**Identifying the optimal location for Indian Restaurant in New York City**

Srinivas Nalla

November 14, 2019

1. Introduction

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** New York City.

Since there are lots of restaurants in in those areas, 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 best possible final location can be chosen by stakeholders.

1. Data Acquisition and Cleaning

Based on definition of our problem, factors that will influence our decision are:

* number of existing restaurants in each of the three neighborhoods (any type of restaurant)
* number of and distance to Indian restaurants in each of the neighborhoods, if any
* distance of neighborhood from Manhattan

We decided to use regularly spaced grid of locations, centered around Manhattan, NY, to define our neighborhoods.

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

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
* number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

1. Exploratory Data Analysis
   1. Neighborhood Candidates

We will create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Manhattan.

We will first find the latitude & longitude of Manhattan, using specific, well known address and Google Maps geocoding API.

We will then create a grid of area candidates, equaly spaced, centered around New york city and within ~6km from Manhattan. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).



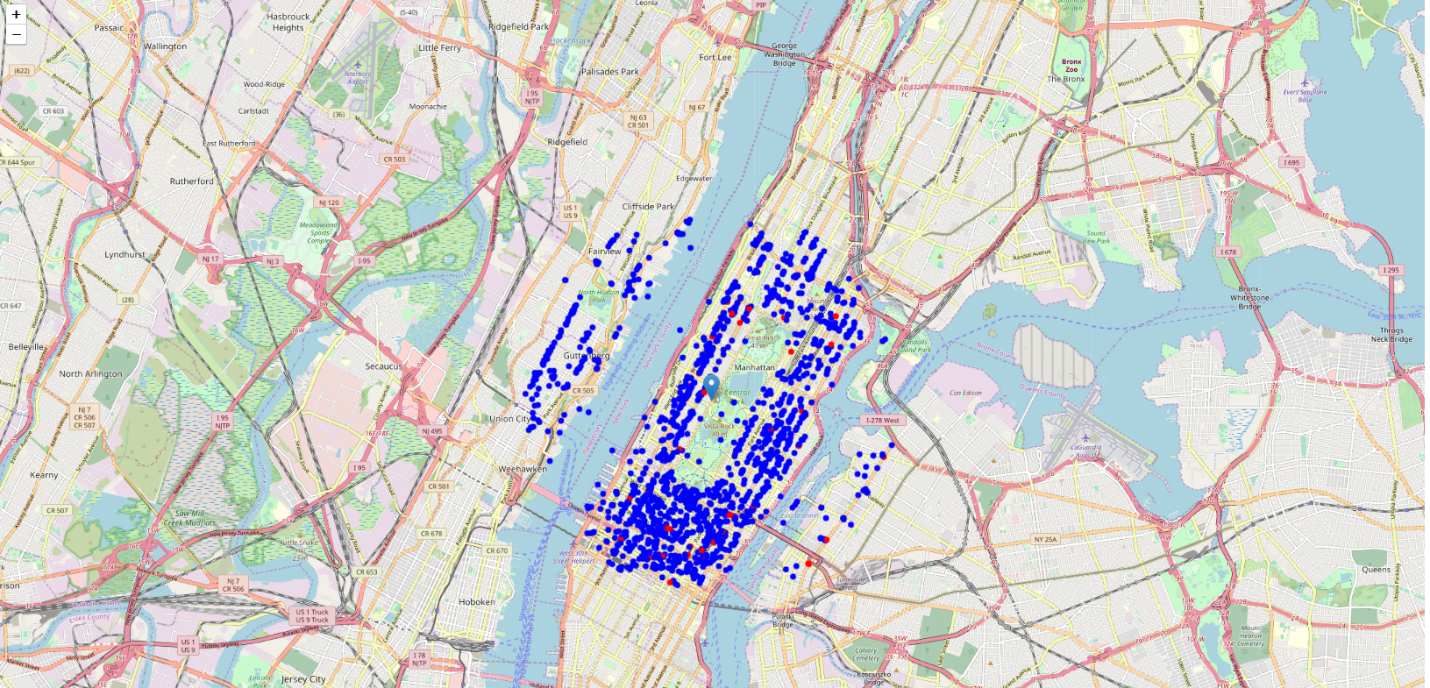
We will then gather all location addresses and their distance from Manhattan:



**Foursquare**

Now that we have our location candidates, we will use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Indian restaurant' category, as we need info on Indian restaurants in the neighborhood.



