

Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

09.10.2019

Kumar Gaurav
CSE
BIT SINDRI
Dhanbad 828123

Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Introduction: Business Problem

Delhi being the capital city of India and one of the oldest city having historical values as well as political and industrial hub of the country attracts a large number of people from other countries as well as different states of the country. And thus, a city having such a diverse population raise a need of diverse taste. Therefore, 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 a full cuisine **Indian Restaurant in Delhi, India**.

Since there are many restaurants in Delhi we will try to find locations which is not already crowded with restaurants. Also, we will in interested in finding area which have no or very less number of indian restaurants in the vicinity which serve the full indian cuisine e.g north, south, east etc. We would also prefer locations **as close to the old Delhi as possible**, assuming that first two conditions are met.

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.

Data

We will use regularly spaced grid of locations, centred around chandni chowk, to define our neighborhoods.

Foursquare API will be used to extract the data required such as centres of candidate areas will be generated and stored in a dataframe with their latitude and longitude values. Also, the number of restaurants and their type and location in every neighborhood will be obtained and pushed into the dataframe.

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 approx. 12x12 kilometers centered around old Delhi city.

Once we have our location candidates, let's use the Foursquare API to get info on restaurants in each neighborhood.

Factors influencing our decision in view of our problem are:

- any type of restaurant existing in the neighborhood.
- If yes, then number of and distance to restaurants in the neighborhood.
- distance of neighborhood from the designated location.

Methodology

In this project we focussed our efforts in finding locations with low restaurants density in the proximity of around 6 km to old Delhi.

In the first step, we have collected the required data i.e location and various restaurants within 6 km from old Delhi using the Foursquare API.

Second step involves the analysis of our extracted data and explore restaurant density across different areas of Delhi. Heatmaps will be used to identify a few promising areas close to the location with low number of restaurants and will focus our attention on those areas.

In the last 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. 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.

Analysis

Some exploratory data analysis is done to derive additional info from our dataset gathered.

Firstly, We counted the number of restaurants in every area candidate. Then, distance to the nearest restaurant from every area is calculated. And used this analysis to find the average distance in which a restaurant could be found.

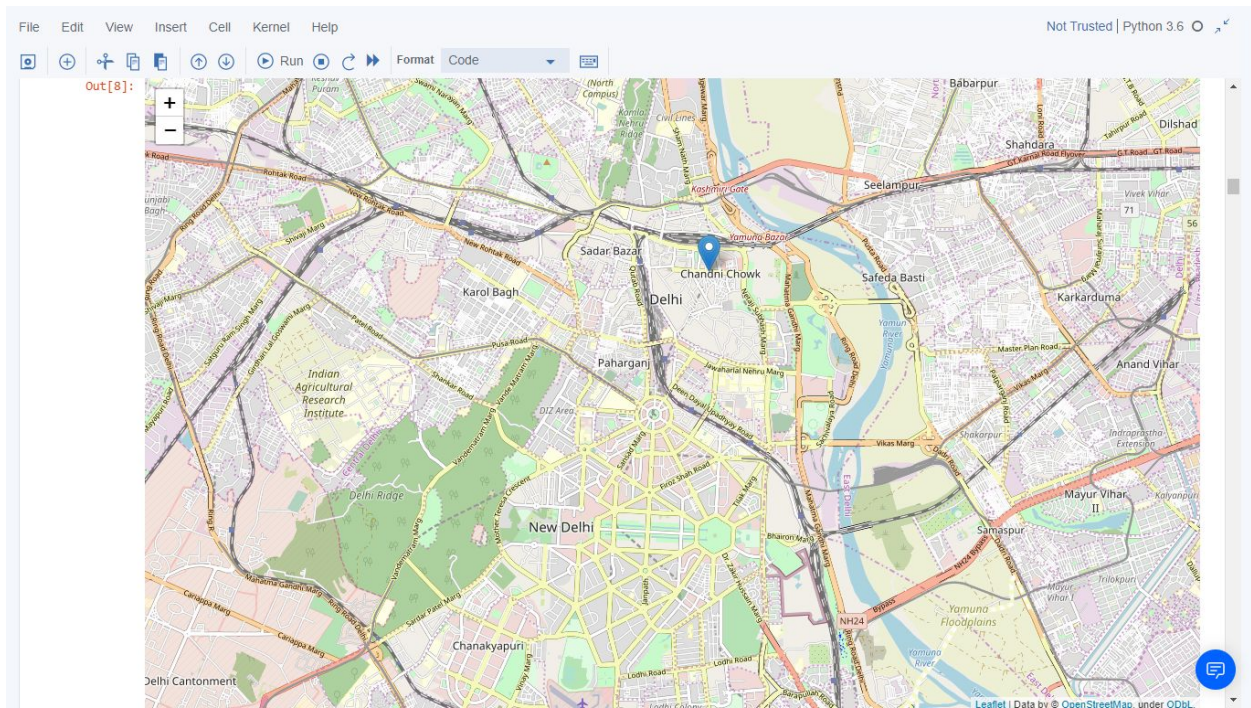
Out[12]:

