Optimal Location for Opening a Restaurant

IBM Data Science Capstone Project

Nishit Chaudhari

Introduction: Business Problem

Asian food is gaining popularity in England lately. Asian restaurants are very often crowded especially in England area. In England center, Thai, Chinese and even Indian restaurants are opening everywhere, and are always full whatever is the hour. The aim of this Project is to find the best localization to open a restaurant business in capital of England and United Kingdom, London City.

Prior launching any restaurant, it's important to know if the business as a good opportunity. In order to do so, this report will try to gather data about other restaurant localization, competitors and best localization.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders. As the goal of this is to create a business plan in the end, we need to make sure data from Api are correct. We also need to check that customer could be interested in this specific business.

Data

The idea is of clustering the city by area and then plot heatmap to find better area.

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 Geocoder reverse geocoding Python package as recently Google started charging for their API.

Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API coordinate of London center will be obtained using Geocoder geocoding Python package of well-known London location (Trafalgar Square)

Following is some change in data:

- Country/City: London, England.
- Goal: Open a restaurant/little shop.

Methodology

I. Neighborhood Candidates

First step is to 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.. 1.5km kilometers centered around London city center.

Let's first find the latitude & longitude of London city center, using specific, well-known address and Geocoder geocoding Python package.

We'll consider the Trafalgar Square to be the city center, as a lot of companies are around.

The geographical coordinate of Trafalgar Square is 51.508037, -0.12804941070724718.

II. City partitioning

Now let's create a grid of area candidates, equally spaced, centered around city center and within ~1.5km from Trafalgar Square. Our neighborhoods will be defined as circular areas with a radius of 100 meters, so our neighborhood centers will be 200 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).

create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all its neighbors.

Let's visualize the data we have so far: city center location and candidate neighborhood centers:

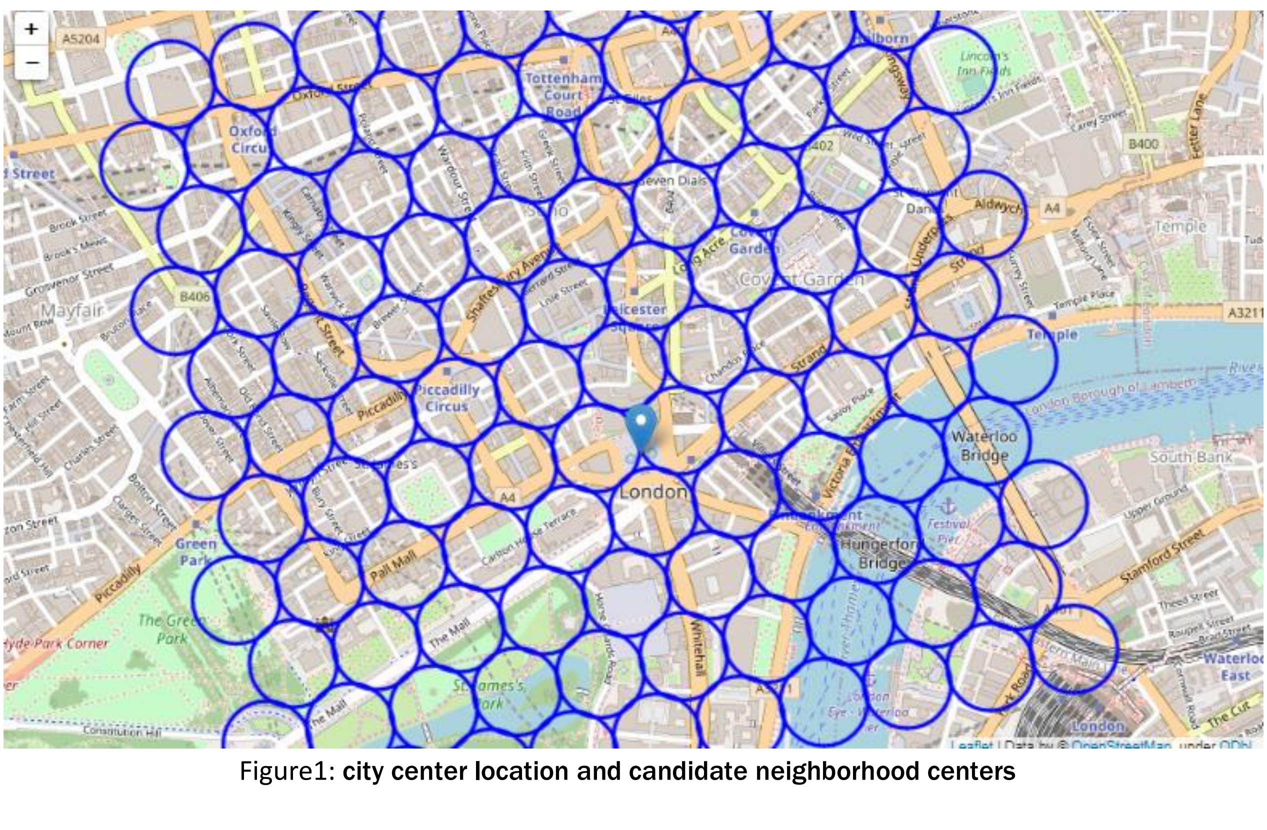

Figure1: **city center location and candidate neighborhood centers**

III. Address Locations

Now use Geocoder Python package to get approximate addresses of those locations and place all this into a Pandas data frame.

	Address	Latitude	Longitude	X	Y	Distance from center
0	Wellington Barracks, Birdcage Walk, Westminste...	51.500719	-0.138645	-547923.445676	5.814989e+06	1111.720843
1	Wellington Barracks, Birdcage Walk, Westminste...	51.501086	-0.135864	-547723.445676	5.814989e+06	957.038784
2	The Royal Parks, Birdcage Walk, Westminster, C...	51.501454	-0.133082	-547523.445676	5.814989e+06	822.145506
3	Birdcage Walk, Westminster, City of Westminste...	51.501821	-0.130301	-547323.445676	5.814989e+06	718.277964
4	Foreign, Commonwealth and Development Office, ...	51.502189	-0.127520	-547123.445676	5.814989e+06	660.244828
5	Laundry Road, Westminster, City of Westminster...	51.502556	-0.124738	-546923.445676	5.814989e+06	660.244828
6	Westminster Pier, Victoria Embankment, Westmin...	51.502923	-0.121956	-546723.445676	5.814989e+06	718.277964
7	Lambeth, London Borough of Lambeth, London, Gr...	51.503290	-0.119175	-546523.445676	5.814989e+06	822.145506
8	Two Southbank Place, York Road, Lambeth, Londo...	51.503657	-0.116393	-546323.445676	5.814989e+06	957.038784
9	Hole In the Wall, 5, Mepham Street, Lambeth, L...	51.504024	-0.113611	-546123.445676	5.814989e+06	1111.720843

Figure2: Address Locations

IV. Foursquare

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

We're interested in venues in 'food' category, but only the ones who can be competitors, this mean food truck, quick food, take away, healthy, not restaurant taking too long.

We will also list all companies and university to evaluate the customer pool. Now see all the collected restaurants in our area of interest on map, and let's also show Asian restaurants in different color

