Determining optimal locations for Indian restaurants in New York City

Rakesh Satpute

Introduction: Business Problem

The objective of this project is 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 City.

Since there are lots of restaurants in Manhattan, we will try to detect locations that are not already very crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity.

We will use our data science to generate a few most promising neighborhoods based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

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

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Indian restaurants in the neighborhood if any

We decided to use regularly spaced grid of locations, centered around Times Square which is like heart of Manhattan NYC, 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

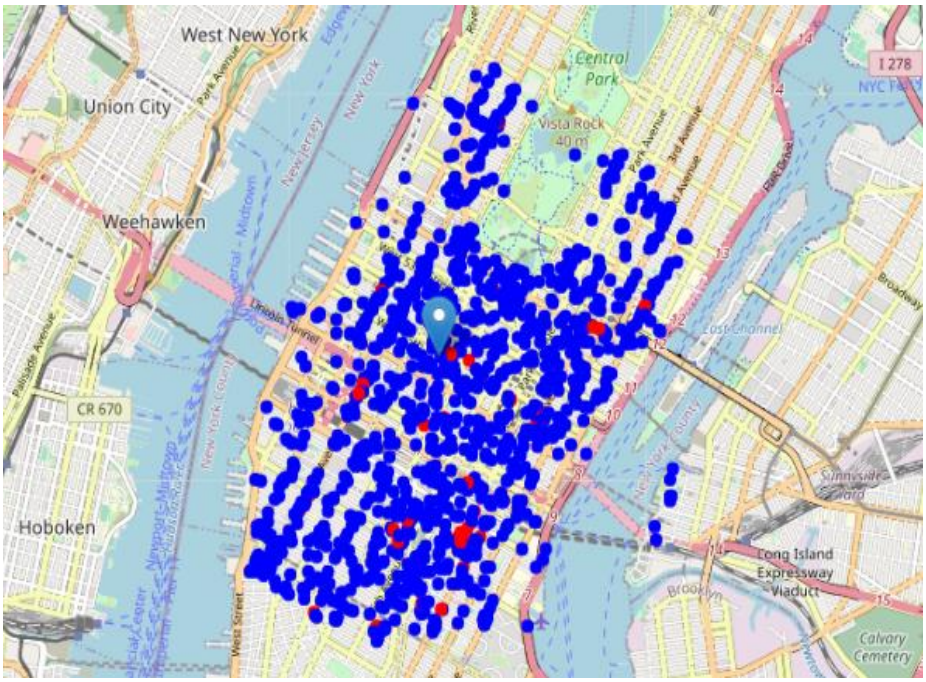
Neighborhood Candidates

Let's 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. 8x8 kilometers centered around Times Square.

Let's first find the latitude & longitude of Times Square, using specific, well known address and Google Maps geocoding API.

Density of restaurants in general and specifically Indian restaurants is determined around 4km of Times Square (below map)

- Total number of restaurants: 1370 (**Blue**)
- Total number of Indian restaurants: 48 (**Red**)
- Percentage of Indian restaurants: 3.50%
- Average number of restaurants in neighborhood: 14.65



Density of Restaurants (blue) & Indian restaurants (Red)

Methodology

Objective is to detect areas of Manhattan that have low restaurant density, particularly those with low number of Indian restaurants. Analysis is limited to area 4km around Times Square. There are 3 main steps.

- Data collection of location and type (category) of every restaurant within 4km from Times Square and identification of Indian restaurants (according to Foursquare categorization).
- Calculation and exploration of '**restaurant density**' across different areas of Manhattan. **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.
- Within those create clusters of locations with **no more than ten restaurants in radius of 250 meters** and **without Indian restaurants in radius of 400 meters**. Clusters of these locations are created (using **k-means clustering**) to identify general zones / neighborhoods / addresses.

Analysis

Some basic explanatory data analysis is performed and some additional info from raw data is derived.