	Address	Latitude	Longitude	X	Y	Distance from center
0	Chandni Chowk, Delhi, Delhi	77.183822	28.532484	832288.241552	8.605746e+06	5992.495307
1	Chandni Chowk, Delhi, Delhi	77.182594	28.556040	832888.241552	8.605746e+06	5840.376700
2	Chandni Chowk, Delhi, Delhi	77.181364	28.579591	833488.241552	8.605746e+06	5747.173218
3	Chandni Chowk, Delhi, Delhi	77.180132	28.603137	834088.241552	8.605746e+06	5715.767665
4	Chandni Chowk, Delhi, Delhi	77.178898	28.626678	834688.241552	8.605746e+06	5747.173218
5	Chandni Chowk, Delhi, Delhi	77.177662	28.650215	835288.241552	8.605746e+06	5840.376700
6	Chandni Chowk, Delhi, Delhi	77.176424	28.673748	835888.241552	8.605746e+06	5992.495307
7	Chandni Chowk, Delhi, Delhi	77.190188	28.501921	831388.241552	8.606266e+06	5855.766389
8	Chandni Chowk, Delhi, Delhi	77.188962	28.525491	831988.241552	8.606266e+06	5604.462508
9	Chandni Chowk, Delhi, Delhi	77.187735	28.549056	832588.241552	8.606266e+06	5408.326913

...and let's now save/persist this data into local file.

OK, so on average restaurant can be found within ~500m from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's create a map showing heatmap / density of restaurants and try to extract some meaningful info from that. Also, let's show borders of Delhi boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Chandni Chowk.



Analysis of popular travel guides and web sites often mention Connaught Place as beautiful, interesting, rich with culture, 'hip' and 'cool', centre of old and new Delhi neighborhoods popular with tourists and businessman.

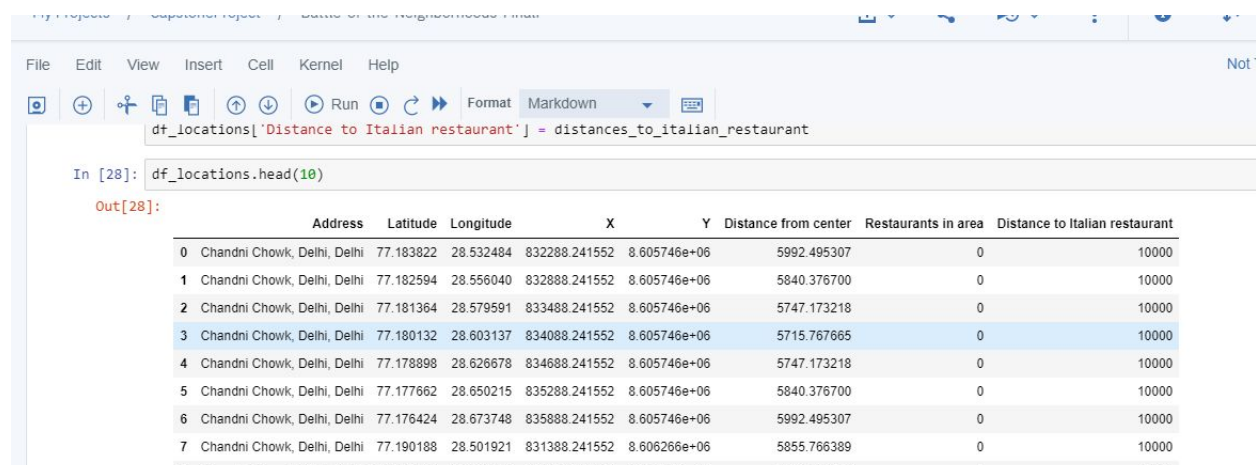
Popular with tourists, booming and trendy, relatively close to city center and well connected, this boroughs appear to justify further analysis.

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of Connaught Place closest to old Delhi.