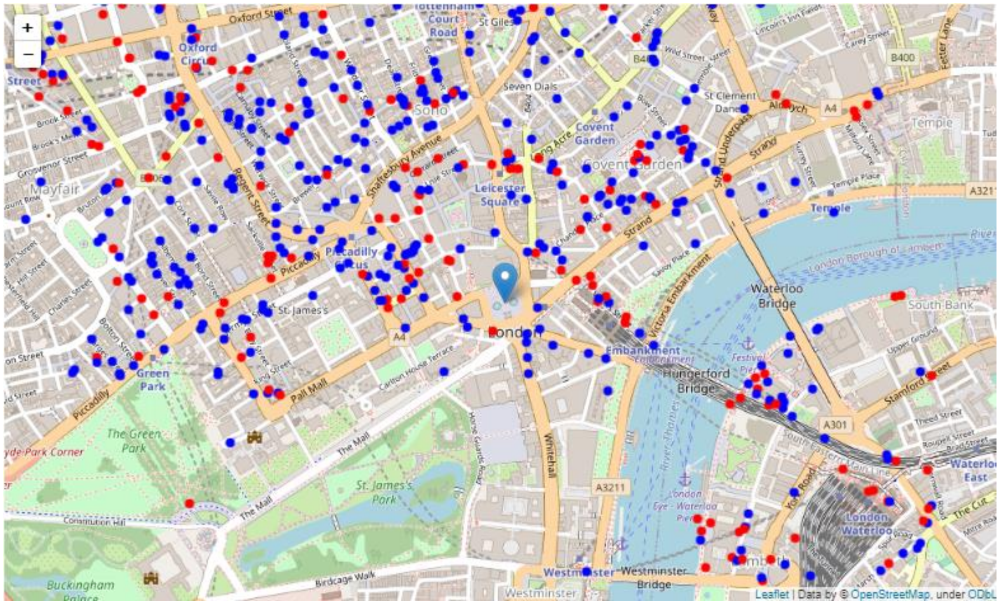

Figure3: collected restaurants in our area of interest on map (Asian restaurants = red)

So, on this map we can see all potential competitors, take away restaurant, healthy restaurant. Asian restaurant is in red.

Now we need to analyze companies/university localization and cluster and cross both analyses.

We need to remember that the target is worker wanting to buy healthy fast food for breakfast and lunch, we don't aim the evening. Also, another target can be universities.

Global map of worker/students versus Restaurants:

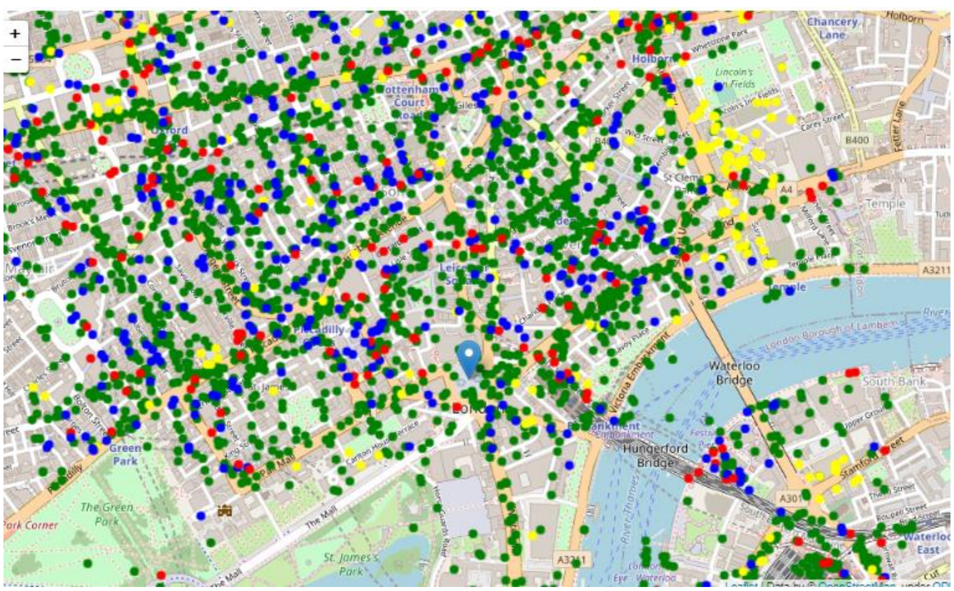

Figure4: Global map of worker/students versus Restaurants (Asian restaurants = red, Universities = yellow, companies = green, other restaurants = blue)

we have a map with the big companies and universities, we can then already see main areas with potential customers.

After visualization of data, we can conclude:

- Competitors are scattered around the city.
- Trafalgar Square is crowded with companies everywhere within the 1.5km region
- Universities are mostly crowded to the east and north-east regions.

We will now focus on spotting worker area more than competitors; we will try to determine where are clusters more of the worker.

