Introduction - Business Problem

Where in London would it be best to open an Italian restaurant. What is the place that has the most reviews, what location seems to be the most popular. Is there a type of restaurant that is more popular than another?

Data

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

* number of existing restaurants in the neighbourhood (any type of restaurant)

* number of and distance to Italian restaurants in the neighbourhood, if any

* distance of neighbourhood from city centre

We decided to use regularly spaced grid of locations, centred around city centre, to define our neighbourhoods.

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

Centres of candidate areas will be generated algorithmically and approximate addresses of centres of those areas will be obtained using Google Maps API reverse geocoding

Number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API

Coordinate of London centre will be obtained using Google Maps API geocoding

The data that will be used is Foursquare data for London city. It will focus on venues of restaurant type. The popularity will be an indicator for how good an area is for a restaurant.

Methodology

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

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

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of London - we will use heatmaps to identify a few promising areas close to centre with low number of restaurants in general (and no Italian 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 Italian 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 / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Results and Discussion

Our analysis shows that although there is a great number of restaurants in London (~2000 in our initial area of interest), which offer a combination of popularity among tourists, closeness to city centre, strong socio-economic dynamics and a number of pockets of low restaurant density.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centres 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.

The results contain the largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian 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 but not crowded with existing restaurants (particularly Italian) - 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 London areas close to centre with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs 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 centres 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 neighbourhoods 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 neighbourhood etc.