- Average number of restaurants in every area with radius=300m: 14.65

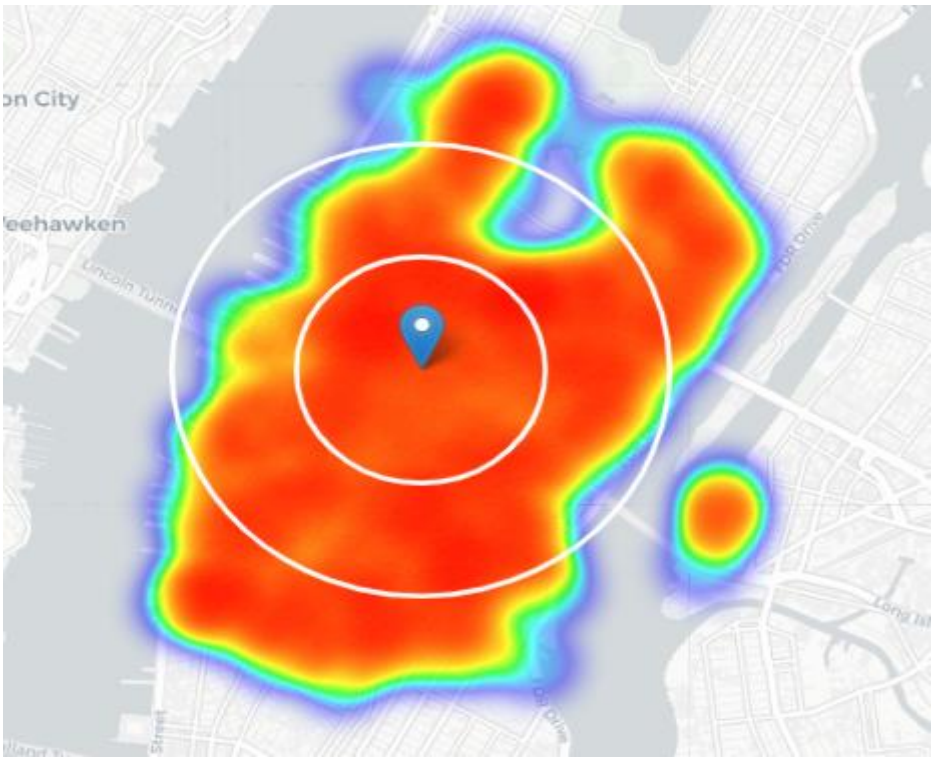
	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	Long Island City, Center Blvd &, 46th Ave, New...	40.747956	-73.957686	-5.816148e+06	9.863532e+06	3973.663297	5
1	Southpoint Park, Roosevelt Island, NY 10044	40.751493	-73.957557	-5.815548e+06	9.863532e+06	3772.267223	0
2	200 W Loop Rd, New York, NY 10044	40.755030	-73.957428	-5.814948e+06	9.863532e+06	3659.234893	0
3	31 Ed Koch Queensboro Bridge Path, New York, N...	40.758567	-73.957299	-5.814348e+06	9.863532e+06	3642.801120	3
4	1194 York Ave, New York, NY 10065	40.762105	-73.957170	-5.813748e+06	9.863532e+06	3724.244890	24
5	404 E 69th St, New York, NY 10021	40.765642	-73.957041	-5.813148e+06	9.863532e+06	3897.435054	7
6	1-01-1-02 54th Ave, Long Island City, NY 11101	40.742736	-73.961906	-5.817048e+06	9.864051e+06	3996.248241	2
7	4749 Center Blvd, Long Island City, NY 11109	40.746272	-73.961777	-5.816448e+06	9.864051e+06	3651.027253	0
8	E Rd, New York, NY 10044	40.749809	-73.961649	-5.815848e+06	9.864051e+06	3377.869151	0
9	1025 E 51st Street Pedestrian Crossing, New Yo...	40.753346	-73.961520	-5.815248e+06	9.864051e+06	3195.309062	0

Distance to nearest Indian restaurant from every area candidate center is calculated

- Average distance to closest Indian restaurant from each area center: 900.21

So, on average Indian restaurant can be found within ~1km from every area center candidate.