City	Address	Locality	Longitude	Latitude
New Delhi	Ground Floor, Plot 482, Haveli Haider Quli, Near Andhra Bank, Ch	Chandni Chowk	77.224214	28.6564057
New Delhi	5080, Ballimaran Road, Katra Nawab, Chandni Chowk, New Delhi	Chandni Chowk	77.2278068	28.6545089
New Delhi	6283, Main Bara Hindu Rao, Near choti Masjid, MD Ismail Marg, S	Chandni Chowk	77.2094826	28.663599
New Delhi	1984-1985, Gali Paranthé Wali, Chandni Chowk, New Delhi	Chandni Chowk	77.2304115	28.6560115
New Delhi	Shop 2202, Bazar Chitli Qabar, Near Jama Masjid, Chandni Chow	Chandni Chowk	77.2335101	28.6494534
New Delhi	251, Near Fatehpuri Masjid, Fatehpuri, Chandni Chowk, New Del	Chandni Chowk	77.22398333	28.65688889
New Delhi	382, Kucha Ghasi Ram, Fatehpuri, Chandni Chowk, New Delhi	Chandni Chowk	77.2244176	28.6566711
New Delhi	376, Near Kucha Ghasi Ram, Chandni Chowk, New Delhi	Chandni Chowk	77.2243841	28.6568732
New Delhi	379, Near Kucha Ghasi Ram, Chandni Chowk, New Delhi	Chandni Chowk	77.2243823	28.656851
New Delhi	Old Delhi Railway Station, Mori Gate, Chandni Chowk, New Delhi	Chandni Chowk	77.2280206	28.6606804
New Delhi	187, Church Mission Road, Fatehpuri, Chandni Chowk, New Delhi	Chandni Chowk	77.2231981	28.6578654
New Delhi	5469, Opposite Town Hall, Chandni Chowk, New Delhi	Chandni Chowk	77.2271781	28.6563302
New Delhi	711, Church Mission Road, Fatehpuri, Chandni Chowk, New Delhi	Chandni Chowk	77.2234955	28.658756
New Delhi	659, Church Mission Road, Fatehpuri, Chandni Chowk, New Delhi	Chandni Chowk	77.2234057	28.6581203
New Delhi	Near Sis Ganj Sahib Gurudwara, Chandni Chowk, New Delhi	Chandni Chowk	77.23421395	28.65606667
New Delhi	Platform 3, Old Delhi Railway Station, Chandni Chowk, New Delhi	Chandni Chowk	77.22333167	28.65664
New Delhi	Main Road Near Central Bank, Opposite Paranthé Wali Gali, Chan	Chandni Chowk	77.2305013	28.6561992
New Delhi	Gali Barf Wali, Near Kinari Bazar, Chandni Chowk, New Delhi	Chandni Chowk	77.23	28.66
New Delhi	Opposite Town Hall, Chandni Chowk, New Delhi	Chandni Chowk	77.2268188	28.6565647
New Delhi	9074, Gali Paranthé Wale, Chandni Chowk, New Delhi	Chandni Chowk	77.2307138	28.6560495
New Delhi	42-B, Gali Parathewali, Opposite Central Bank, Chandni Chowk, N	Chandni Chowk	77.2305013	28.6561097
New Delhi	453, Khari Baoli, Chandni Chowk, New Delhi	Chandni Chowk	77.2155012	28.6625615
New Delhi	1894, Main Road, Opposite Dariba Kalan, Chandni Chowk, New D	Chandni Chowk	77.2324772	28.6562981
New Delhi	105-110, Bazar Matia Mahal, Opposite Jama Masjid Gate 1, Chan	Chandni Chowk	77.233555	28.6491441
New Delhi	Gali Paranthé Wali, Chandni Chowk, New Delhi	Chandni Chowk	77.2306151	28.6559026
New Delhi	Shop 1462, Near Sheesh Ganj Sahib Gurudwara, Chandni Chowk	Chandni Chowk	77.2320281	28.656524
New Delhi	934, Kucha Pati Ram, Sitaram Bazar, Chandni Chowk, New Delhi	Chandni Chowk	77.2269535	28.6476635

OK. Now let's calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use radius of 250 meters) and distance to closest restaurant.