Analysis

I. Average number of customers

Perform some basic explanatory data analysis and derive some additional info from our raw data. First count the number of businesses in every area candidate and then find average number of customers in every area with certain radius. Putting it all in pandas' data frame for better visualization:

	Address	Latitude	Longitude	X	Y	Distance from center	Customers in area
9	Hole In the Wall, 5, Mepham Street, Lambeth, L...	51.504024	-0.113611	-546123.445676	5.814989e+06	1111.720843	29
77	itsu, 2-4, Neal Street, Covent Garden, City of...	51.513631	-0.124133	-546623.445676	5.816201e+06	688.029485	26
66	22, Floral Street, Covent Garden, City of West...	51.511944	-0.125014	-546723.445676	5.816028e+06	489.348060	25
83	Princi, 135, Wardour Street, Soho, London, Isl...	51.513848	-0.134381	-547323.445676	5.816375e+06	792.027538	24
51	Standbrook House, 2-5, Old Bond Street, St. Ja...	51.508419	-0.139805	-547823.445676	5.815855e+06	827.972509	23
81	Hamleys, 188-196, Regent Street, St. James's, ...	51.513113	-0.139945	-547723.445676	5.816375e+06	1013.561849	22
99	Superdrug, 232-233, High Holborn, Holborn, Lon...	51.517372	-0.119588	-546223.445676	5.816548e+06	1208.813745	22
101	15-17, Great Portland Street, Fitzrovia, City ...	51.516120	-0.140965	-547723.445676	5.816721e+06	1286.527776	21
67	Penhaligon's, Covent Garden, City of Westminst...	51.512311	-0.122232	-546523.445676	5.816028e+06	632.029686	20
65	Leicester Square Platforms 1-2, Cranbourn Stre...	51.511576	-0.127796	-546923.445676	5.816028e+06	399.326338	19

Figure5: **average number of customers in every area with radius=100m**

The mains areas are scattered around Trafalgar Square usually around 500m - 1000m away.

II. Heatmap/density of customers.

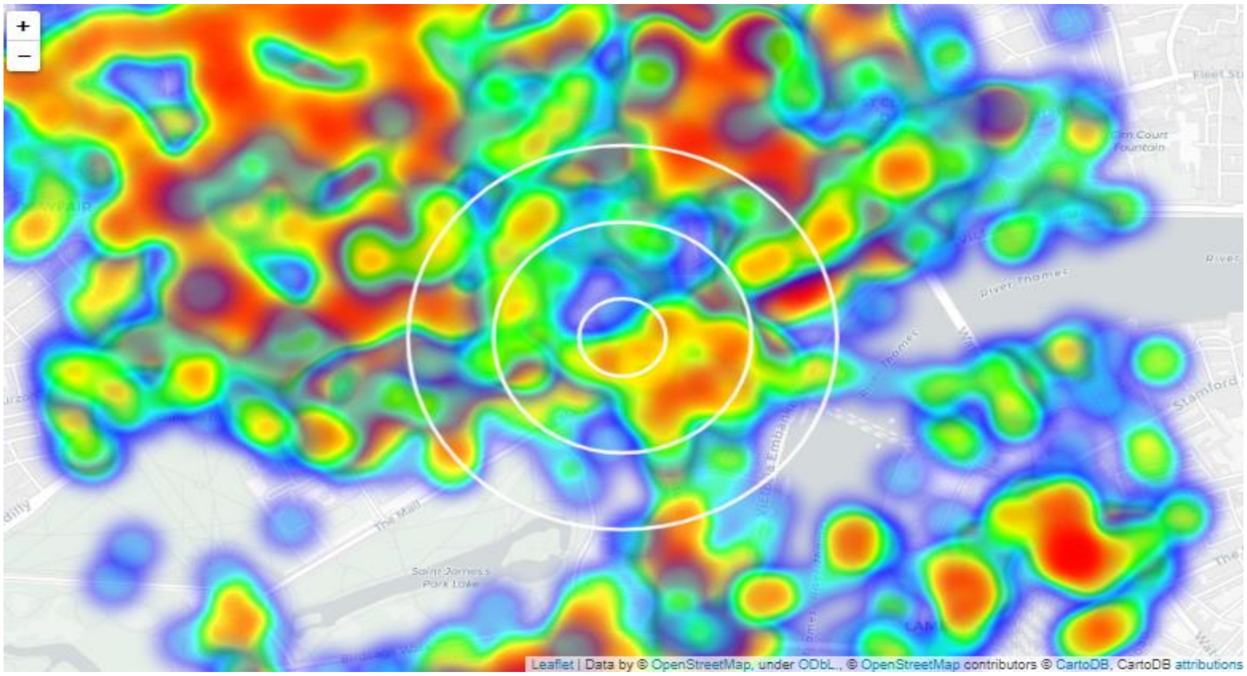

Figure6: Heatmap map showing heatmap/density of customers.

