**The best location to open an Italian Restaurant in New York City**

**Introduction**

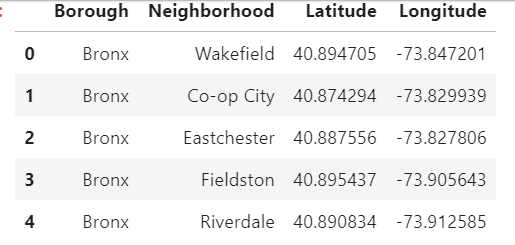
The business problem I would like to discuss in this report is the following one: Which is the best place to open an Italian Restaurant in New York City (NYC)?

Clearly the answer to the above question is not trivial and it should be the result of a careful evaluation that the business owner should make.

The problem I introduced may be of interest for anybody who wants to start the above business activity and would like to select the best location to start it.

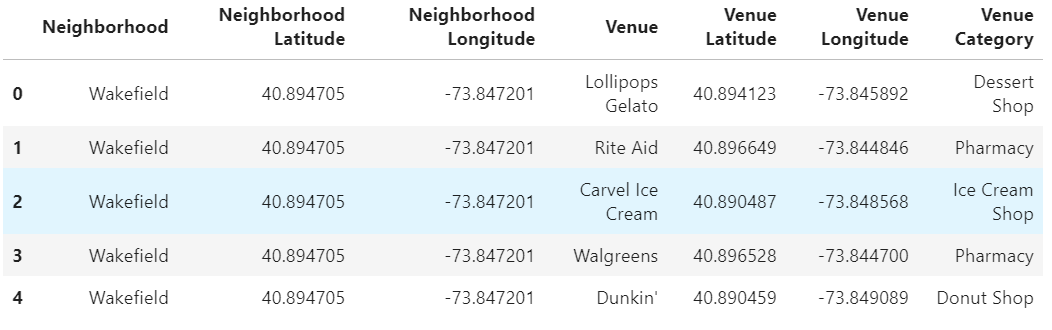
**Data**

The data I am going to use for this report are the geographical data of NYC, in particular the data set used during the Course laboratories and contained in the file “newyork\_data.json”. NYC data set contains a list of neighborhoods which include Boroughs, Neighborhoods along with a position described by geographical coordinates (i.e. Latitude and Longitude). These data can be simply transformed into a pandas dataframe, as shown in the figure below.



**Figure 1: Neighborhhoods of NYC**

Through these data, using the Foursquare API, it’s possible to explore neighborhoods and get a picture of the venues for each neighborhood. This allows to make an analysis which is propaedeutic to the evaluation of the best place to start the business activity. As shown in the figure below, for each neighborhhood a number of venues is found thanks to Foursquare along with its position (latitude and longitude) and category.



**Figure 2: Neighborhhoods of NYC with their venues.**

**Methodology**

The methodology used for this project is based on the following elements:

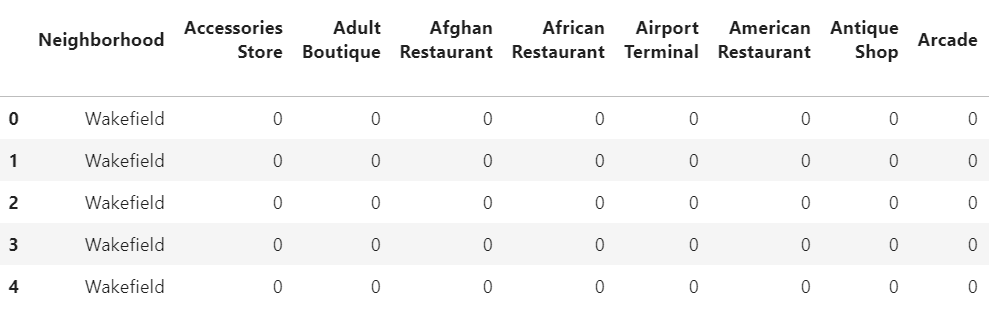
1. Preparation and analysis of data set
2. Use of Foursquare API to explore neighborhoods in the City
3. Analysis of each Neighborhood through Pandas
4. Definition of the Features for the choice of the best location to open an Italian Restaurant in NYC.
5. Selection of the best location

For what concerns point 1, as mentioned earlier, the reference data set is the one used for the course laboratory, named “newyork\_data.json”, and concerns neighborhoods of NYC. It can be download and conveniently transformed into a pandas dataframe, looking like the one in Figure 1.

As it can be seen, each row contains a neighborhood of NYC with the following information: Borough, Neighborough, Latitude and Longitude.

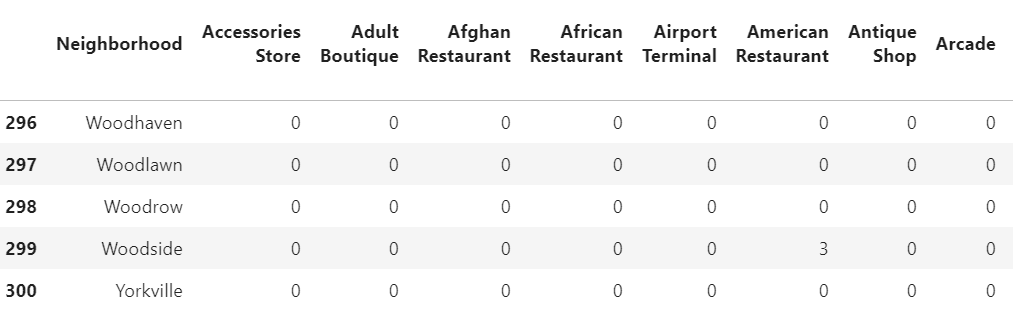
Starting from this set, it’s then possible to go to point 2, i.e. use of Foursquare API to explore neighborhoods in the City. This means that, for each neighborhood, a number of venues is found within a certain radius. Both the number of venues and the radius can be set as parameters. For this project, the number of venues and radius have been set to 100 and 500 m respectively. The resulting dataframe looks like the one shown in Figure 2.