1. Methodology

## Methodology

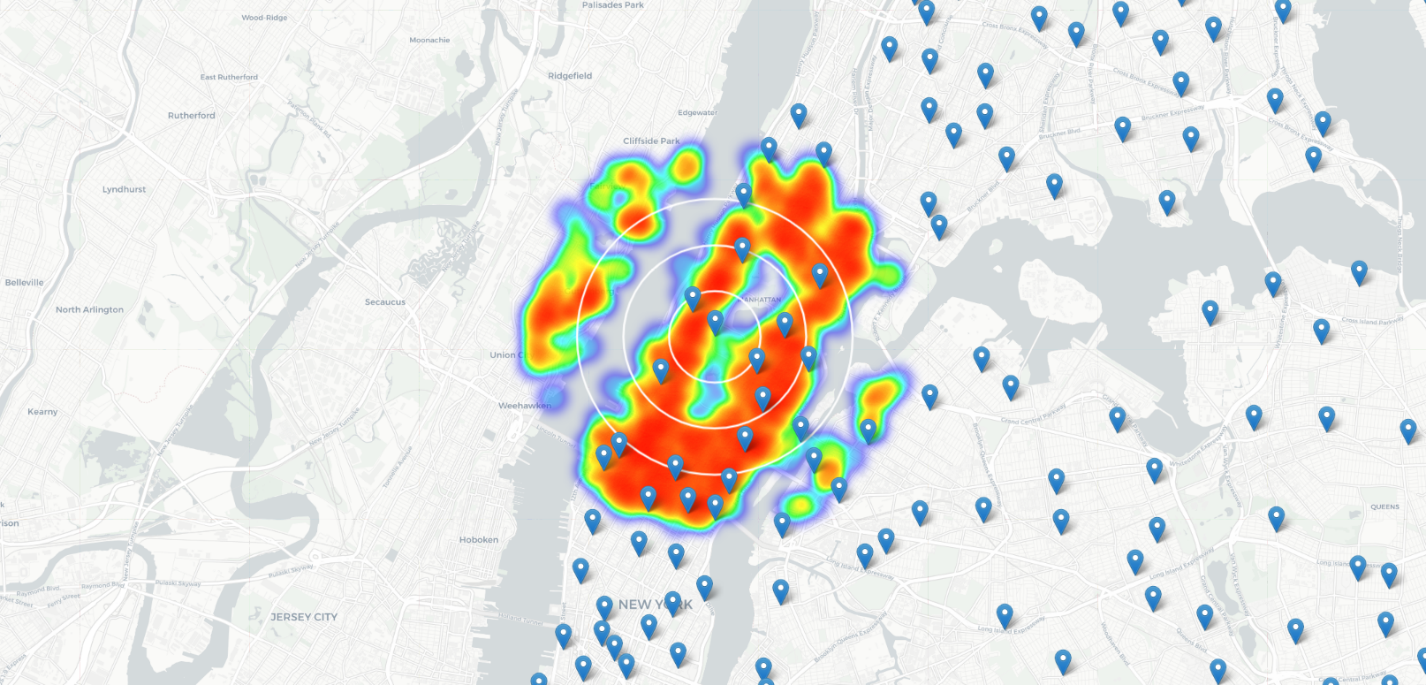
In this project we will direct our efforts on detecting areas of New York city that have low restaurant density, particularly those with low number of Indian restaurants. We will limit our analysis to area ~6km around Manhattan.

In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Manhattan** . We have also **identified Indian restaurants** (according to Foursquare categorization).

