Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Opening a new Indian Restaurant in Manhattan, New York

By

Divya Singh

Table of contents

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Indian restaurant** in **Manhattan**, New York. Since there are lots of restaurants in **Manhattan** we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Indian restaurants in vicinity**. We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, the factor that will influence our decision is the number of existing Indian restaurants in the neighborhood. We will select the Cluster of neighborhood which has the lowest density of Indian restaurants.

To Analyze the neighborhoods of Manhattan, we will have to first analyze the New York City data.

New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists on the web at
https://geo.nyu.edu/catalog/nyu_2451_34572

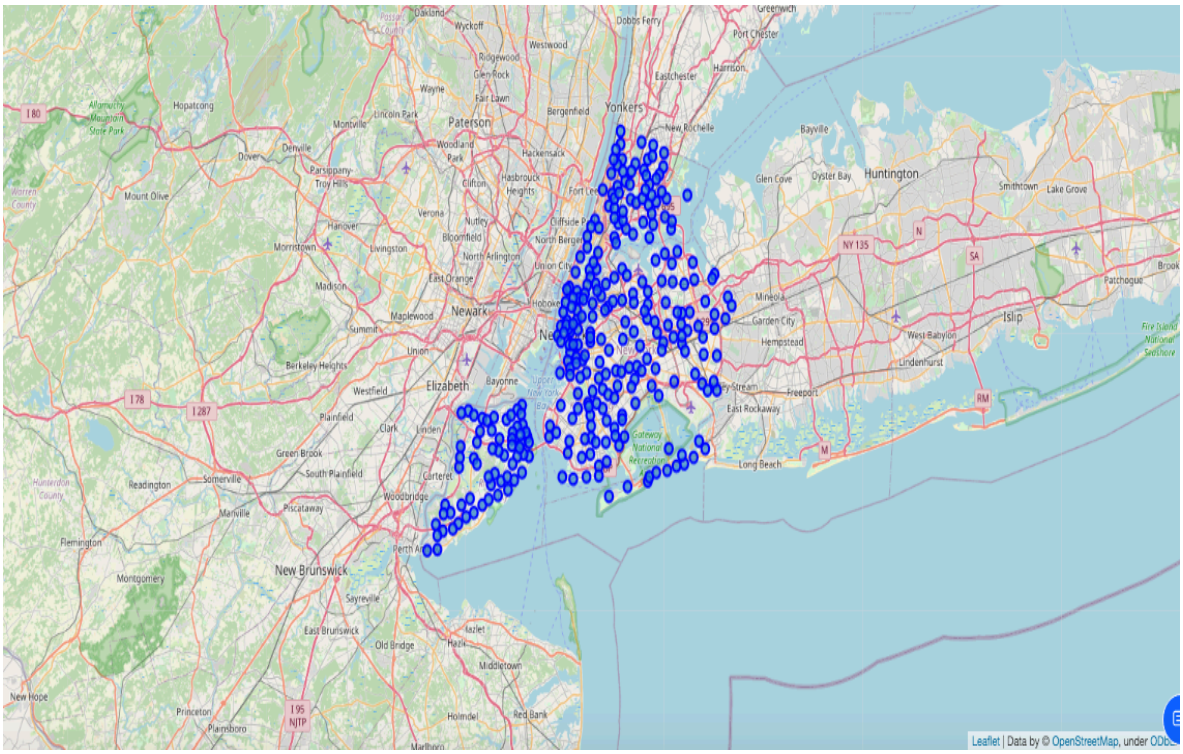
Following data sources will be needed to extract/generate the required information:

- Number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- Coordinate of New York City/Manhattan neighborhoods will be obtained using **geocoder**

Methodology

In this project we will direct our efforts on detecting areas of Manhattan that have low density of Indian restaurants.