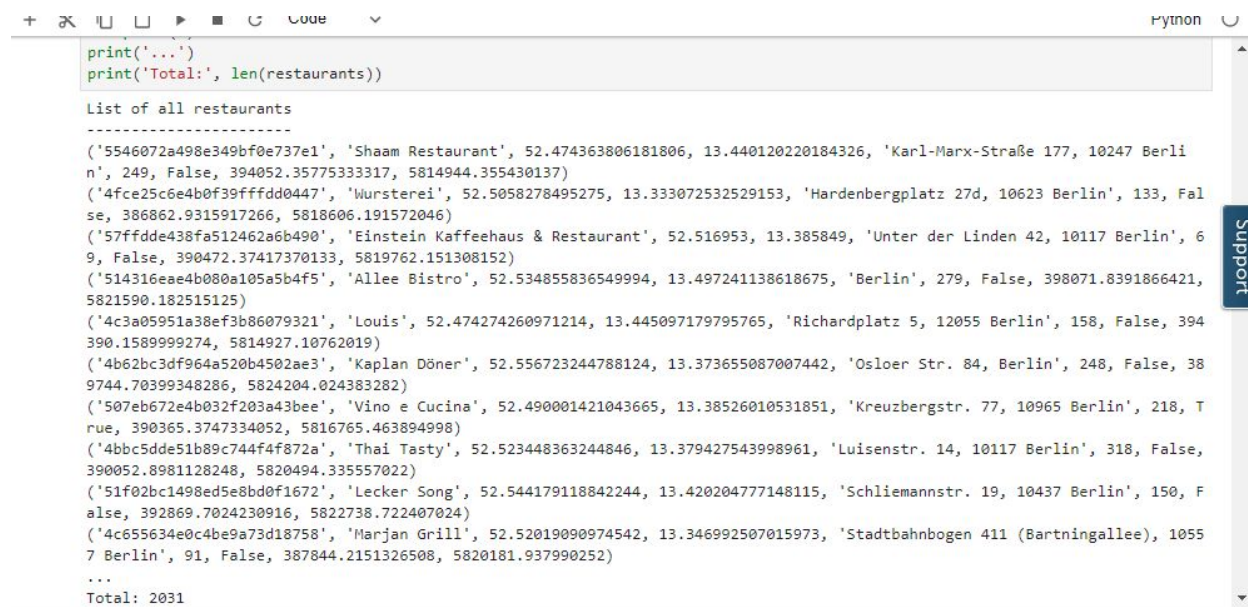
OK. Let us now filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters.



Let's now show those good locations in a form of heatmap:

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and *no* restaurants at all nearby.

Let us now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.



Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

IBM Watson Studio

My Projects / CapstoneProject / Battle-of-the-Neighborhoods-FinalF

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6

Out[26]:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	Chandni Chowk, Delhi, Delhi	77.183822	28.532484	832288.241552	8.605746e+06	5992.495307	0
1	Chandni Chowk, Delhi, Delhi	77.182594	28.556040	832888.241552	8.605746e+06	5840.376700	0
2	Chandni Chowk, Delhi, Delhi	77.181364	28.579591	833488.241552	8.605746e+06	5747.173218	0
3	Chandni Chowk, Delhi, Delhi	77.180132	28.603137	834088.241552	8.605746e+06	5715.767665	0
4	Chandni Chowk, Delhi, Delhi	77.178898	28.626678	834688.241552	8.605746e+06	5747.173218	0
5	Chandni Chowk, Delhi, Delhi	77.177662	28.650215	835288.241552	8.605746e+06	5840.376700	0
6	Chandni Chowk, Delhi, Delhi	77.176424	28.673748	835888.241552	8.605746e+06	5992.495307	0
7	Chandni Chowk, Delhi, Delhi	77.190188	28.501921	831388.241552	8.606266e+06	5855.766389	0
8	Chandni Chowk, Delhi, Delhi	77.188962	28.525491	831988.241552	8.606266e+06	5604.462508	0
9	Chandni Chowk, Delhi, Delhi	77.187735	28.549056	832588.241552	8.606266e+06	5408.326913	0

OK, now let's calculate the **distance to nearest Italian restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).

```
In [27]: distances_to_italian_restaurant = []
for area x, area y in zip(xs, ys):
```

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no restaurants nearby, all zones being fairly close to city center (all less than 4km from Chandni Chowk). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Connaught Place boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a great number of restaurants in Delhi (~500 in our initial area of interest which was 12x12km around Chandni Chowk), there are pockets of low restaurant density fairly close to way to Connaught Place. Highest concentration of restaurants was detected in the middle of Chandni Chowk, so we focused our attention to areas moving out of Chandni Chowk corresponding to Kasmiri Gate, Chawri Bazaar, Connaught Place. Our attention was focused on Connaught Place and Chawri Bazaar which

offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km from Chandni Chowk) we first created a dense grid of location candidates (spaced 100m apart); those locations were filtered further to remove neighborhood which have more than 2 restaurants in the vicinity of ~200 or more than 4 in ~500.

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 these 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 old Delhi but not crowded with existing restaurants - 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

The main aim of the project was to find areas close to old Delhi with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant preferably Indian. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Connaught Place), 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 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.