Data frame is then one-hot encoded to have a shape like the following one:



**Figure 3: Neighborhhoods of NYC with their venues – one hot encoded**

After that, all data are grouped for Neighborhood, therefore the resulting dataframe will have each row representing a neighborhood and the columns the venue category. The data are the number of venues falling into that category (see figure below).



**Figure 4: Neighborhhoods of NYC with their venues – grouped by Neighborhoods and summed for each category**

This dataframe will be finally used for the selection of the best location to open an Italian Restaurant, according to the following procedure:

First of all, the neighborhoods list must be filtered according to some criteria which have been identified as follows:

* No Italian Restaurant already existing
* There must be at least one bus stop or one metro station (in order to ease people that want to reach the restaurant and don’t have or don’t want to use their own car)
* There must be at least one entertainment venue belonging to the following categories:
  + Theater
  + Comedy Club
  + Art Gallery
  + Art Museum
  + Museum
  + Music Venue
  + Jazz Club
  + Concert Hall

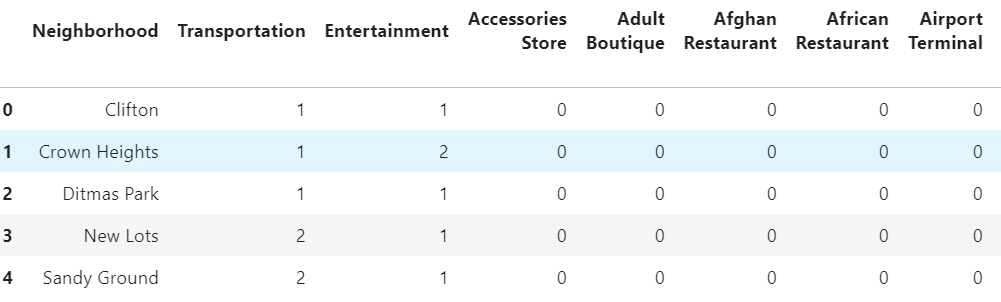
This requirement is important because lunch or dinner a tour restaurant may be encouraged by the presence of the above venues such that a person can combine lunch or dinner with other entertainment activities.

After filtering out the neighborhoods according to the above requirements, in order to make the final choice of the location where it’s better to open an Italian Restaurant, two further columns (“Transportation” and “Entertainment”) have been added to the dataframe and put at the beginning of it. “Transportation” column contains the sum of bus stop and metro station whereas “Entertainment” contains the sum of entertainment venues (i.e. Theater, Comedy Club, Art Gallery, Art Museum, Museum, Music Venue, Jazz Club, Concert Hall). The criterion used to select the optimal neighborhood is the presence of transportation facilities, as they have been considered with higher weight than the Entertainment venues.

Results are discussed in the next section.

**Results**

After filtering the neighborhood data according to the criteria defined in the previous section, the resulting data frame is the one shown in the next figure:



**Figure 5: Neighborhhoods of NYC satisfying the requirements**

According to the figure, only five candidates still remain. However, two neighborhoods (New Lots and Sandy Ground) have more transportation means than the remaining ones. In particular, New Lots has both a bus stop and a metrostation, that’s why it has been considered more suited than the other candidate neighborhoods.

**Discussion**

The methodology discussed in this report aimed at finding the best location in NYC where to open an Italian Restaurant. Starting from the venues of each neighborhood, some criteria have been defined to select the candidate neighborhood. These criteria represented the requirements each neighborhood had to satisfy in order to be considered as a candidate for the final choice.

Then, the final selection has been done on a further criterion introduced to select the preferred neighborhood. The used methodology can obviously lead to different results in case different criteria are used. In fact, it has been set as a constrained optimization problem where the constraints were the requirements to be satisfied (i.e. No Italian Restaurant already existing, at least one transportation mean to reach the neighborhhod and at least one entertainment venue in the neighborhhood) while the objective function was the feature considered to select the optimum neighborhood, i.e. the total number of transportation means and their types (the presence of bus and metro was preferred over only bus or only metro).

Different criteria could in any case be defined and, as a result, different locations could be considered the best one to open an Italian Restaurant in NYC.

**Conclusions**

In this report, an analysis based on location data has been carried out. The goal of the analysis was the identification of the best location in NYC where to open an Italian Restaurant. The problem has been stated and the source of data has been described. After that, the methodology used for the analysis has been explained in detail and all the steps of the used procedure have been highlighted. Finally, results have been presented along with some remarks.

Starting from a data set containing the Neighborhood of NYC, the source of data for the analysis came from Foursquare API, which allowed to get a given number of venues within a predefined radius for each neighborhood. Venues and their categories have been used to define some criteria needed to select the neighborhood candidates for the final choice and, subsequently, to select the most suited neighborhood to open an Italian Restaurant on NYC.