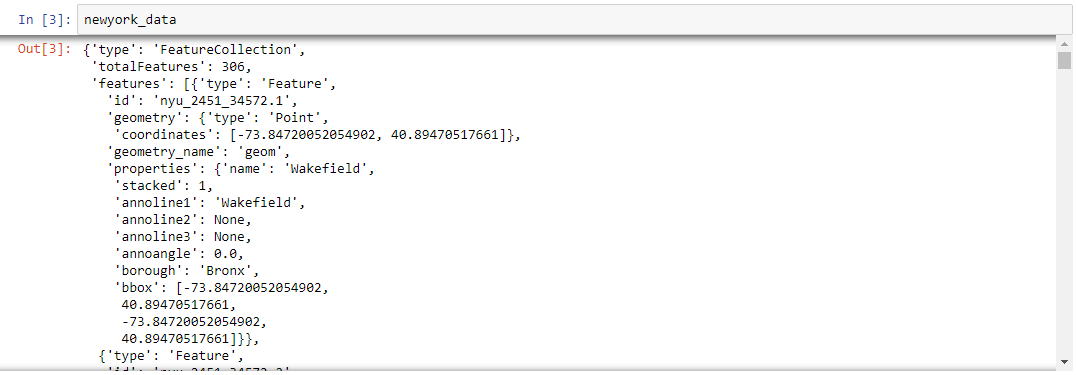
The source data contains recommended neighborhood in New York City. I will retrieve the most recommended neighborhood from New York City dataset with some parameter like populated area. For this demonstration, I will simplify the analysis by using conditional function.

**Data download and convert:**

Data collect from  <https://geo.nyu.edu/catalog/nyu_2451_34572>. I downloaded the files and open as jason format. The retrieved dataset contains jason data. I will convert it in pandas dataframe.

By the following code I just explore the dataset. Then Define the data with the ‘features’.





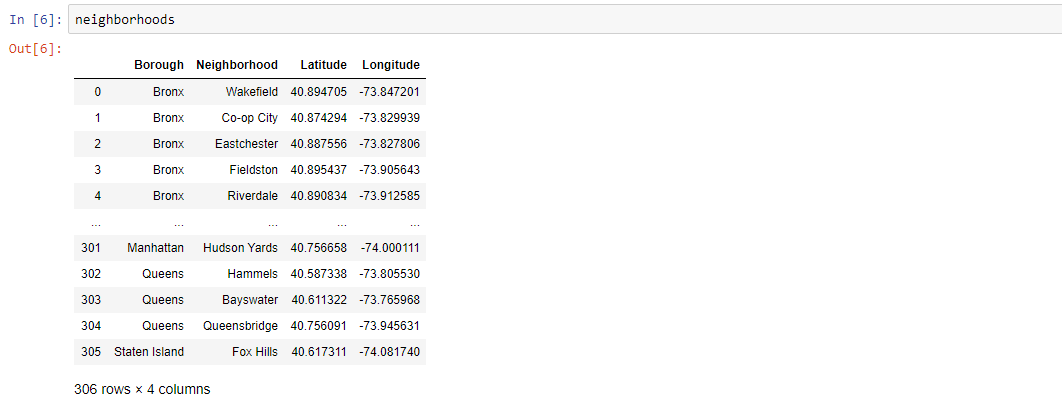


**Convert data in ‘Pandas Dataframe’:**

The next task is essentially transforming this data of nested Python dictionaries into a *pandas* dataframe. So, let's start by creating an empty dataframe. At first define columns then transform data into dataframe.