III. Heatmap/density of restaurants.

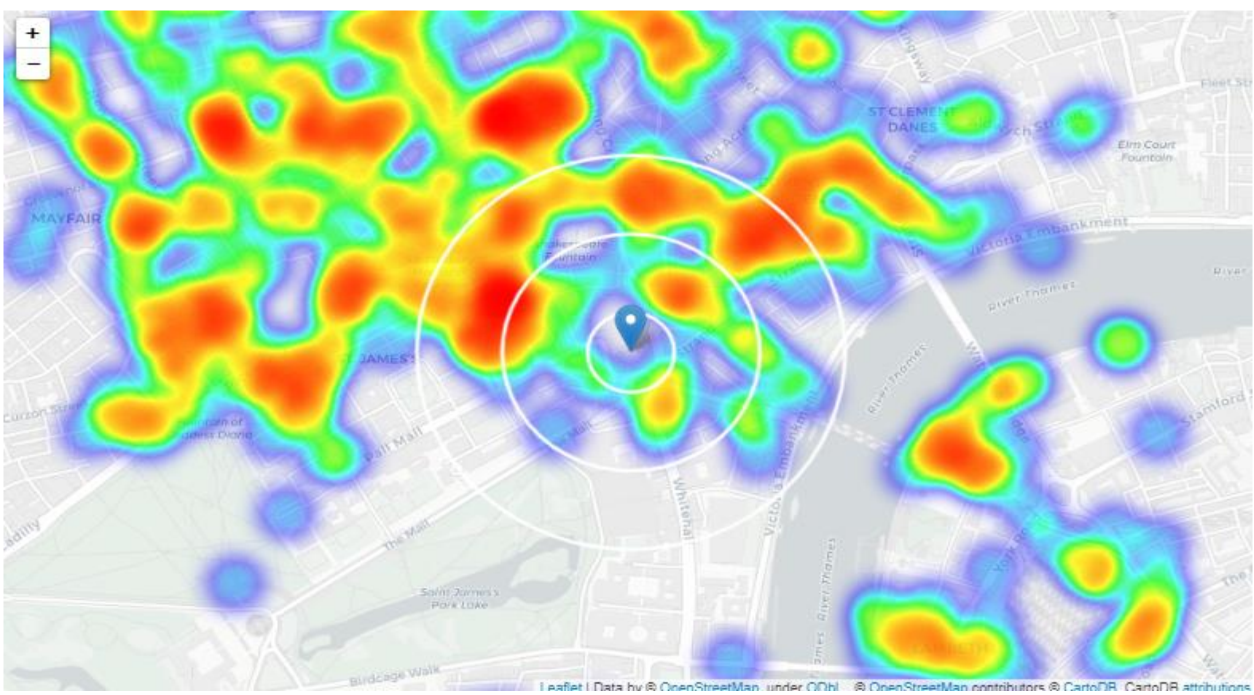

Figure7: heatmap map showing heatmap/density of restaurants.

We can clearly identify here an area with a lot of customers and few competitors >>> See the **green region**.

The innermost white circle indicates an area 100m, middle being 300m and outermost 500m wide.