Then a map is created showing **heatmap / density of restaurants** and try to extract some meaningful info from that. Also, neighborhoods of Manhattan are depicted on the map and a few circles indicating distance of 1km and 2km from Times Square.

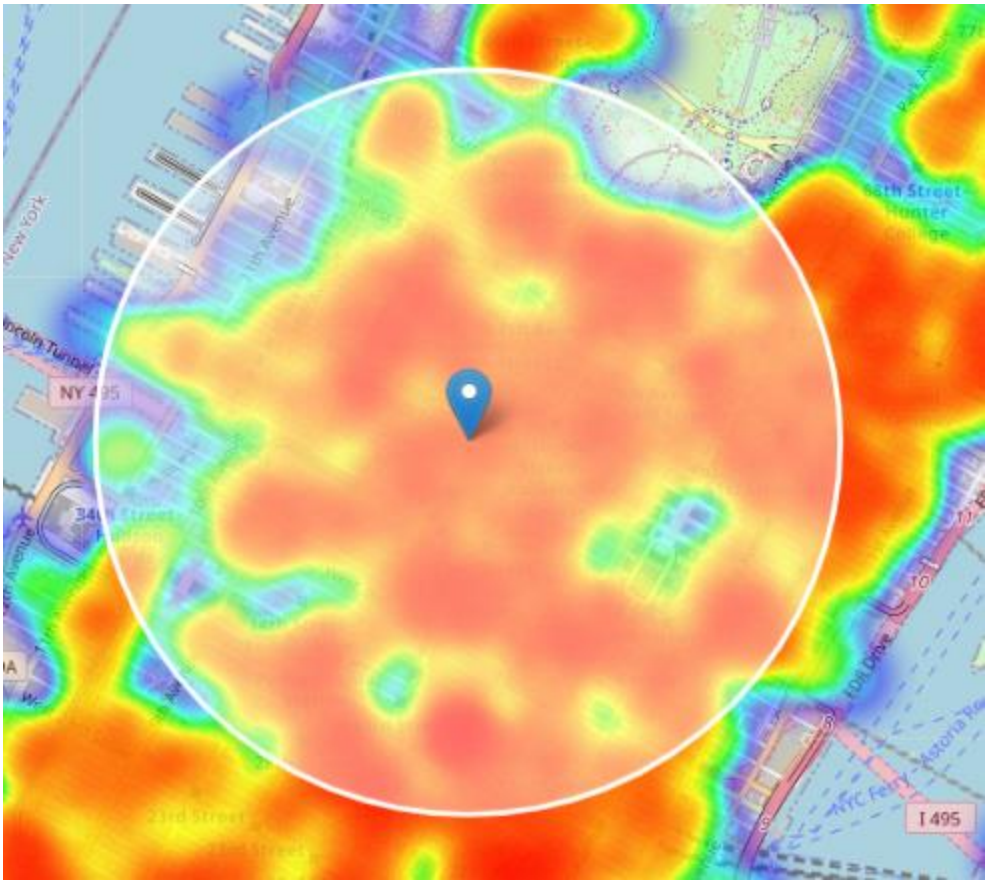


Heat map of restaurants around Times Square

Heat map of Indian restaurants around Times Square

This map is not so 'hot' (Indian restaurants represent a subset of ~3.5% of all restaurants in Manhattan). The Indian restaurants density is low pretty much everywhere except west of Times Square. And in 1.5 km around times square the density is pretty low.

Based on this, we will focus our analysis on areas around Times Square within 1.5 km radius.