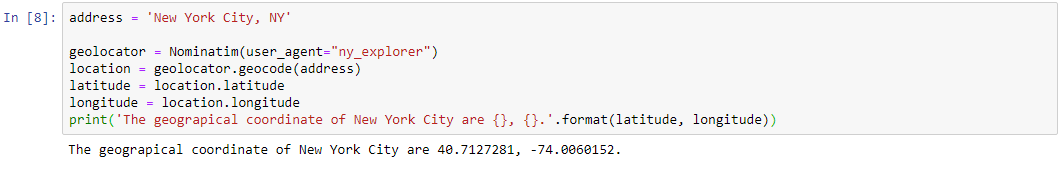


The resultant dataframe is,



**Use geopy library to get the latitude and longitude values of New York City:**

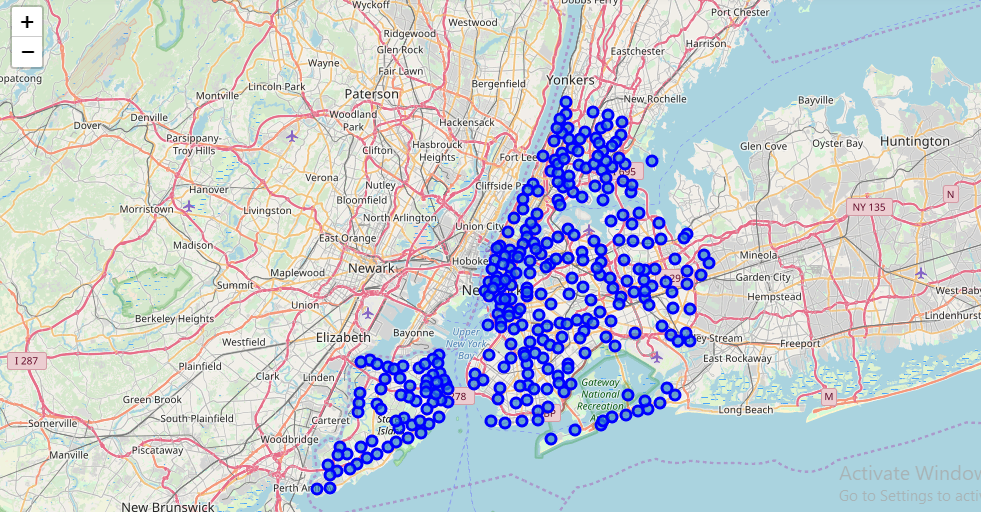
In order to define an instance of the geocoder, have to define user\_agent. I will name this agent ny\_explorer, as shown below.

****

Create a map of New York with neighborhoods superimposed on top.

****

The output of the code

****

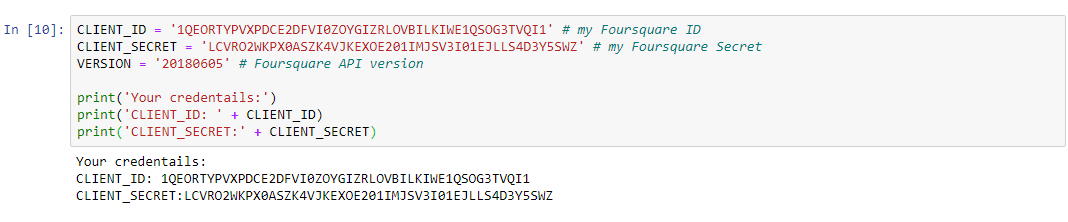
**Retrieving FourSquare Places of interest:**

Using the Foursquare API, the **explore** API function was be used to get the most

common venue categories in each neighborhood, and then used this feature to group

the neighborhoods into clusters. Here by this API I will get the information about New York neighborhoods then get preferable places for restaurant business. To get data from Foursquare process will describe below.