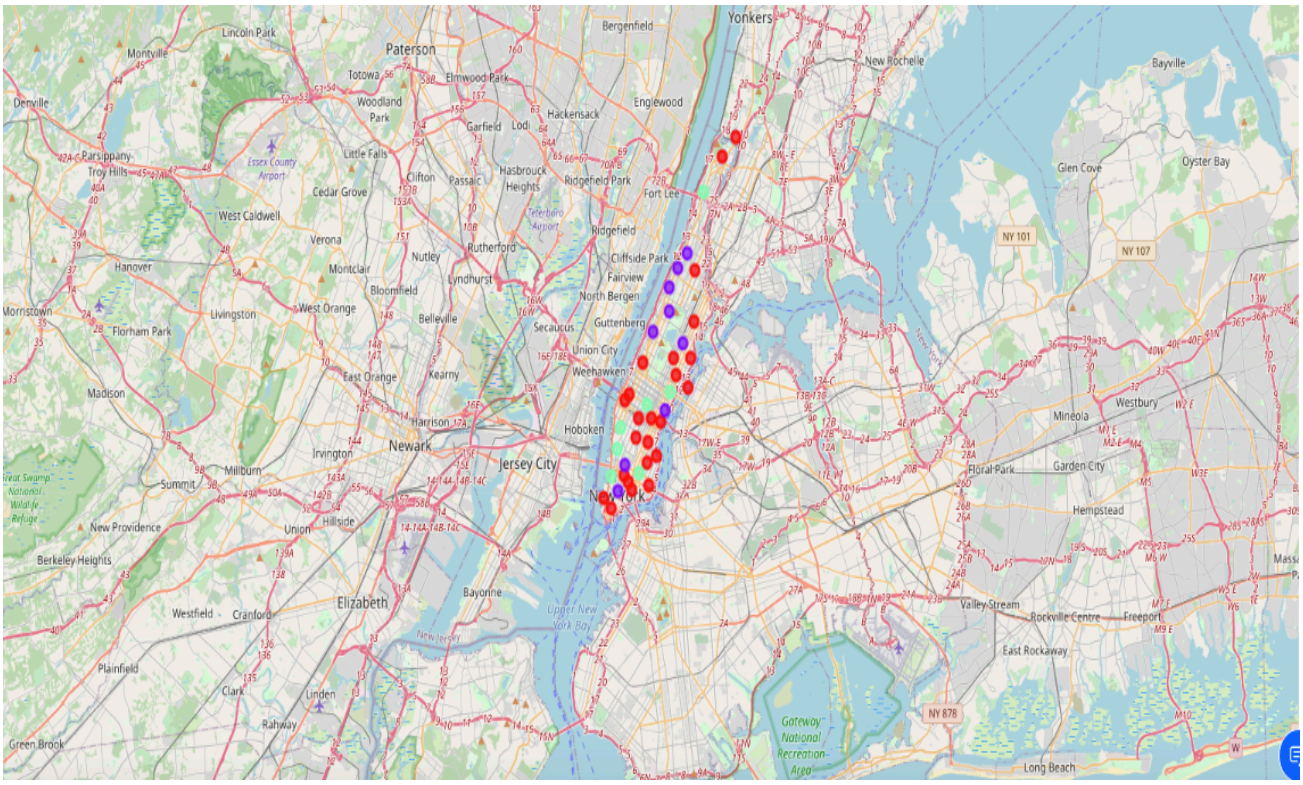
In first step we have collected the New York City **data**. We have then extracted the information about the 5 Boroughs and 306 neighborhoods in the city.
Second step in our analysis will be the exploration of the various neighborhoods in Manhattan. From the various businesses/stores in Manhattan, we are mainly interested in **Indian Restaurants**. Thus we have calculated the locations of Indian restaurants in different neighborhoods in Manhattan.
In third and final step we will focus on most promising areas and within those created **clusters of locations that meet basic requirement** established in discussion with stakeholders: i.e., Neighborhoods with **lowest density** of Indian restaurants. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.



Results and Discussion

The results from the k-means clustering shows that we can categorize the neighborhoods into 3 clusters according to the density of the Indian restaurants:

- **Cluster 0**: Neighborhoods with **lowest/non-existent concentration** of Indian restaurants.
- **Cluster 1**: Neighborhoods with **highest density** of Indian restaurants.
- **Cluster 2**: Neighborhoods with **moderate density** of Indian restaurants.



Thus, Cluster 0 could be a possible area to consider for opening an Indian Restaurant due to negligible competition.

This, of course, does not imply that those clusters are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas where the density of Indian restaurants is very low - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended areas should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify neighborhoods in Manhattan, NYC with low number of Indian restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general neighborhoods that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major neighborhoods of interest (containing greatest number of potential locations) .

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended neighborhood, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.