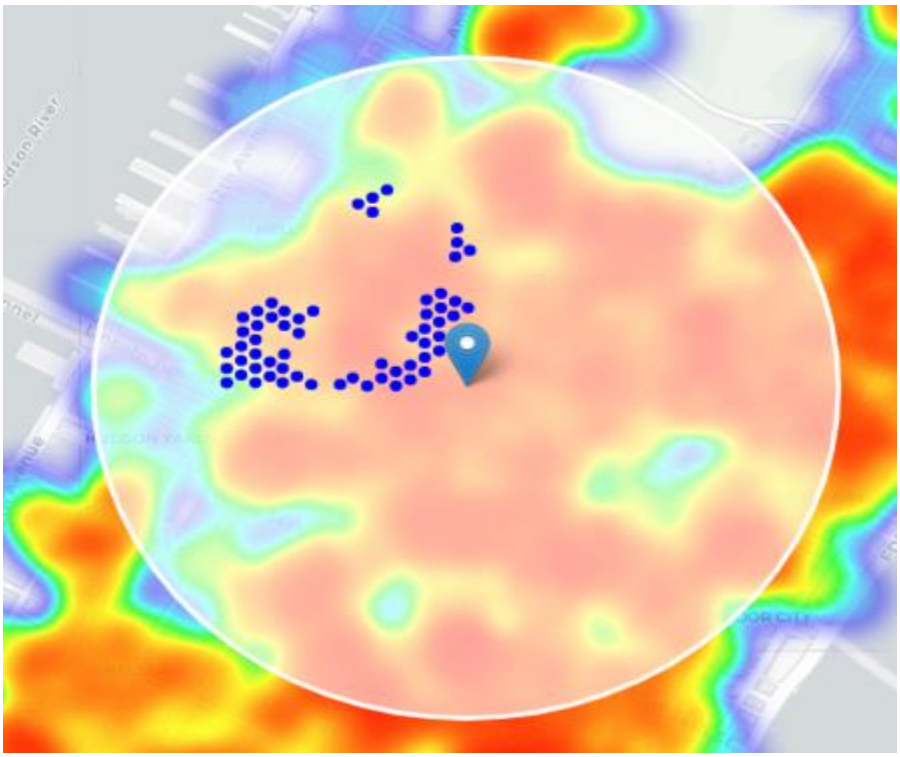


Heat map of restaurants around 1.5 km radius of Times Square

Further analysis is focused on west of Times Square which has very low density of Indian restaurants and within 1.5 km radius.

Locations with no more than ten restaurants in radius of 250 meters, and no Indian restaurants in radius of 400 meters.

- Locations with no more than ten restaurants nearby: 91
- Locations with no Indian restaurants within 400m: 154
- Locations with both conditions met: 60



Locations satisfying both conditions

- Clear indication of zones with low number of restaurants in vicinity, and no Indian restaurants at all nearby is determined.
- These locations are clustered using K-means clustering create centers of zones containing good locations.
- Those zones, their centers and addresses are the final result of analysis.

Clusters of desired locations of Indian restaurants

Results

The candidate area centers are reverse geocoded to get the addresses which can be presented to stakeholders.

- 560 10th Ave, New York, NY 10036 => 1.3km from Times Square
- 825 8th Ave, New York, NY 10019 => 0.7km from Times Square
- 711 8th Ave, New York, NY 10036 => 0.4km from Times Square
- 444 W 52nd St, New York, NY 10019 => 1.4km from Times Square
- 605 10th Ave, New York, NY 10036 => 1.3km from Times Square

Locations of candidate area centers

Results and Discussion

Our analysis shows that there is a great number of restaurants in Manhattan (~1400 in our initial area of interest which was 4X4 km around Times Square), there are pockets of low Indian restaurant density fairly close to city center. So we focused our attention to these areas which are particularly to the west of Times Square.

After directing our attention to this more narrow area of interest (covering approx. 1.5X1.5km west from Times Square) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than ten restaurants in radius of 250m and those with an Indian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 5 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Indian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Times Square but not crowded with existing restaurants (particularly Indian) - 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 zones 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 Manhattan areas close to Times Square (heart of New York City) 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 areas that justify further analysis (west of Times Square), 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 zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission 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.