IV. K mean clustering

Calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use radius of 150 meters) and number of customers.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Asian restaurant	Customers
288	51.510964	-0.123229	-546623.445676	5.815894e+06	13	31.091094	83
388	51.513971	-0.124248	-546623.445676	5.816241e+06	1	170.787947	83
338	51.512468	-0.123739	-546623.445676	5.816067e+06	2	99.800232	83
387	51.513788	-0.125640	-546723.445676	5.816241e+06	2	226.407080	75
289	51.511148	-0.121838	-546523.445676	5.815894e+06	11	37.124291	75
413	51.514758	-0.124239	-546604.445676	5.816327e+06	1	123.854904	74
263	51.510248	-0.122710	-546604.445676	5.815808e+06	11	79.329996	73
250	51.507859	-0.140792	-547904.445676	5.815808e+06	7	50.079146	72
337	51.512284	-0.125130	-546723.445676	5.816067e+06	3	99.780475	72
402	51.512737	-0.139541	-547704.445676	5.816327e+06	8	99.748425	72

Figure8: **number of restaurants in vicinity and number of customers.**

Now filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters, and no Asian restaurants in radius of 100 meters, and more than 10 customers.

see how this looks on a map.

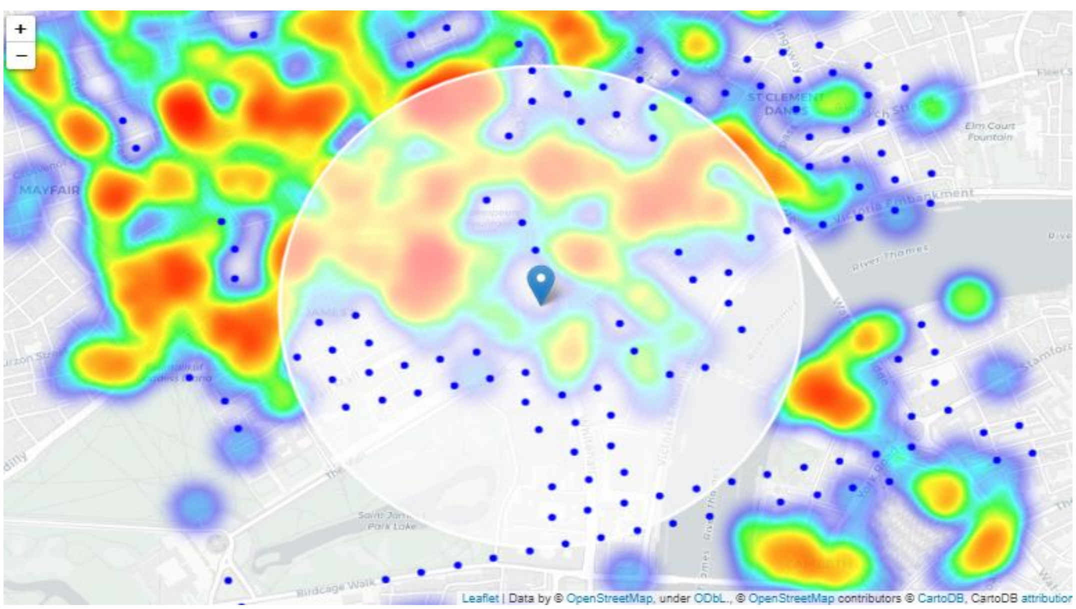

Figure9: **locations with no more than two restaurants and no Asian restaurants and more than 10 customers.**

We can identify mains areas which are east, north-east, south, south-east and even some in west as well.

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

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.

Figure10: **zones containing good locations**

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.

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.

Finally, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

Results and Discussion

This analysis shows that we must consider other criteria than just number of restaurants.

London is a very big city, the concentration of restaurant at the center is quite high due to the various famous landmarks situated there, in this analysis the idea was to correlate the number of restaurant and quantity of potential customer based on the factors of number of companies and universities nearby.

We first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Asian restaurant closer than 100m were removed (we had to keep the range small i.e., 100m because of the overly crowded restaurants at the center which was 655 restaurants in just 1.5X1.5km region around the London Centre).

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 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Asian 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 London center but not crowded with existing restaurants (particularly Asian) - 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

This project can be reused for other cities, just think about changing clustering size to adapt to your city.

Purpose of this project was to identify London areas close to center with potential customer count in order to aid stakeholders in narrowing down the search for optimal location for a new Asian restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general locations 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 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.

It's very far from being perfect, a lot of work can be done, other source of data can be found, but in the end the result seams to correlate with the real world, when we know the city, the area predicted seams correct.