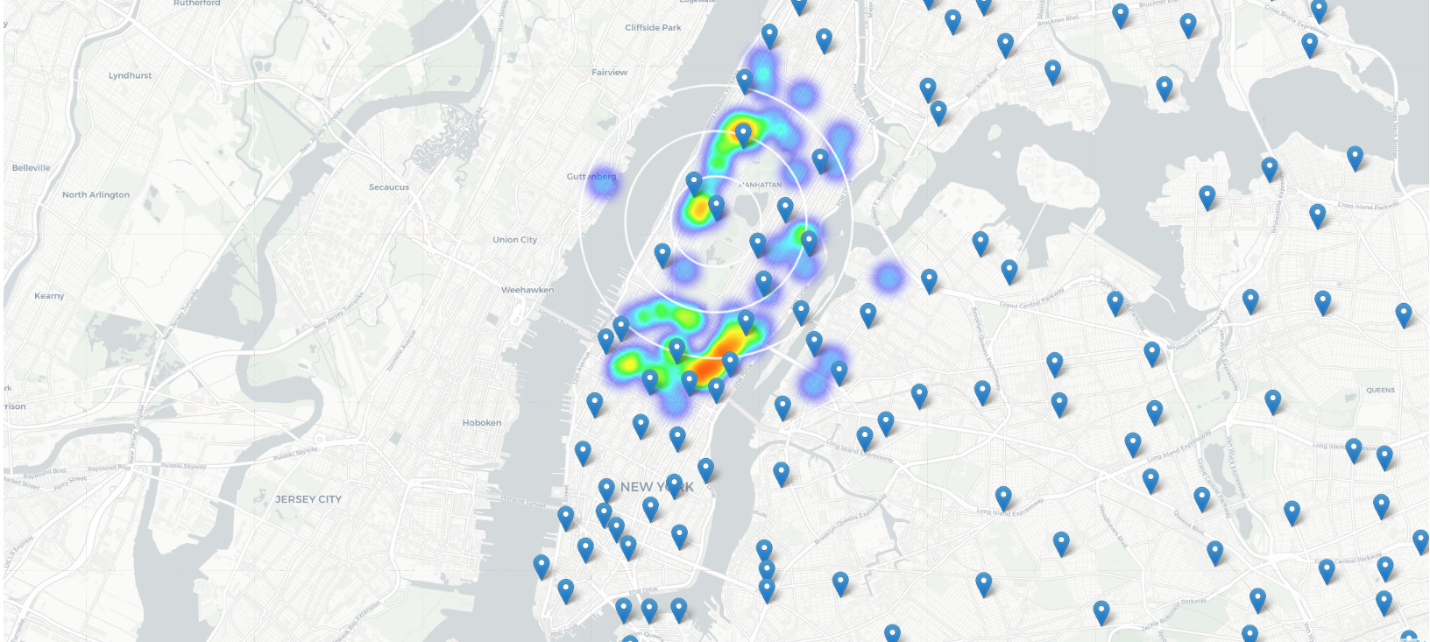
Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Berlin - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (and no Indian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Indian restaurants in radius of 400 meters**. 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.

**Heat Map of all restaurants**



**Heat map of Indian Restaurants**

****

1. Conclusion

## Results and Discussion

Our analysis shows that although there is a great number of restaurants in New York City(~1800 in our initial area of interest which was 12x12km around Manhattan), there are pockets of low restaurant density fairly close to Manhattan. Highest concentration of restaurants was detected most of the Manhattan area with least density around East Harlem. The highest density of Indian restaurants are in between 42nd street and 57th street, close to 2nd and 3rd ave. Again East Harlem area had low density of Indian restaurants compared to other parts of New York city.

## Conclusion

Purpose of this project was to identify NYC areas close to Manhattan with low number of restaurants (particularly Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. Based on our analysis of restaurant density in general around Manhattan and Indian restaurant density in particular, we recommend East Harlem as the preferred location for new Indian Restaurant.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, 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.