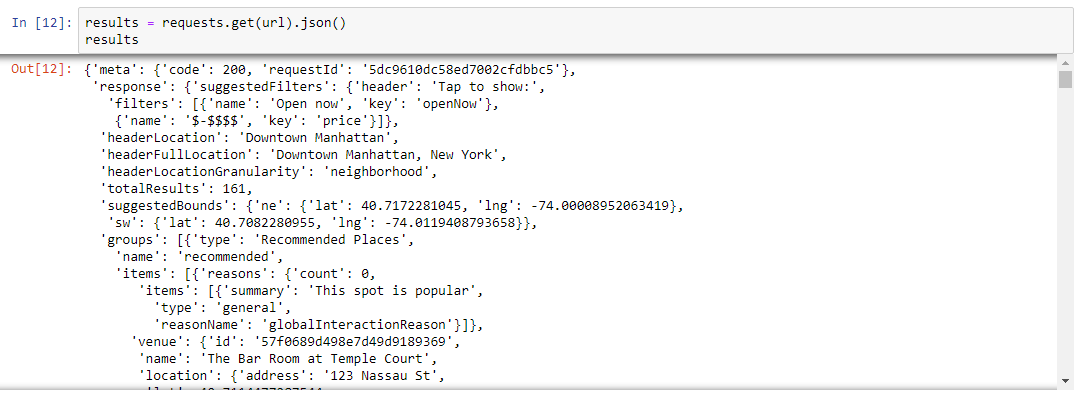
Define Foursquare Credentials and Version:



Let’s explore top 100 venue within a radius of 500 meters:



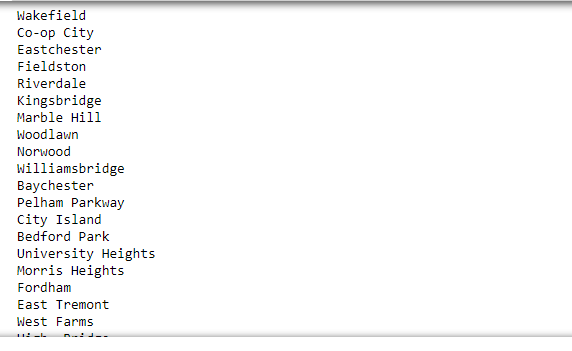
Send the GET request and examine the results:



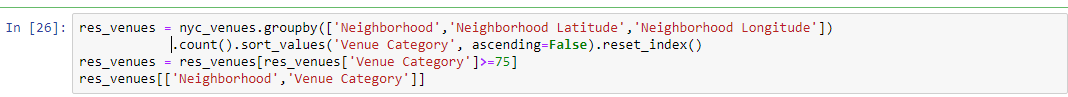
Here is main work. I will define venue of crowded area without restaurant and create dataframe. By this I will get listed venue where suitable for restaurant.

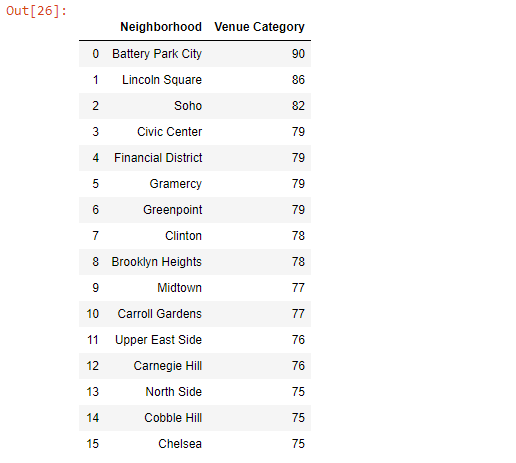


Sample of result:



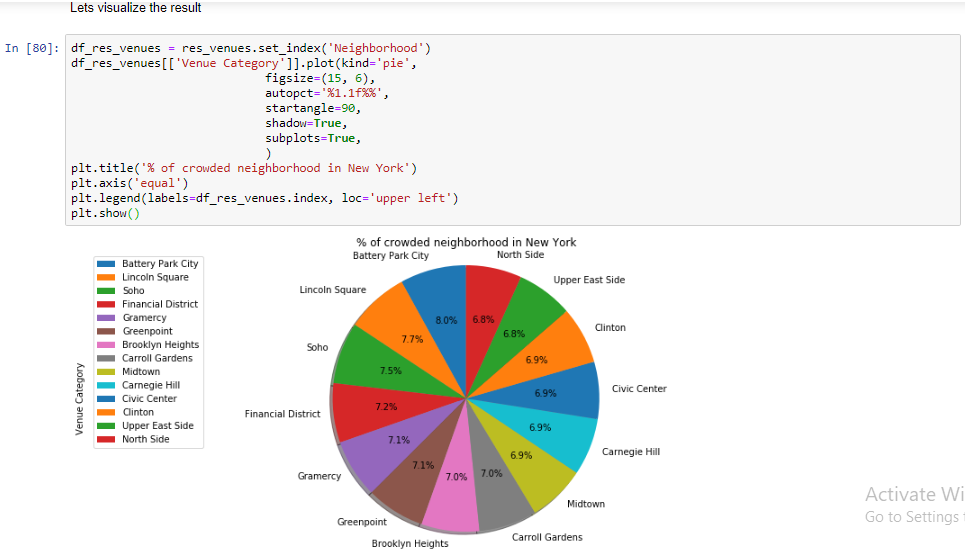
Here showing resulted neighborhood where restaurant preferable. I select preferable place which are most surrounding with venue and define this whose listed value bigger then 75.

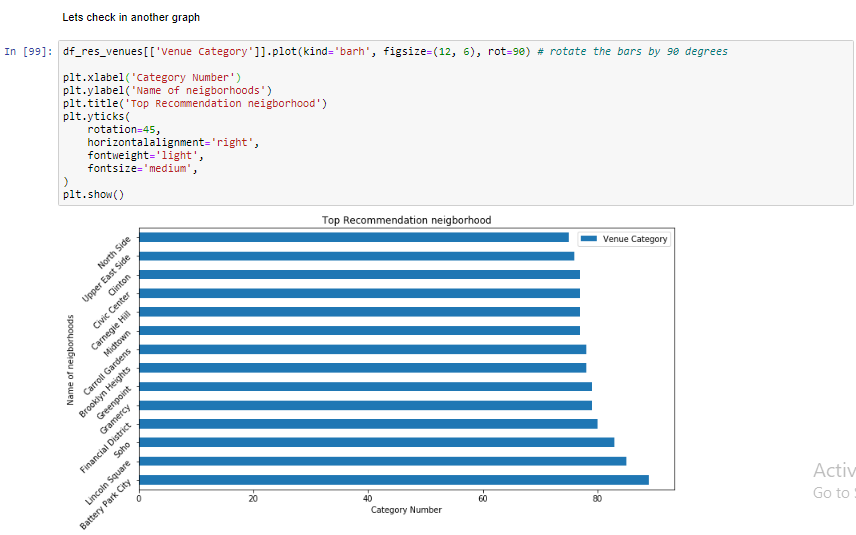




**Let’s show the result with the visualization:**

Here I am showing 2 type graphical view for the result. One is pie and another one is bar chart. Its help us to get decision that which places is suitable for the restaurant business.





**Discussion and Conclusion**

On this notebook, Analysis of best town venue recommendations based on restaurant venue category has been presented. Recommendations based on people go there mostly. As New York city is a big city with a whole host of interesting venues scattered around the city, the information extracted in this notebook present on the city areas, will be a good supplement to web based recommendations for visitors to find out nearby venues of interest and be a useful aid in deciding a place to stay or where to go during their visits. Using Foursquare API, we have collected a good amount of venue recommendations in New York City. Sourcing from the venue recommendations from FourSquare has its limitation; The list of venues is not exhaustive list of all the available venues is the area. Furthermore, not all the venues found in the area has crowded. For this reason, the number of analyzed venues has bigger than 75 number categories are available venues initially collected. The results therefore may significantly change, when more information is collected on those with missing data. The generated neighborhood from our results very good and interesting places located in areas. This kind of results may be very interesting restaurant business. Our results also yielded some interesting findings. For instance, the initial assumption among websites providing recommendations is that the Central Area. I will be providing another supplementary Inferential Statics in the future about on these data collected and also update in a new notebook using other categories. For now, this completes the requirements for this task.

Thank you.

Shah Alam Sumon

email: sacsesumon@gmail.com

Created For: COURSERA **IBM Applied Data Science Capstone Project**