**Richard Kornblith**

[**Rkornblith@me.com**](mailto:Rkornblith@me.com)

**IBM Data Science Professional Sequence**

**Capstone Project [Week 1]**

**Battle of the Neighborhoods**

**February 24, 2019**

**Project: Finding Characteristics among New York Neighborhoods Experiencing Rapid Development of “Hot” Restaurants**

Introduction/Business Problem/Working Hypothesis

At the request of my client, an investment group wishing to establish a neighborhood restaurant in an affluent, dynamic neighborhood of Manhattan, I will provide advice and insight on the question of location. The client has lined up an executive chef who is well-known in New York and is somewhat flexible in cuisine so long as it is exciting and *nouvelle*; he anticipates that, if placed in the right neighborhood, the restaurant will attract a significant base of young (25-40 years old) finance, legal, academic and other professionals living in the immediate vicinity. The price point of his anticipated menu will be in the mid- to upper-range for the location, full liquor service is essential and, at least initially, only dinner service will be provided. The client, in principle, is in full agreement with the chef.

The client is very well funded and is not concerned with rental rates or initial working capital requirements, provided that the resulting customer base at the chosen location reasonably could be expected to produce revenues sufficient to support the client’s business plan. Accordingly, the client has specified that consideration of the availability of specific properties, and the cost of rent and maintenance of those properties, and other local considerations be deferred until a later date.

Upon a brief review of business literature to gain knowledge of the domain, I understand that the decision on the neighborhood in which to locate a new restaurant is dependent upon a number of factors: the demographics of the neighborhood, the proximity of other new and successful restaurants, in an urban area such as Manhattan, the proximity of transportation and so on. Based upon that review, I have reached the following working hypothesis, with which the client is in agreement.:

The likelihood of success of a “neighborhood restaurant” (that is, of a restaurant the expected success of which is premised on the assumption that a major portion of its clientele will reside in the surrounding neighborhood) turns primarily upon the location of that restaurant. The likelihood of development of a ‘hot area’ for restaurants, particularly in a neighborhood of New York attractive to the target clientele, in turn is a function of a precedent rapid residential development of that neighborhood. Neighborhood restaurants in such ‘hot areas’ find rapid success and mature, typically, during a roughly five-year period that commences as the development of the neighborhood reaches a critical point sufficient to support those restaurants. Following that period, most restaurants (and the neighborhoods in which they are located) have matured: the residential population has matured, married, perhaps started families and either move on or find that other activities displace eating out at hot restaurants; the restaurants also reach a similar stage of sociocultural maturity and are change to attract the maturing neighborhood residents, are sold or go out of business: rarely such a restaurant does continue to experience profitability, by, e.g., drawing in a stable and loyal clientele in substantial portion from without the neighborhood. My client investor group is focused on achieving rapid success and high profit during the initial period and is heavily discounting the potential long-term value of the enterprise.

Available Data Sources

With that working hypothesis, I have reviewed various sources of data as resources for my analysis of features that might have been characteristic of neighborhoods that currently are or recently have supported the development of restaurant areas of the type envisioned. There are many possible features that could be considered: however, the client has demanded a report based upon only two weeks’ work. Limited by this to only a preliminary exploration of available data, I have determined to perform a cluster analysis of neighborhoods in mid-town and down-town Manhattan.

In an initial exploration of available data, I have uncovered a number of resources that provide potentially relevant location-based information: fortuitously, the bulk of this data is indexed not only by geographical coordinate but also by United States Census Bureau census tract and, in many cases, census block-group label. With its dense population, Manhattan conveniently comprises a correspondingly dense distribution of census tracts, most comprising four to five block groups that typically include no more than two to four city blocks. For convenience, I will refer to each such geographical area as a “target area”.

The datasets to be used for this project are included as datafiles in the GitHub repository, to which the reader is directed. These may be supplemented, amplified or discarded as the project progresses. From these data, I will be extracting information characterizing each target area based upon what I believe are the most relevant factors addressed by that data:

1. *Local density of popular restaurants similar to the one envisioned by the client*.

For this, I have drawn upon three sources:

* 1. the Foursquare database of restaurant venues by location, menu price categorization and favorable ‘tips’. The Foursquare data will be generated on an as needed basis through calls to the venues API’s including, particularly, venues/EXPLORE (as was done in prior exercises) and venues/VENUE\_ID. Searches in the VENUE\_ID databases are premium searches and are subject to a daily Foursquare limitation: for this early exploration, I am working around that limitation, but it nevertheless may lead to a contraction of the number of target areas that are explored;

* 1. the open database of the New York State Liquor Authority (the “NYSLA”) cataloguing recently issued liquor licenses. These data are found in the file ‘*Excerpted\_NYSLA\_Quarterly\_List\_of\_Active\_Licenses*’; and,
  2. the open databases published by the City of New York (“NYC”) at ‘NYC.com/data’ (the ‘NYC Open Data Project’), more specifically the records of the New York City Department of Consumer Affairs cataloguing the operating licenses of local restaurants. These data are found in the file ‘*NYC\_DOB\_Manhattan\_2016-present*’;

1. *Target Area construction permits for new or converted housing*. For this, I have drawn upon the NYC Open Data Project, more specifically the records of the City’s Department of Buildings that catalogue project approval and permit applications by location and type of construction project, number of involved residential units and so on. These data are found in the file ‘*NYC\_DOB\_Manhattan\_2016-present*’;
2. *Residential rental rates*. For this, in order to get a sense of rental rate trends, I have drawn upon the open data published by the United States Census Bureau (the “USCB”) in its five-year American Community Surveys (2010-2014), (2011-2015) and (2013-2017). The USCB provides substantial discussion and enumeration of the error ranges for this data and also cautions against relying too heavily five-year studies that have substantial overlapping years. Nevertheless, these are the only free and open sources of data that I found that might reveal patterns of rental changes on a target area basis and so for purposes of this exploratory study I have considered them. These data are found in the subdirectories ‘*ACS2014*’, ‘*ACS2015*’ and ‘*ACS2017*’; and
3. *Demographics of local residents.* For this, I again have drawn upon the American Community Surveys (2010-2014) and (2013-2017) [subdirectories ‘*ACS2014’* and ‘*ACS2017’*], including particularly tabulations of household income, householder level of education, size of household, and tenure of the householder in the community. I also have drawn upon the NYC Open Data Project tabulation of dog licenses by census tract (2016-2017). That latter data are found in the file ‘*NYC\_Dog\_Licensing\_Dataset*’.

Supporting the analysis, I also have drawn upon several mapping services, including the Google developer API’s for geocoding and venue searches, and folio for map visualization. The value and feasibility of use of additional or different geocoding and venue services and visualization techniques may emerge from the exploration of the data described above.